IKERBASQUE UPVEHU

MOL2NET: FROM MOLECULES TO NETWORKS (PROCEEDINGS BOOK), Vol. 1.

sciforum

H. González-Díaz S. Arrasate N. Sotomayor
E. Lete F.P. Cossío E. Domínguez
D. Bonchev S.C. Basak D. Quesada
A. Duardo-Sanchez J.J. Chou C.M. Romeo-Casabona

Basel, Switzerland
ISBN: 978-3-03842-817-6
ISSN: 2624-5078
Year: **2018**

MOL2NET-01, doi: 10.3390/mol2net-02-05189, http://sciforum.net/conference/mol2net-01

[2015] [Committee] [Facebook] [Welcome Videos] [官话] [हिन्दी] [Euskera] [Castellano] [Português] [Français]

MOL2NET 2015, International Conference on Multidisciplinary Sciences

(1st edition)

MOL2NET: FROM MOLECULES TO NETWORKS (PROCEEDINGS BOOK), Vol. 1.

YEAR-ROUND CONFERENCE,
15 January–15 December 2016

Editors:

Prof. H. González-Díaz	Prof. S. Arrasate	Prof. N. Sotomayor
(UPV/EHU, Ikerbasque)	(UPV/EHU)	(UPV/EHU)
Prof. E. Lete	Prof. F.P. Cossío	Prof. E. Domínguez
(UPV/EHU)	(UPV/EHU, Ikerbasque)	(UPV/EHU)
Prof. D. Bonchev	Dr. S.C. Basak,	Prof. D. Quesada
(VCU, USA)	(UMN, USA)	(STU, USA, Miami)
Dr. A. Duardo-Sanchez	Prof. J.J. Chou	Prof. C.M. Romeo-Casabona
(UPV/EHU)	(Harvard Med School, USA)	(UPV/EHU)

Book Publication Date: **2018**-April-09
ISSN: 2624-5078, ISBN: 978-3-03842-817-6

MDPI Sciforum platform
http://sciforum.net/

MOL2NET, International Conference on Multidisciplinary Sciences

5–15 December 2015

Subscribe to Conference News

Subscribe to Conference Series News

Sections

All Contributions

00. Editorial from the Chairman

A. Research in Inorganic, Analytical, Physical, and Organic Chemistry

B. Medicinal Chemistry, Pharmacology, Biotechnology, and Drug Discovery

C. Soft Matter Physics, Polymers, Materials, and Nanosciences

D. Clinical Medical Sciences, Biomedical Engineering, and Medical Informatics

E. Statistics, Artificial Intelligence, Data Science, Complex Networks Analysis

F. Scientific Software

Conference Menu

Call for Papers

Instructions for Authors

Organizers

Sponsors

Full List of Presentations

Conference Discussions

Sister Conferences

Other Editions in This Series

Home » MOL2NET-1

MOL2NET, International Conference on Multidisciplinary Sciences

Welcome to the MOL2NET International Conference, 5–15 December 2015, Online

The full title of this conference is MOL2NET, the 1st International Conference on Multidisciplinary Sciences. MOL2NET (the conference's running title) will be held from 5-15 December 2015 on the Sciforum platform. This running title is inspired by the possibility of multidisciplinary collaborations in science between experimentalists and theoretical scientists; represented disciplines will encompass the molecular and biomedical sciences, social networks analysis, and beyond. More specifically, this conference aims to promote scientific synergies between groups of experimental molecular and bio-medical scientists. Relevant fields include chemistry, pharmacology, cancer research, proteomics, the neurosciences, the nanosciences, and epidemiology. Moreover, the conference welcomes computational and social sciences experts from different areas, such as computational chemistry, bioinformatics, social networks analysis, big data predictive analytics, biostatistics, *etc*.

The conference per se is the result of the synergy between the Department of Organic Chemistry, University of Basque Country (UPV/EHU), and IKERBASQUE, Basque Foundation for Sciences, with the Faculty of Informatics, University of Coruña (UDC). Accepted papers will be published in the proceedings of the conference, and selected papers will be considered for publication in the journal International Journal of Molecular Sciences (IJMS), ISSN 1422-0067, JCR Impact Factor 2014 IF = 2.86. This is an open access publication journal of MDPI in the field of Molecular and Biomedical Sciences (http://www.mdpi.com/journal/ijms). The link to the call for papers of this special issue is: http://www.mdpi.com/journal/ijms/special_issues/QSAR_QSPR_Chemistry. In addition, promotional videos in different languages and additional materials of related topics are expected to be released in the YouTube channel of MOL2NET http://bit.do/mol2net, in MOL2NET's Twitter account @mol2net, and in the Facebook group of the conference https://www.facebook.com/groups/chembioinfo.networks/, which has +8000 followers at this moment. We also invite all colleagues to share the conference website through social media. Some of the main topics of interest are explained in the section details.

Sincerely yours,

Conference Chairman
Prof. Humberto González-Díaz
Ikerbasque Senior Professor, humberto.gonzalezdiaz@ehu.eus

(1) Department of Organic Chemistry II, University of Basque Country / Euskal Herriko Unibertsitatea (UPV/EHU), 48940, Leioa, Sarriena w/n, Bizkaia.
(2) Ikerbasque - Basque Foundation for Science, 48011, Bilbao, Bizkaia, web: http://www.ikerbasque.net/humberto.gonzalez

Researchgate: http://www.researchgate.net/profile/Humbert_Gonzalez-Diaz

Prof. González-Díaz H. holds a position as Senior Ikerbasque Research Professor at the Department of Organic Chemistry II, University of the Basque Country UPV/EHU, Bizkaia Campus. This position is a research chair endowed by Ikerbasque, Basque Foundation for Science of the Basque Government / Eusko Jaurlaritza. Prof. González-Díaz H. obtained a PhD in Organic Chemistry in 2005 from the University of Santiago de Compostela (USC), and was supervised by Profs. E. Uriarte and L. Santana. He also received a Lic. degree in Pharmaceutical Sciences from the Central University of Las Villas (UCLV), where he was supervised by Prof. E. Estrada. Prof. González-Díaz H. is a Senior Research Professor with a Hirsch Index H > 50 (Google Scholar), >150 publications, >10 PhD theses supervised, >10 software developed, and 2 patents in nanoscience and neuroscience. He is also the Europe Editor of the journal *Curr Top Med Chem* and guest editor of several special issues. He has served as a pro bono consulting reviewer of research project proposals for various public agencies, including the US National Foundation for Science (FSA), the UK Biotechnology and Biological Sciences Research Council (BBSRC), and the German Federal Ministry of Education & Research (BMBF), *etc*.

Prof. González-Díaz H.'s research interests encompass multidisciplinary studies in chemoinformatics, networks, and data science. Specific interests include, but are not limited to, the application of network sciences and data analysis tools to the study of structure-property relationships in molecular structures, biological networks and complex systems in cheminformatics, bioinformatics, systems biology, neurosciences, nanosciences, omics, ecology, and epidemiology.

Sponsors

Find Us on Facebook Follow Us on Twitter Read our Blog

Sciforum is a platform maintained by MDPI.

Conference Organizers

Conference Chairman

Prof. Humberto González-Díaz
Ikerbasque Senior Professor, humberto.gonzalezdiaz@ehu.eus

(1) Department of Organic Chemistry II, University of Basque Country / Euskal Herriko Unibertsitatea (UPV/EHU), 48940, Leioa, Sarriena w/n, Bizkaia.
(2) Ikerbasque - Basque Foundation for Science, 48011, Bilbao, Bizkaia, web: http://www.ikerbasque.net/humberto.gonzalez

Researchgate: http://www.researchgate.net/profile/Humbert_Gonzalez-Diaz

Advisory Committee

Presidents (Experimental Sciences)
Prof. Esther Lete, esther.lete@ehu.eus
Department of Organic Chemistry II,
University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.
Researchgate: https://www.researchgate.net/profile/Esther_Lete

Prof. Nuria Sotomayor, nuria.sotomayor@ehu.es
Department of Organic Chemistry II,
University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.
Researchgate: https://www.researchgate.net/researcher/35044059_Nuria_Sotomayor

President (Theoretical Sciences)
Prof. Alejandro Pazos, Ph.D., D.M., alejandro.pazos@udc.es,
Chair and Director of Department of Computer Sciences,
University of Coruña (UDC), Coruña, Spain.
Researchgate: https://www.researchgate.net/profile/Alejandro_Pazos

President (Law, Ethics, Biosciences)
Prof CM Romeo Casabona, Full Professor of Law, carlosmaria.romeo@ehu.es
University of Basque Country (UPV/EHU), Bilbao, Bizkaia.
Director of Law & Human Genome Interuniversity Chair,
UPV/EHU-University of Deusto, Bilbao, Bizkaia.
Chairweb: http://www.catedraderechoygenomahumano.es/

Members of Honor

Prof. Fernando Cossío, President of IKERBASQUE, Basque Foundation for Science, Prof. Department of Organic Chemistry I, University of Basque Country (UPV/EHU), Donostia - San Sebastián Campus, Gipuzkoa.

Full Professor Esther Domínguez Pérez, Dean of Faculty of Science and Technology, Department of Organic Chemistry II, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Prof. Claudio Palomo Nicolau, Department of Organic Chemistry I, University of Basque Country (UPV/EHU), Donostia - San Sebastián Campus, Gipuzkoa.

Prof. Allen B. Reitz, Ph.D., CEO Fox Chase Chemical Diversity Center, Inc., Doylestown, PA, USA. Editor-in-Chief of the journal Current Topics in Medicinal Chemistry, Adjunct Professor at Drexel University College of Medicine, Moore Fellow in the Management of Technology University of Pennsylvania (Wharton, Penn Engineering), Founder CEO of ALS Biopharma, LLC.

Prof. James J. Chou, Department Biological Chemistry & Molecular Pharmacology (BCMP), Harvard Medical School, Boston, MA 02115, USA.

Prof. Dr. Peter Langer, Full Professor (C4) of Organic Chemistry, Vice Director Institute of Chemistry, University of Rostock, Head of the Department of Organic Chemistry. Universität Rostock Institut für Chemie Abteilung für Organische Chemie Albert-Einstein-Straße 3a 18059 Rostock.

Prof. Danail Bonchev, Director of Research on Bioinformatics, Center for the Study of Biological Complexity. Professor, Department of Mathematics and Applied Mathematics, Virginia Commonwealth University (VCU), USA.

Prof. Ernesto Estrada, Department of Mathematics and Statistics, Department of Physics, Chair in Complexity Sciences, Institute of Complexity Systems, University of Strathclyde Glasgow, G1 1XQ, UK. Editor-in-Chief of Journal of Complex Networks, Oxford Academic Press.

Prof. Jose María Pitarke, Full Professor of Condesed Matter Physics, UPV/EHU, Director of Nanomaterials Cooperative Research Center (CICNanoGune), Tolosa Hiribidea, 76, E-20018 Donostia – San Sebastian, Gipuzkoa.

Porf. Jesús Jimenez Barbero, Ikerbasque Professor, Scientific Director of Center for Cooperative Research in Biosciences (CICBiogune), Bizkaia.

Prof. Luis M Liz-Marzán, Ikerbasque Senior Professor, Scientific Director of Center for Cooperative Research in Biomaterials (CICbiomaGUNE), Gipuzkoa.

Prof. Victor M. Preciado, Raj and Neera Singh Term Assistant Professor Electrical and Systems Engineering (ESE), Penn Engineering, University of Pennsilvania, USA.

Prof. Roberto I Vazquez Padron, Ph.D, D.M., Research Associate Professor of Surgery, Molecular and Cellular Pharmacology, Miller School of Medicine, University of Miami, USA.

Prof. Mariano Provencio, Ph.D., D.M., Head of Medical Oncology Service, Universitary Hospital Puerta de Hierro (HUPH), Autonomous University of Madrid (UAM), Madrid, Spain.

Dr. A Rodríguez-Antigüedad Zarrans, Head of Medical Service of Neurology Hospital of Basurto, Bilbao, President Spain Society of Neurology (SEN). Prof. Department of Neuroscience, Faculty of Medicine UPV/EHU, Leioa, Sarriena w/n, Bizkaia.

Prof. Daniel Graham, Professor of Chemistry, Loyola University of Chicago, USA.

Prof. Roberto Todeschini, Head of Milano Chemometrics and QSAR Research Group. Professor of Chemometrics, Department of Environmental Sciences of the University of Milano-Bicocca, Milano, Italy.

Prof. Francesc Illas Riera, Director of Institute of Theoretical and Computational Chemistry (IQTCUB), Physical Chemistry Department, Faculty of Chemistry, University of Barcelona.

Prof. Kuo Chen Chou, Computational Biology, Gordon Life Science Institute, Belmont, MA 02478, USA, Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah 21589, Saudi Arabia.

Prof. Bairong Shen, Director of Center of Complex Systems, University of Soochow (SUDA), PRC, China. China Coordinator of SUDA-UVP/EHU Collaboration Agreement.

Prof. María Isabel Loza, Department of Pharmacology, USC, Pharmatools Digital Interactive Services, Allelyus, USEF Drug Screening Platform, President R+D Network for Medicines in Galicia, Santiago de Compostela, Spain.

Prof. Javier Luque Garriga, Department of Physical Chemistry, Faculty of Pharmacy and Institut of Biomedicine (IBUB), Universitat de Barcelona.

Prof. Jorge Galvez, Department of Physical Chemistry, Faculty of Pharmacy, Universitat de Valencia, Spain.

Prof. Subhash Chandra Basak, Ph.D. Senior Scientist and Adjunct Professor, Department of Chemistry & Biochemistry, University of Minnesota Duluth, Duluth, MN, USA.

Full. Prof. Yiyu Cheng, Director of Pharmaceutical Informatics Institute, Zhejiang University (ZJU), China.

Full. Prof. Miguel Ángel Gutiérrez Ortiz, Department of Chemical Engineering, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Prof. Pilar Goya, Institute of Medicinal Chemistry (IQM), CSIC, Juan de la Cierva st., Madrid.

Prof. Florencio M. Ubeira, M.D., Ph.D. Department of Microbiology and Parasitology, University of Santiago de Compostela (USC), Santiago, Spain.

Prof. Ramon García Domenech, Department of Physicalchemistry, Faculty of Pharmacy, Universitat de Valencia, Spain.

Prof. Luis Manuel Leon Isidro, Department of Physical Chemistry, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Prof. Néstor Etxebarria Loizate, Director Department of Analytical Chemistry, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Prof. Julio Seijas Vasquez, Department of Organic Chemistry, University of Santiago de Compostela (USC), Campus Lugo, Lugo, Spain. Chairman of Electronic Conference on Synthetic Organic Chemistry (ECSOC), SciForum, MDPI Switzerland.

Prof. Yagamare Fall, Department of Organic Chemistry, University of Vigo (UVIGO), Vigo, Spain. Africa Editor of Current Topics Medicinal Chemistry.

Neurosciences, University of the Basque Country (UPV/EHU), Head of Psychiatry Research of Osakidetza (Basque Public Health System), Head of Medical Psychiatry Service, University Hospital Santiago Apostol de Vitoria-Gasteiz, Vitoria.

Assoc. Prof. Jose Ramon Rey Caeiro, M.D., Head of Traumatology and Orthopedic Surgery, University Hospital of Santiago de Compostela (USC). SERGAS, Xunta de Galicia, University of Santiago de Compostela (USC).

M.D. Javier Castro Alvariño, Head of Department of Gastroenterology, Ferrol University Hospital Complex, SERGAS, Xunta de Galicia.

Scientific Committee

President (Experimental Sciences)
Assoc. Prof. Sonia Arrasate Gil, sonia.arrasate@ehu.eus
Department of Organic Chemistry II,
University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.
Researchgate: https://www.researchgate.net/profile/Sonia_Arrasate

Presidents (Theoretical Sciences)
Prof. Juan Ruso, juanm.ruso@usc.es
Faculty of Physics, University of Santiago de Compostela (USC), Spain.
Researchgate: https://www.researchgate.net/profile/Juan_Ruso

Prof. Natalia Cordeiro, Associate Professor with Habilitation, ncordeir@fc.up.pt
Theoretical Chemistry Network REQUIMTE / Department of Chemistry and Biochemistry, University of Porto, Portugal.
Researchgate: https://www.researchgate.net/profile/Natalia_Cordeiro2

Members

Assoc. Prof. Jose Luis Vicario, Department of Organic Chemistry II, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Assoc. Prof. Maria Luisa Carrillo Fernández, Department of Organic Chemistry II, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Assoc. Prof. Kazuhiro Takemoto, Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Iizuka, Fukuoka 820-8502, Japan.

Assistant Prof. William H Bisson, Ph.D. (Sr Res) Department of Environmental and Molecular Toxicology (EMT), College of Agriculture & Life Sciences Building, Oregon State University (OSU), Corvallis, OR, USA.

Assist. Prof. Efraim Reyes Martín, Department of Organic Chemistry II, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Assoc. Prof. Inmaculada Arostegui, Department of Applied Mathmatics and Statistics, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Assoc. Prof. David Quesada, PhD, Assoc. Prof. of Physics, School of Science, Technology, and Engineering Management, Saint Thomas University, Miami, USA.

Assoc. Prof. Aresatz Usobiaga, Academic Sec.of Department of Analytical Chemistry, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Assoc. Prof. Kunal Roy, Ph.D., Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India. Associate Editor Molecular Diversity, Springer. Research Fellow at Manchester Institute of Biotechnology, University of Machester (MIB), Manchester, United Kingdom.

Assoc. Prof. Angeles Sánchez-Glez Guerra, Department of Inorganic Chemistry, University of Santiago de Compostela (USC), Santiago, Spain.

Assoc. Prof. Jose M. Amigo, Prof. Department of Food Science, Quality & Technology. University of Copenhagen. Denmark.

Prof. Alessandro Giuliani, Instituto Superiore da Sanita (ISS), Rome, Italy.

Assoc. Prof. Xerardo García Mera, Department of Organic Chemistry, University of Santiago de Compostela (USC), Santiago, Spain.

Assoc. Prof. Maite Insausti, Department of Inorganic Chemistry, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Assoc. Prof. Julio Caballero, Center of Bioinformatics and Moleclar Simulation, Director of Department of Applied Sciences, University of Talca, Chile.

Ph.D. Sonsoles Martin-Santamaría, Staff Scientist, CIB-CSIC, Mardid, Spain.

Assoc. Prof. Jose Luis Vilas, Department of Physical Chemistry, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Assoc. Prof. Cristian Robert Munteanu, FIC, University of Coruña (UDC), Spain.

Assoc. Prof. Izaskun Gil de Muro, Academic Sec. Department of Inorganic Chemistry, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Ph.D. Jose Manuel Laza Terroba, Researcher Department of Physical Chemistry, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Ph.D. Andrey Toropov, Senior Researcher at Mario Negri Institute for Pharmacological Research, Milan, Italy.

Ph.D. Ailette Prieto Sobrino, Department of Analytical Chemistry, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Ph.D. M.D. Marta Arrasate, Dr. Psychiatry Osakidetza, Adjunct Prof. Department of Neuroscience, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Ph.D. Jose Manuel Laza Terroba, Researcher Department of Physical Chemistry, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Ph.D. Ricardo Grau-Crespo, Lecturer Department of Chemistry, University of Reading, Reading, United Kingdom.

Ph.D. Robersy Sánchez, Computational Biology Research Assistant Professor, Department of Agronomy & Horticulture, University of Nebraska–Lincoln, Lincoln, NE, USA.

Assoc. Prof. Marcus Tullius Scotti, Departamento de Engenharia e Meio Ambiente Centro de Ciências Aplicadas e Educação Universidade Federal da Paraíba, Campus IV, Paraíba, Brasil.

Prof. Carolina Horta Andrade, Head of LabMol Group, Faculdade de Farmacia, Universidade Federal de Goias, Setor Leste Universitario, Goiania, Brazil.

Ph.D. Anuraj Nayarisseri, Principal Scientist, Eminent Biosciences, Vijaynagar, Indore, India.

Assoc. Prof. Maité Sylla, Enseignant Chercheur, Conservatoire National des Arts et Métiers (CNAM), Paris, France

Assoc. Prof. Fernanda Borges, Department of Chemistry and Biochemistry, University of Porto, Portugal.

Prof. José Maria Monserrat, Instituto de Ciências Biológicas (ICB), Universidade Federal do Rio Grande - FURG, Rio Grande, RS, Brazil.

Assoc. Prof. Jose Luis Ayastuy Arizti, Department of Chemical Engineering, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Prof. Yovani Marrero-Ponce, Universidad Tecnológica de Bolívar (UTBVirtual), Colombia.

Prof. Alcidez Perez Bello, Fac. Veterinary Medicine, UCLV, Cuba / ESUDER, Universidade Eduardo Mondlane, Mozambique.

Full Prof. PhD. Francisco Javier Prado Prado, Division of Health Sciences, University of Qintana Roo (UQROO), Chetumal, Mexico.

Assoc. Prof. Julian Dorado, PhD. RNASA-IMEDIR Group. Universidade da Coruña, Institute of Biomedical Research of Coruña (INIBIC), Coruña University Hospital (CHUAC), A Coruña, Spain.

Assoc. Prof. PhD. Ana Porto Pazos. RNASA-IMEDIR Group, University of Coruña (UDC), Institute of Biomedical Research of Coruña (INIBIC), Coruña University Hospital (CHUAC), A Coruña, Spain.

Assoc. Prof. David Perez del Rey, PhD. Biomedical Informatics Group, Polytechnic University of Madrid, Spain.

Prof. Marcos Gestalt Pose, PhD. RNASA-IMEDIR Group. University of Coruña (UDC), Coruña University Hospital (CHUAC), A Coruña, Spain.

Prof. Salvador Fernández Pita, MD. PhD., Unit of Clinical Epidemiology and Biostatistics C. Coruña University Hospital (CHUAC), INIBIC, University of Coruña.

Organizing Committee

Coordinator (Theoretical Studies)
Assoc. Prof. Cristian Robert Munteanu, muntisa@gmail.com, Information and Communications Technologies Department, Faculty of Computer Science, University of A Coruna, Campus de Elviña s/n, 15071 A Coruña, Spain.
ResearchGate: http://www.researchgate.net/profile/Cristian_Munteanu

Coordinator (Experimental Studies)
Adjunct Prof. Uxue Uria Pujana, uxue.uria@ehu.eus
Department of Organic Chemistry II,
University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.
Researchgate: https://www.researchgate.net/researcher/38894030_Uxue_Uria

Coordinator (TICs, Legal Issues, Ethics & Biosciences)
PhD. Aliuska Duardo-Sánchez, PhD TICs, DEA & Lic. Law degrees, aliuska.duardo@usc.es
Department of Public Law, Faculty of Law, University of Santiago de Compostela (USC),
Santiago de Compostela, 15782, Spain.
Researchgate: https://www.researchgate.net/researcher/994320_Aliuska_Duardo-Sanchez

Committee Members

PhD. Jose A. Seoane, Stanford Cancer Institute, Stanford School of Medicine, Stanford University, Stanford, 94305, USA.

Ph.D. Vanessa Aguiar-Pulido, Florida International University, School of Computing and Information Sciences, Miami, FL, USA.

Adjunct Prof. Irantzu Barrio, Ph.D. Department of Applied Mathematics and Statistics, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Ph.D. Jesus vicente De Julián Ortiz, Staff Researcher I+D+i at Foundation Center of Innovation and Technologic Demonstration, Valencia, Spain.

Ph.D. Yasset Perez-Riverol, Post-Doc Researcher, Proteomics Services Team, PRIDE Group, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK.

Ph.D. Santiago Vilar, Research Associate, Department of Systems Biology, Columbia University, USA.

Ph.D. Hugo Gutierrez-de-Terán, Department of Cell and Molecular Biology, Uppsala University , Box 596, BMC, SE-751 24 Uppsala, Sweden.

PhD. Advait Apte, Scientific programmer, Department of Biology, City College of New York, New York, NY, USA.

Adjunct Prof. Lazaro Pino Rivero,Department of Chemistry, Physics, and Earth Sciences, Kendall Campus, Miami Dade College (MDC), Miami FL, USA.

Adjunct Prof. PhD. Eider Aranzamendi, Department of Organic Chemistry II, University of Basque Country (UPV/EHU), Bizkaia Campus, Bizkaia, Spain.

Ph.D. Diana Herrera Ibatá, Department of Computer Sciences, FIC, University of Coruña (UDC), Spain.

Ph.D. A Duardo-Sanchez, Department of Public Law, USC, Santiago de Compostela, Spain.

Ph.D. Verónica Ortiz de Elguea, Department of Organic Chemistry II, University of Basque Country (UPV/EHU), Bizkaia Campus, Bizkaia, Spain.

Ph.D. Oana Chis, Technological Institute for Industrial Mathematics (ITMATI), University of Santiago de Compostela (USC), Spain.

Ph.D. Ricardo Riccardo Concu, FCUP-Faculty of Sciences, University of Porto (UPORTO), Porto, Portugal.

M.D. Berkis Turiño Guerra, Aralia Servicios Sociosanitarios SA, Madrid

M.Sc. Enrrique Barreiro, Department of Computer Sciences, FIC, University of Coruña (UDC), Spain.
Maria Galvez-Llompart, Research Associate, Department of Physicalchemistry, Faculty of Pharmacy, Universitat de Valencia, Spain.

Riccardo Zanni, Research Associate, Department of Physicalchemistry, Faculty of Pharmacy, Universitat de Valencia, Spain.

Ph.D. Yong Liu, Visiting Post-Doc University of A Coruna, Spain. Research Collaborator Institute of Subtropical Agriculture, Chinese Academy of Sciences, Changsha, Hunan, China. Visiting Post-Doc University of Basque Country UPV/EHU, Leioa, Bizkaia.

Adjunct. Prof. Gerardo M. Casañola Martin. Facultad de Ingenieria Ambiental, Universidad Estatal Amazonica (UEA), Puyo, Ecuador.

PhD. Juan Alberto Castillo Garit, Visting Prof. of Bioinformatics Research Carleton University, Ottawa, Canada.

MSc. Michel González-Durruthy, Instituto de Ciências Biológicas (ICB), Universidade Federal do Rio Grande - FURG, Rio Grande, RS, Brazil.

Prof. Maykel Cruz-Monteagudo, CIQ/REQUIMTE, Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal and Institute of Biomedical Investigations (IIB), University of Las Américas , 170513 Quito, Pichincha, Ecuador.

Ph.D. Virginia Mato Abad, LAIMBIO, University Rey Juan Carlos, Madrid, Spain.

PhD. Javier Pereira Loureiro. RNASA-IMEDIR Group. University of Coruña (UDC), Institute of Biomedical Research of Coruña (INIBIC), Coruña University Hospital (CHUAC), A Coruña, Spain.

M.D. Paula Peleteiro Higuero, Oncology and Radiotherapy, University Hospital of Santiago de Compostela, SERGAS, Xunta de Galicia, Santiago de Compostela

M.D. José Luis Ulla-Rocha, Department of Gastroenterology, University Hospital of Pontevedra. SERGAS. Xunta de Galicia

M.D. Xavier Romero Duran, Service of Neurology, IMQ Zorrotzaure Clinic, Bilbao, Bizkaia, Spain.

PhD. MD. Javier Saavedra. Family doctor. EOXI - SERGAS. Galician Government, Institute of Biomedical Research (INIBIC), Coruña, Spain.

PhD. Juan Ramon Rabuñal Dopico. RNASA-IMEDIR Group, University of Coruña (UDC), Coruña University Hospital (CHUAC), A Coruña, Spain.

PhD. Carlos Fernandez Lozano. RNASA-IMEDIR Group, University of Coruña (UDC), Coruña University Hospital (CHUAC), A Coruña, Spain.

PhD. Rajeev K. Singla, SERB-Young Scientist/PI, Division of Biotechnology, Netaji Subhas Institute of Technology, Sec-3, Dwarka, New Delhi-110078, India.
Editor of Indo Global Journal of Pharmaceutical Sciences.

PhD. Enrique Fernández Blanco, RNASA-IMEDIR Group, University of Coruña (UDC), Coruña University Hospital (CHUAC), A Coruña, Spain.

PhD. José Manuel Brea Floriani, BioPharma, REGID- Innopharma, CIMUS, University of Santiago de Compostela, Spain.

PhD. Daniel Torrecilla, BioPharma, REGID- Innopharma, CIMUS, University of Santiago de Compostela, Spain.

Sponsors

A: Research in Inorganic, Analytical, Physical, and Organic Chemistry

This section covers: both experimental and theoretical research in Chemistry including, but not limited to, Inorganic, Analytical, Physical, and Organic Chemistry. The topics are expected to be of wide scope. For instance, in experimental studies: Organic synthesis, Chemical reactivity, Catalysis, Solid State Chemistry, Inorganic Crystals, Crystal Symmetry, and Complexes. Physicochemistry and Analytical chemistry techniques; Spectroscopy (X-Ray, NMR, IR, EPR, Mass Spectroscopy), Chromatography and sample preparation techniques, TEM and SEM Microscopy. Also, in theoretical and computational studies: Computational Chemistry, Quantum Mechanics (*Ab inition*, DFT, MP3, AM1 methods), Monte Carlo (MC) algorithm, Quantitative Structure-Reactivity, Structure-Property, or Structure-Retention Relationships (QSPR/QSRR) models in organic, inorganic, physical and analytical chemistry. Chemometrics, Experimental Design, and Data Analysis in Analytical chemistry.

List of presentations (15)

a001 **Influence of synthetic conditions in the hydrothermal preparation of TiO_2 nanotubes** show abstract
by Izaskun Gil de muro *, Amaia Ereño , Maite Insausti

a002 **HPLC-qTOF-MS platform as valuable tool for the exploratory characterization of phenolic compounds in guava leaves at different oxidation states** show abstract
by Elixabet Díaz-de-Cerio *, Vito Verardo , Ana María Gómez-Caravaca , Alberto Fernández-Gutiérrez , Antonio Segura-Carretero

a003 **Chiral imines on the wave: reactivity of *tert*-butyl acrylate and stereoselectivity determination using NMR in liquid crystals** show abstract
by Lucie VANDROMME , Li CHEN , Lai WEI , Franck LE BIDEAU , André LOUPY , Olivier LAFON , Philippe LESOT , Marie-Elise TRAN HUU DAU , Pierre CHAMINADE , Françoise DUMAS *

a004 **A Palladium NCP Pincer Complex as an Efficient Catalyst for Intramolecular Direct Arylation** show abstract
by Nerea Conde , Fátima Churruca , Raul Sanmartin *, María Teresa Herrero , Esther Domínguez

a005 **Efficient aerobic oxidation of arylcarbinols and arylmethylene compounds mediated by Nickel (II)/1,2,4-triazole ligand catalyst system** show abstract
by Garazi Urgoitia , Raul Sanmartin *, María Teresa Herrero , Esther Domínguez

a006 **Regioselective Friedel-Crafts hydroalkylation using friendly conditions: Application to the synthesis of unsymmetrical triarylmethanes** show abstract
by Maité Sylla-Iyarreta Veitía *, Céline Rampal , Clotilde Ferroud

a007 Asymmetric Mizoroki-Heck reactions: generation of quaternary stereocenters and cascade cyclizations show abstract
by Ane Rebolledo Azcargorta , Iratxe Barbolla , Estíbaliz Coya , Nuria Sotomayor , Esther Lete *

a008 Diastereoselective formation of tertiary stereocenters *via* Mizoroki-Heck reaction show abstract
by Ane Rebolledo Azcargorta , Esther Lete , Nuria Sotomayor *

a009 **2-Nitromethylacrylates as Useful Dinucleophiles for the Enantioselective Organocatalytic Michael/Henry Cascade Reaction** show abstract
by Naiara Fernández , Iker Riaño , Uxue Uria , Efraim Reyes , Luisa Carrillo , Jose Luis Vicario *

a010 **Uptake of different organic pollutants by carrot** show abstract
by Ekhiñe Bizkarguenaga Bizkarguenaga *, Luis Ángel Fernández , Olatz Zuloaga , Ailette Prieto

a011 Performance of the NOF theory in the description of the four-electron harmonium atom in the singlet state show abstract
by Mario Piris *

a012 **Fluorinated Nucleosides (Mini Review)** show abstract
by Mohamed Ibrahim Elzagheid *

a013 **Perturbation Theory Modeling of Intramolecular Carbolithiation Reactions** show abstract
by Asier Gómez-SanJuan , Sonia Arrasate , Garazi Urgoitia *, Nuria Sotomayor , Esther Lete

a014 **Synthesis and Platinum (II) Complexes of Different Polyazacyclophane Receptors** show abstract
by Begoña Verdejo *, Javier Pitarch-Jarque , Estefanía Delgado-Pinar , Lorena Magraner-Pardo , Jesus Vicente de Julián-Ortiz , Enrique García-España

a015 Molecular Rearrangement of an Aza-Scorpiand Macrocycle Induced by pH. A Computational Study show abstract
by Jesus Vicente de Julián-Ortiz *, Begoña Verdejo , Víctor Polo , Emili Besalú , Enrique García-España

Sponsors

About Contact Terms of Use Privacy Policy

Find Us on Facebook Follow Us on Twitter Read our Blog

Sciforum is a platform maintained by MDPI.

B: Medicinal Chemistry, Pharmacology, Biotechnology, and Drug Discovery

This section covers: Experimental and theoretical methods applied to drug discovery, biomarkers and target validation, vaccine design, in biosciences. In experimental studies: Pharmacological assays, Toxicity and Cytotoxicity studies, Molecular Biology and Biotechnology. Proteomics, Genomics, and Metabolomics (OMICS methods) like Sequencing, Cloning, DNA microarrays, and Mass Spectroscopy in Clinical Proteomics. Also, theoretical studies: Molecular Mechanics and Molecular Dynamics (MM/MD) for Drug-protein Docking studies, Quantitative Structure-Activity / Toxicity Relationships (QSAR / QSTR) models. Bioinformatics analysis of Disease Biomarkers and Computational vaccine design (Alignment and Alignment-free techniques). Determination of the 3D proteins structure using NMR and X-ray techniques. Experimental and computational study of RNA (Rnomics), secondary RNA structure prediction, miRNA biomarkers.

List of presentations (38)

b001 Chemoinformatics Profiling of Ionic Liquids Cytotoxicity—From Machine Learning to Network-Like Similarity Graphs show abstract
by Maykel Cruz-Monteagudo *, Eduardo Tejera , Cesar Paz-y-Miño , Yunierkis Perez-Castillo , Aminael Sánchez-Rodríguez , Fernanda Borges ,María Natália Días Soeiro Cordeiro

b002 **Complex networks of anti-HIV drugs activity vs. prevalence of AIDS in US Counties using symmetry information indices** show abstract
by Diana Maria Herrera-Ibatá *, Ricardo Alfredo Orbegozo-Medina

b003 **Computational study of mycobacterial promoters with low sequence homology.** show abstract
by Alcides Pérez-Bello *

b004 Fatty Acids Distribution Networks in Ruminal Membrane by Computational and Experimental Studies show abstract
by Yong Liu *, Zhi Liang Tan , Claudia Giovanna Peñuelas-Rivas , Esvieta Tenorio-Borroto

b005 *In Silico* Design of New Drugs for Myeloid Leukemia Treatment show abstract
by Washington Pereira , Ihosvany Camps *

b006 **Interdependence of Influenza HA and NA and possibilities of new reassortments** show abstract
by Ashesh Nandy *, Subhas Basak

b007 QSPR-perturbation models for the prediction of B-epitopes from immune epitope database: an interesting route for predicting *"in silico"* new optimal peptide sequences and/or boundary conditions for vaccine development. show abstract
by Severo Vázquez-Prieto *, Esperanza Paniagua , Florencio M. Ubeira

b008 **Towards computational prediction of Biopharmaceutics Classification System: a QSPR approach** show abstract
by Hai Pham-The *, Huong Le-Thi-Thu , Teresa Garrigues , Marival Bermejo , Isabel González-Álvarez , Miguel Ángel Cabrera-Pérez

b009 Information Signatures of Viral Proteins: A Study of Influenza A Hemagglutinin and Neuraminidase show abstract
by Daniel J. Graham *, Samuel Barlow , Diego F. Cuculon , Jordan C. Hauck

b010 **^{14}N NMR Spectroscopy for detection of binding interaction between sodium azide and hydrated Fullerene by titration method** show abstract
by Tamar Chachibaia *, Manuel Martin Pastor

b011 Appying a novel web-tool for performing virtual screening experiments show abstract
by Karina S. Machado , Vinicius Rosa Seus *, Jorge Gomes

b012 **Prediction of the total antioxidant capacity of food based on artificial intelligence algorithms** show abstract
by Estela Guardado Yordi *, Raul Koeling , Yailé Caballero Mota , Maria João Matos , Lourdes Santana , Eugenio Uriarte , Enrique Molina

b013 **A Proposal Tool for Manipulation of a Set of Protein Structures from PDB** show abstract
by Vinicius Rosa Seus *, Karina S. Machado , Adriano Velasque Werhli

b015 **An insight to segment based genetic exchange in Influenza A virus: an *in silico* study** show abstract
by Antara De *, Ashesh Nandy

b016 Phylogenetic and genetic analysis of envelope gene of the prevalent Dengue serotypes in India in recent years show abstract
by Ashesh Nandy , Sumanta Dey *

b017 **Phosphorylated Sites on the Disordered Interface Signatures the Interacting Behavior of Proteins - A Comparative Mapping of Phosphorylation Propensities on Disordered Interfaces of Interactome and Negatome** show abstract
by Srinivas Bandaru , Deepika Ponnala , Chandana Lakkaraju , Chaitanya Kumar Bhukya , Uzma Shaheen , Anuraj Nayarisseri *

b018 **Docking Studies and ADMET Profile of Streblusol E, Anti-hepatitis B viral Agent of *Streblus Asper*** show abstract
by Rajeev K Singla *, Rohit Gundamaraju , Baishakhi De , Varadaraj Bhat G

b019 **Homology modeling, Molecular Dynamic Simulation and in silico screening of Activator for the Intensification of human sirtuin type 1 (SIRT1) by novel 1, 3, 4-thiadiazole derivatives-A potential antiaging approach** show abstract
by Vinit Raj *, Sudipta Saha , amit rai , Mahendra Singh , Durgesh Kumar , Anil Kumar Sahdev

b020 **Multi-target Prediction of Neuroprotective Drugs, Synthesis, Assay, and Theoretical Study of Rasagiline Carbamates** show abstract
by Francisco J Romero Durán *, Nerea Alonso , Olga Caamaño , Matilde Yañez , Xerardo Garcia-Mera

b021 **Conception, synthesis, characterization and antimicrobial evaluation of new ferrocene-based derivatives inspired by the bisacodyl lead structure** show abstract
by Meral Görmen , Maité Sylla-Iyarreta Veitía *, Fatma Trigui , Mehdi El Arbi , Clotilde Ferroud

b022 Effect of neuronal nitric oxide synthase inhibitors and antioxidants on the development of tolerance by different opioid agonists in the rat locus coeruleus show abstract
by Patricia Pablos , Aitziber Mendiguren , Joseba Pineda *

b023 **Pharmacological characterization of the prostanoid receptor EP3 in locus coeruleus neurons by single-unit extracellular recordings in the rat brain in vitro** show abstract
by Amaia Nazabal , Aitziber Mendiguren *, Joseba Pineda *

b024 **DPPH• Free Radical Scavenging Activity of Coumarin Derivatives. *In Silico* and *in Vitro* Approach** show abstract
by Elizabeth Goya Jorge , Anita Maria Rayar , Stephen Jones Barigye , María Elisa Jorge Rodríguez , Maité Sylla-Iyarreta Veitía *

b025 **QSAR for the characterization of drug resistance: Differential QSAR (DiffQSAR) using mathematical chemodescriptors** show abstract
by Subhash C. Basak *

b026 **Pharmacokinetics and Toxicological Profiling of Surfactin A: An *In silico* Approach** show abstract
by Rajeev K Singla , Ashok K Dubey *

b027 ***New insights from the CoMSIA analysis within the framework of Density Functional Theory.*** show abstract
by Alejandro Morales-Bayuelo *, Julio Caballero

b028 **Microwave Activated Synthesis of Benzalacetones and Study of Their Potential Antioxidant Activity Using Artificial Neural Networks Method** show abstract
by Anita Maria Rayar , Elizabeth Goya Jorge , Stephen Jones Barigye , María Elisa Jorge Rodríguez , Clotilde Ferroud , Maité Sylla-Iyarreta Veitía *

b029 Fragment-based approach for affinity and selectivity for dUTPase: Insights for design of new anti-malarial agents show abstract
by Marilia Nunes Nascimento , Marina Rocha Martins , Rodolpho Campos Braga , Bruno Júnior Neves , Vinicius Medeiros Alves , Carolina Horta Andrade *

b031 **Improved virtual screening performance through docking scoring fusion in the discovery of dual target ligands for Parkinson's disease** show abstract
by Yunierkis Pérez-Castillo *, Aliuska Morales-Helguera , M. Natália D. S. Cordeiro , Eduardo Tejera , Cesar Paz-y-Miño , Aminael Sánchez-Rodríguez , Fernanda Borges , Maykel Cruz-Monteagudo

b032 Dengue NS5 global consensus sequence development to find conserved region for antiviral drug development show abstract
by Shahid Mahmood , Usman Ali Ashfaq *

b033 **Virtual screening tailored ensembles of QSAR models for the discovery of dual A_{2A} Adenosine Receptor Antagonists / Monoamine Oxidase B Inhibitors** show abstract
by Aliuska Morales-Helguera , Yunierkis Pérez-Castillo , M. Natália D. S. Cordeiro , Eduardo Tejera , Cesar Paz-y-Miño , Aminael Sánchez-Rodríguez, Marta Teijeira-Bautista , Evys Ancede-Gallardo , Fernando Cagide , Fernanda Borges *, Maykel Cruz-Monteagudo *

b034 **Development of QSAR models for identification of CYP3A4 substrates and inhibitors** show abstract
by Flávia Cristina Silva , Ekaterina Varlamova , Rodolpho Campos Braga , Carolina Horta Andrade *

b035 **Histones Bind, Aggregate and Fuse Phosphoinositides Containing Bilayers** show abstract
by Marta G. Lete *, Hasna Ahyayauch , Jesús Sot , Felix M. Goni , Alicia Alonso

b036 **Two QSAR Paradigms- Congenericity Principle versus Diversity Begets Diversity Principle- analyzed using computed mathematical chemodescriptors of homogeneous and diverse sets of chemical mutagens** show abstract
by Subhabrata Majumdar , Subhash C. Basak , Subhash C. Basak *, Subhash C. Basak *, Subhash C. Basak *

b037 **Intrinsic dimensionality of chemical space: Characterization and applications** show abstract
by Gregory D. Grunwald , Subhash C. Basak , Subhash C. Basak *

b038 Hierarchical quantitative structure-activity relationships (HiQSARs) for the prediction of physicochemical and toxicological properties of chemicals using computed molecular descriptors show abstract
by Subhabrata Majumdar , Subhash C. Basak , Subhash C. Basak , Subhash C. Basak , Subhabrata Majumdar , Subhash C. Basak *

b039 Study of Dried Blood Spot reliability for quantitative drug analysis by UHPLC-PDA-FLUO show abstract
by Beatriz Uribe , Oihane Elena Alboniga , Oskar Gonzalez *, Rosa María Alonso *

b040 **An unprecedented revolution in medicinal science** show abstract
by Kuo-Chen Chou *

List of Accepted Abstracts (5)

The microRNA regulatory network: a far-reaching approach to theregulate the Wnt signaling pathway in number of diseases show abstract
by Shahid Mahmood , Attya Bhatti *

Anxiolytic-like effects of 7*H*-benzo[*e*]perimidin-7-one derivatives through elevated plus-maze test in mice show abstract
by José Ángel Fontenla , Seyed M. Nabavi , Eduardo Sobarzo-Sánchez *

Computational prediction of thermolysin inhibitors using multiple linear regression according to *OECD* principles show abstract

by Yudith Cañizares-Carmenate , Juan Alberto Castillo-Garit *, Yovani Marrero-Ponce , Dayan Machado Aguila , Francisco Torrens

Modeling the interactions between mitochondrial voltage-dependent anion cannel (VDAC) and single walled carbon nanotubes using molecular docking simulation with virtual screening framework show abstract

by Michael González Durruthy *, Vinicius Rosa Seus , Adriano Velasque Werhli , Karina S. Machado , Luisa Cornetet , José Montserrat

Two QSAR Paradigms- Congenericity Principle versus Diversity Begets Diversity Principle- analyzed using computed mathematical chemodescriptors of homogeneous and diverse sets of chemical mutagens. show abstract

by Subhash C. Basak *, Subhabrata Majumdar

Sponsors

About Contact Terms of Use Privacy Policy

Find Us on Facebook Follow Us on Twitter Read our Blog

Sciforum is a platform maintained by MDPI.

C: Soft Matter Physics, Polymers, Materials, and Nanosciences

This section covers: experimental and/or theoretical analysis of artificial polymers, biopolymers, materials, nanomaterials, etc. Experimental and theoretical study of carbon nanomaterials (Graphene, Fullerenes, Nanotubes, Diamonoids), Ceramic materials, Alloys. Biopolymeric Nanomaterials for biosciences (drug carriers, diagnosis tools, medical imaging) including Dendrimer, Protein nanoparticle. Shape Memory Polymers, Nanopatterning, and Surface Imprinting.

List of presentations (13)

c001 **Kinetic study of activated carbon synthesis from Marabou Wood** show abstract
by Pedro Julio Villegas *

c002 The symmetry-adapted configurational ensemble approach to the computer simulation of site-disordered solids show abstract
by Ricardo Grau-Crespo *, Said Hamad

c003 **Enhancement of photovoltage generation, storage capacity and energy conversion efficiency of photoelectrochemial cell of mixed dye system: Role of oxidized multi-walled carbon nanotubes** show abstract
by Poonam Bandyopadhyay *, Ruma Basu , Sukhen Das , Papiya Nandy

c004 **Application of triturated copper nanoparticles as an agent for remediation of an azo dye, methyl orange** show abstract
by Monalisa Chakraborty *, Ruma Basu , Sukhen Das , Papiya Nandy

c005 **Cobalt Alumino Silicate Ceramic(CASC) nanocomposite, a material with moderately high dielectric constant and low tangent loss at a critical concentration in high frequency range.** show abstract
by Biplab Kumar Paul , Smarajit Manna , Debasis Roy , Papiya Nandy , Sukhen Das *

c006 Green processing of nanoporous biodegradable carriers of bioactive agents for pharmaceutical and biomedical applications show abstract
by Jorge Santos , Pasquale del Gaudio , Mariana Landin , Carlos A García-González *

c007 **CORAL: The dispersion of SWNTs in different organic solvents** show abstract
by Alla P. Toropova , Andrey A. Toropov *

c008 **3D hierarchically scaffolds for bone repair: at the crossroads of experimental and computational outlooks** show abstract
by Paula Messina , Juan M. Ruso *

c009 Nanoparticulate Fe$_2$O$_3$ and Fe$_2$O$_3$/C Composites as Anode Materials for Li-Ion Batteries show abstract
by Amaia Iturrondobeitia , Aintzane Goñi , Luis Lezama *

c010 **Combination of microscopic and spectroscopic techniques to study the presence and the effects of microplastics in mussels** show abstract
by Mireia Irazola Duñabeitia *, Larraitz Garmendia Altuna , Beñat Zaldibar Aranburu , Urtzi Izagirre Aramayona , Eider Bilbao Castellanos , Sara Danielsson , Anders Bignert , Kepa Castro Ortiz de Pinedo , Nestor Etxebarria Loizate , Manu Soto Lopez , Ionana Marigomez Allende

c011 Synthesis and characterization of shape memory polyurethanes show abstract
by Míriam Sáenz-Pérez *, José Manuel Laza , Luis Manuel León , Jorge García-Barrasa , José Luis Vilas

c012 Shape memory behaviour of a gamma-irradiated commercial polycyclooctene. show abstract
by Nuria García-Huete *, José Manuel Laza , José María Cuevas , José Luis Vilas , Luis Manuel León

c013 Building a New High-Selective Molecular Imprinted Polymer show abstract
by Riccardo Concu *, Maria Natalia Dias Soeiro Cordeiro *

Sponsors

About Contact Terms of Use Privacy Policy
Find Us on Facebook Follow Us on Twitter Read our Blog
Sciforum is a platform maintained by MDPI.

D: Clinical Medical Sciences, Biomedical Engineering, and Medical Informatics

This section covers: experimental and/or computational medical diagnostic tools in cancer research, neurosciences, clinic and biomedical engineering. Including, but not limited to, EEG and structural NMR in clinical diagnosis in neurology and brain research. EEG, fNMRI, microscopy, tomography, study for tissue connectivity analysis, including the use of experimental techniques and complex networks computational analysis in neurosciences, bone tissue connectivity, vascular system connectivity, etc.

List of presentations (3)

d001 **Prognostic value of affective symptomatology in first-admitted psychotic patients** show abstract
by Marta Arrasate *, Itxaso González-Ortega , Adriana García-Alocén , Susana Alberich , Iñaki Zorrilla , Ana González-Pinto

d002 **Synthesis and characterization of Carbon nanotube/Hydroxyapatite/Clay based Hybrid antimicrobial biomaterial for potential tissue engineering application** show abstract
by subrata kar *, Papiya Nandy , Ruma Basu , sukhen Das

d003 Do we use well benzodiazepines in elderly ?: a case report show abstract
by MARIA JOSE DIAZ GUTIERREZ *

List of Accepted Abstracts (2)

Biometeorology of asthma: A multi-scale and complexity approach. show abstract
by David Quesada *

Multiple drug resistant malaria and its effects on hemoglobin and CD4-lymphocytes of HIV-seropositive pregnant women at Kaduna state, Nigeria show abstract
by Idris Abdullahi Nasir *, Maryam Muhammad Aliyu

Sponsors

About Contact Terms of Use Privacy Policy
Find Us on Facebook Follow Us on Twitter Read our Blog

Sciforum is a platform maintained by MDPI.

E: Statistics, Artificial Intelligence, Data Science, Complex Networks Analysis

This section covers: connectivity analysis in biology, environment, epidemiological, and social networks; including the computational analysis of metabolic pathways in Metabolomics, Protein interaction networks in proteomics, food webs, and other biological-ecological networks like host-parasite, prey-hunter, etc. Geographical Information Systems (GIS), land covering networks, atmospheric reactions networks. Study of social collaboration, electronic social networks (Facebook, Twitter, etc.), disease spreading networks and epidemiology, vaccination models in epidemic networks, legal and law citing networks, networks in sociology and criminology, etc. This section covers also: technological, industrial, and economic connectivity, including the analysis of computer connectivity, Internet, wireless networks, satellite networks, electrical networks, airport and other transport networks, financial networks, trade networks, etc. In addition, we cover pure theoretical aspects in network science and data analysis theory, including but not limited to theoretical studies in network sciences, topological indices, node centrality, network robustness, multiplex networks, network attack, and new spatial statistical analysis, time series analysis, biostatistics, machine learning and big data analysis methods.

List of presentations (12)

e001 Scheduler for SANN Analysis of U.S. Supreme Court Network Based on Markov-Shannon Entropy show abstract
by Aliuska Duardo-Sanchez *

e002 Pairwise Ortholog Detection in Related Yeast Species by Using Big Data Supervised Classifications show abstract
by Deborah Galpert Cañizares , Sara del Río García , Francisco Herrera , Evys Ancede Gallardo , Agostinho Antunes , Guillermin Agüero-Chapin *

e003 Genome-wide Discriminatory Information Patterns of Cytosine DNA Methylation show abstract
by Robersy Sanchez *, Sally A Mackenzie

e004 Categorisation of continuous variables in a logistic regression model using the R package CatPredi show abstract
by Irantzu Barrio *, María Xosé Rodríguez-Álvarez , Inmaculada Arostegui

e005 **Multi-viral targets entropy QSAR for antiviral drugs** show abstract
by Francisco Javier Prado Prado *, Xerardo García-Mera

e006 **A Proposal about Normalization of Experimental Designs in Computational Intelligence** show abstract
by Carlos Fernandez-Lozano *, Julián Dorado , Marcos Gestal

e007 An alternative approach to structure specification based on fuzzy multidimensional membership function using forward selection rule show abstract
by SREYASI GHOSH *, SARBARI GHOSH , pradip kumar sen , subhra chatterjee

e008 **Machine Learning and Atom-Based Quadratic Indices for Proteasome Inhibition Prediction** show abstract
by Gerardo Maikel Casañola-Martin *, Huong Le-Thi-Thu , Facundo Perez-Gimenez , Concepción Abad

e009 **Computational Models of the Brain** show abstract
by Lucas A Pastur-Romay , Francisco A Cedrón , Alejandro Pazos , Ana B Porto-Pazos *

e010 **Iterative Kernel K-means for Metagenomic Sequences** show abstract
by Isis Bonet *, Andrea Mesa-Múnera *, Adriana Escobar , Juan Fernando Alzate

e011 **New theoretical model for the study of new β-secretase inhibitors** show abstract
by Jan-carlo Miguel Díaz-González , Francisco Javier Aguirre-Crespo , Xerardo García-Mera , Francisco Javier Prado Prado *

e012 A Computer-Aided SAS Macro for the Evaluation of the Simulation Performances in Missingness Settings show abstract
by Urko Aguirre *, Inmaculada Arostegui , Jose M. Quintana

List of Accepted Abstracts (1)

A revision of statistical methodology in experimental sciences show abstract
by Nerea Gutiérrez *, Irantzu Barrio , José M. Lacave , Amaia Orbea , Inmaculada Arostegui

Sponsors

About Contact Terms of Use Privacy Policy
Find Us on Facebook Follow Us on Twitter Read our Blog
Sciforum is a platform maintained by MDPI.

F: Scientific Software

This section is aimed at presenting the most commonly used software tools in Multidisciplinary Science. Include, but is not limited to, new scientific software, web servers, databases, etc. with applications in Chemistry (all branches), Bioinformatics, Proteomics, Biotechnology, Medical Informatics and Biomedical Engineering, Computer Science, etc.

The short communications should present computational tools that may be desktop/web/mobile applications/scripts, open code or private software. The tool may be original or a pipe of other tools. It should contain a software description, case uses in order to understand how to employ it, links to the open repositories (GitHub, GitLab, Personal Webs, etc.) or official Webs of the private products, and references of the publications where the tools have been applied. The authors may include in the communication a link to their personal webs, web servers, repositories, databases, etc.

Special attention will be paid to the links to tutorials (blogs, videos, etc.), print screens with the tools in action, pseudocodes, examples of input and outputs, script examples while using the tools, and links to the social network posts for the tools. The emphasis of this section is on the software per se. Communications that make use of a software to solve a practical problem but do not put emphasis on describing it could be suitable for other sections.

Enjoy programming for science!

SECTION COORDINATORS:

Assoc. Prof. Cristian Robert Munteanu, muntisa@gmail.com, Information and Communications Technologies Department, Faculty of Computer Science, University of A Coruna, Campus de Elviña s/n, 15071 A Coruña, Spain.

PhD. Yasset Perez Riverol, Bioinformatician - Hermjakob team, yperez@ebi.ac.uk, EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK.

IKERBASQUE Prof. Humberto González-Díaz, humberto.gonzalezdiaz@ehu.eus, Department of Organic Chemistry II, University of Basque Country UPV/EHU, 48940, Leioa, Sarriena w/n, Bizkaia, Spain.

List of presentations (14)

F001 Evaluation of computational tools for thermodynamics and structural analysis of protein stability upon point mutation prediction show abstract
by Alex Camargo *, Karina Machado *, Adriano Werhli *

F002 **In silico computer simulation risk assessment of triazole fungicides on human cytochrome p450 aromatase enzyme: cyp19a1 inhibition by triazoles using autodock software.** show abstract
by Tamar Chachibaia *, Joy Harris Hoskeri

F003 ***MetAlgNet :***Metabolic pathway network reconstruction from algae genome annotation data. show abstract
by Kirtan Dave *, Darshan Choksi , Hetalkumar Panchal

F004 **Bio-AIMS Chemoinformatics Web tools for proteins** show abstract
by Cristian R. Munteanu , Humberto González-Díaz , Carlos Fernandez-Lozano , José Antonio Seoane Fernández , José M. Vázquez-Naya *, Mabel Loza , Alejandro Pazos

F005 **Genetic Algorithms with Fine Tuning** show abstract
by Marcos Gestal *, Julián Dorado

F006 **Law, Software, & Cheminformatics: Copyright, Taxes, and Legal Issues** show abstract
by Aliuska Duardo-Sanchez *, Antonio Lopez-Diaz

F007 **AlzPred-SVV: Free Web Tool for Alzheimer Prediction using Spectroscopy Voxel Volume** show abstract
by Virginia Mato Abad *, Cristian Robert Munteanu , Carlos Fernández-Lozano , Alejandro Pazos

F009 Using the RRegrs R package for automating predictive modelling show abstract
by Georgia Tsiliki *, Cristian R Munteanu , Jose A Seoane , Carlos Fernandez-Lozano , Haralambos Sarimveis , Egon L Willighagen

F010 Prot-SSP: a tool for amino acid pairing pattern analysis in secondary structures show abstract
by Miguel de Sousa *, Cristian Robert Munteanu , Alexandre Lopes Magalhães

F011 TI2BioP: Topological Indices to BioPolymers show abstract
by Guillermin Agüero-Chapin *, Reinaldo Molina Ruiz , Agostinho Antunes

F012 **Solvent Accessible Surface Area Hot-Spot Detection Method** show abstract
by Cristian R Munteanu *, António C. Pimenta , Carlos Fernandez-Lozano , André Melo , Maria Natália Cordeiro , Irina S. Moreira *

F013 SISTEMAT X - A web tool to manage databases of secondary metabolites show abstract
by Marcus Tullius Scotti *, Roberto Oliveira Da Silva Junior , Silas Yudi , Rafael Brayner , Luciana Scotti

F014 **ASD Module: a software to support the personal autonomy in the daily life of children with autism spectrum disorder** show abstract
by Betania Groba *, Javier Pereira , Laura Nieto , Thais Pousada , Susana Falcón , Cristian Robert Munteanu , Alejandro Pazos

F015 **Pro-ChInt: Machine Learning Methods for Identifying Dual- / Multi- Protein Chain Interactions with Python** show abstract
by Yong Liu *

Sponsors

Sponsors

About Contact Terms of Use Privacy Policy

Find Us on Facebook Follow Us on Twitter Read our Blog

Sciforum is a platform maintained by MDPI.

SciForum
Mol2Net

Scheduler for SANN Analysis of U.S. Supreme Court Network Based on Markov-Shannon Entropy

Aliuska Duardo-Sanchez

Department of Special Public Law, University of Santiago de Compostela (USC), 15782 Santiago de Compostela, Spain; E-Mail: aliuska.duardo@usc.es

Received: 10 September 2015 / Accepted: 14 September 2015 / Published: 2 December 2015

Abstract: Many complex systems may be represented as complex networks of ith parts or nodes (ni) interconnected by some kind of bonds, ties, relationships, links (Lij). For instance, Fowler *et al.* represented all case citations (Lij) in the U.S. Supreme Court as a network of nj cases citing and/or cited by other. These huge collections of nodes/links are impossible to remember and rationalize by a single person in order to assign correct links in new situations. Fortunately, Artificial Neural Networks (ANNs) can help us in this task. If we want use ANNs to predict links in complex networks, first we need to transform all the information into numerical input parameters to feed ANNs, second: we need to find the best ANN to predict our network. We can solve the first problem quantifying the structural information of the complex system (Brain, Ecological, Social, *etc.*) with universal information measures known as Shannon entropy (Sh). We can quantify topological (connectivity) information of both the complex networks under study and a set of ANNs trained using Shannon measures. Then using both sets of information parameters as inputs we can develop a dual QSPR model to discriminate between SANNs and not efficient ANN topologies. Here we used this QSPR method to develop potential HPC schedulers for complex systems. We studied 663072 citations to majority opinions in 43 sub-networks; each one with 5,000 (5K) citations to U.S. Supreme Court decisions (5KCNs). The overall accuracy of the ANN found was of >85% for 5KCNs; in training and validation series.

Keywords: SANN Scheduler; Markov-Shannon Entropy; U.S. Supreme Court; Social Network Analysis.

1. Introduction

Many important systems, in center of attention of modern science, may be approached as complex networks of ith parts or nodes (*ni*) interconnected by some kind of links (*Lij*), bonds, ties, or relationships [1-7]. The diversity of systems susceptible to be studied by complex networks is very high; e.g.,: Human brain [8], Ecosystems [9-11], or the citations to U.S. Supreme Court decisions [12]. All these collections of nodes and links are so large that it is impossible for a person to remember and rationalize all possible connections. We can solve this problem using Quantitative Structure-Activity/Property Relationships (QSAR/QSPR) models [13-18]. In QSAR/QSPR modeling we can represent the system as a graph of interconnected nodes and use as inputs theoretic-information parameters that quantify information about the structure of the graph. In this context, Shannon entropy quantifies the information contained in a message, usually in units such as bits [19-36]. The software MARCH-INSIDE (Markov Chains Invariants for Network Simulation and Design) has become a very useful tool for QSAR/QSPR studies [37-45].

In this occasion we selected the dataset of Fowler *et al.* [12], represented all case citations (Lij) in the U.S. Supreme Court as a network of nj cases citing and/or cited by other. Fowler used a complex network approach to quantify links in citations between cases and unravel the most relevant precedents. The work opens the door to the use of complex network structural parameters like topological indices and/or information measures to predict the future citation behavior of state courts, the U.S. Courts of Appeals, the U.S. Supreme Court, as well as other legal systems [45, 60-62].

The number of nodes and connections in complex systems is very large and the problem of prediction of correct links may become a computationally expensive task for large collections of complex systems. Artificial Neural Networks (ANNs) can help us in this task. ANNs are powerful bio-inspired algorithms able to learn/infer large datasets. There many examples of applications of ANNs to seek QSPR-like models [63-66]. ANNs can mange, for example, to learn to discriminate the correct collections of nodes (nj) and links present in complex systems (Lij) from other connectivity patterns not correct and/or distributed at random. We have at least two major problems if we want use ANNs to predict links in bio-systems and complex other networks. First, we need to transform all the information into numerical input parameters to feed ANNs. Next, we have to train many ANNs to detect which topology is better learning the structure of the system under study.

In this work, we introduce a new type of algorithm to solve this problem. The idea is simple: if ANNs are networks with nodes (neurons) and links (functions) we should treat them as such. In so doing, we can quantify topological (connectivity) information of both the complex networks under study and a set of ANNs trained using Shannon measures. Using both sets of information parameters as inputs we can develop a dual Quantitative Structure-Property Relationship (QSPR) model to discriminate between SANNs and not efficient ANN topologies. Here we used this QSPR method to develop potential HPC scheduler for complex systems. We studied 663072 pairs in 43 sub-networks; each one with 5,000 (5K) citations to U.S. Supreme Court decisions (5KCNs). The overall accuracy of the SANN-HPC schedulers found was of >81% for 5KCNs; in training and validation series (see Figure 1). This report of QSPR models potentially useful as task schedulers for HPC or Cloud Computing of ANNs with the subsequent can help to safe time and computational resources in the prediction of Complex Networks.

Linear Discriminant Analysis (LDA) models: Once the values of the Markov-Shannon entropies were obtained, we carried out a Linear Discriminant Analysis (LDA) by means of the STATISTICA software [76]. Let be qS(Lij) the output variable of a HPC schedule model used to score the ability of a given ANNq to predict correctly the link Lij between two nodes i-th and j-th (Lij = 1). We can use LDA to seek a linear equation with coefficients aik, bjk, cqk, dijqk, and e0. These are the coefficients of the Shannon entropies for the first node (Shik), for the second node (Shjk), and for the ANN graph (Shk(ANNq)), used as input for the LDA model. The k subindex indicates that this Shk value codify information for all nodes placed at least at topological distance d = k from the node of reference. We can use different statistical parameters to evaluate the statistical significance and validate the goodness-of-fit of LDA equation: n = number of cases, $\chi 2$ = Chi-square, p = the error level, as well as the Accuracy, Specificity, and Sensitivity of both train and external validation series [77]. We can write the LDA equation with the parameters mentioned above in the following form, see also **Figure 1**.

2. Results and Discussion

Social Network Analysis (SNA) emerged in 1930 to become one of the more powerful tools in socials sciences [80]. With the rise of network search, commerce, consume, and socialization companies like Google, Facebook, Twitter, LinkedIn, Amazon, and others, SNA have become a very important tool for the analysis of the high amount of information of users interactions in the web [refs]. However, the application of these methods in legal studies is still at the beginning [81, 82]. Network tools may illustrate the interrelation between the different law types/judicial cases and help to understand law/judicial cases effect in the legal system and its effectiveness to regulate aspects of necessity in society or not. We have applied the present methodology the design new schedulers for HPC of ANN models useful to predict one important legal network. The example selected was the USSCC network and the best model found was:

Where Shk(Lti) and θk(Lti+1) are the entropy parameters that quantify information about the Legal norms (Laws) of type L introduced in the Spanish legal system at time ti and ti+1 with respect to the previous or successive kth norms approved. The model behaves like a time series embedded within a complex network. This is because it predicts the recurrence of the Spanish law system to a financial norm of class c when socio-economical conditions change at time ti+1 given that have been used a known class of norm in the past at time *ti*. The model correctly re-constructed the network of the historic record for the Spanish financial system with high Accuracy, Specificity, and Sensitivity (**Table 1**).

Table 1. Results of models for USSC network.

Model	Training Series				Model	Cross-Validation Series			
Param.[a]	%	Class	$L_{ij} = 0$	$L_{ij} = 1$	Param.[a]	%	Class	$L_{ij} = 0$	$L_{ij} = 1$
Sp	92.8	$L_{ij} = 0$	**219919**	17161	Sp	92.8	$L_{ij} = 0$	**74552**	5780
Sn	73.8	$L_{ij} = 1$	41449	**116831**	Sn	73.0	$L_{ij} = 1$	13817	**37391**
Ac	85.2	Total			Ac	85.1	Total		

Rows: Observed classifications; Columns: Predicted classifications; C_{ij} = Calculation with high priority; NC_{ij} = No C_{ij}.

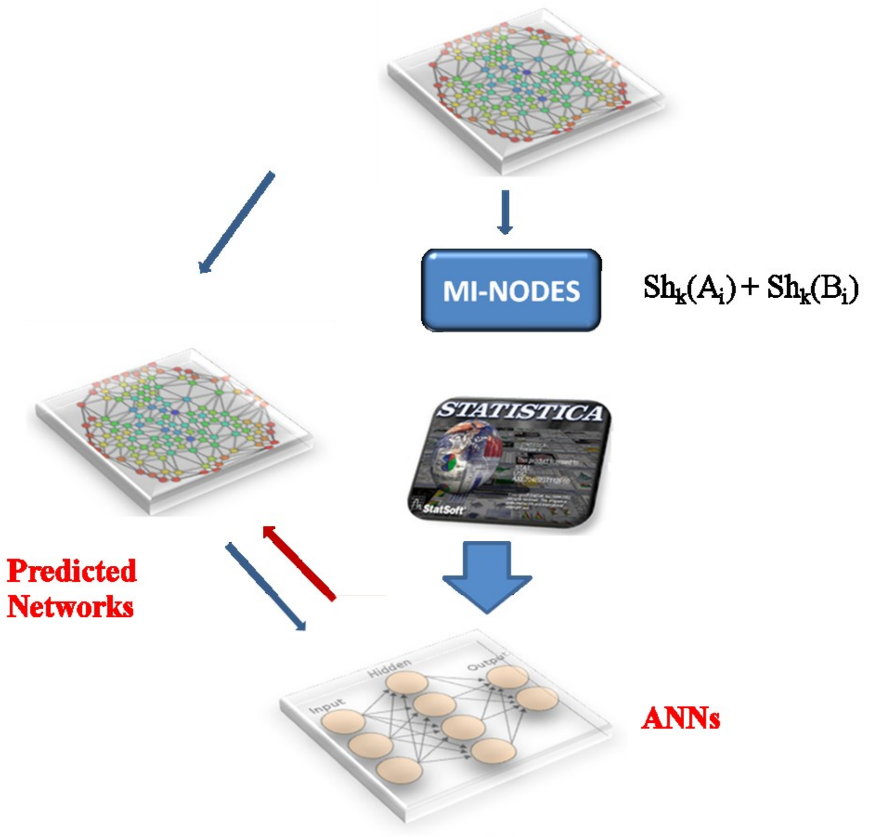


Figure 1. General workflow used in this work to develop new ANN for USSC network.

3. Materials and Methods

Datasets: U.S. Supreme Court (USSC) Network. We used a complex network constructed by Fowler et al. [75]. The authors included 26,681 majority opinions written by the U.S. Supreme Court. The dataset contains all cases that cite this U.S. Supreme Court decisions from 1791 to 2005. In this network each case is represented by a node. The links between two nodes Ai and Bj (arcs) express that the case jth cites the ith case previous to it (precedent). In order to both make more tractable the dataset for computation of Shk(Ai) and Shk(Bj) values and focused on specific intervals of time we split the data in 43 sub-networks. Each one represent one slot of 5000 (5K) citing cases, 5K-Citations Network (5KCNs) for > 22,000 cases of the U.S. Supreme court.

Computational Methods: Markov-Shannon Entropy Centralities for nodes.

We construct the classical Markov matrix (1Π) for each network as follows. First, we download from public resources the connectivity matrix L or obtain the data about the links between the nodes to assemble L (n by n matrix, where n is the number of vertices). Next, the Markov matrix Π is built. It contains the vertices probability (pij) based on L. The probability matrix is raised to the power k, resulting (1Π)k, and multiplied by the vector of the initial probabilities (0pj). The resulting vectors contain the absolute probabilities to reach the nodes moving throughout a walk of length k from node ni (kpj) for each k and are the base for the entropy centrality (Shk) calculation:

4. Conclusions

In this work we confirm that it is possible to combine Markov Chains and Shannon Entropy in order to calculate higher order entropy parameters. We also show that these parameters can be used to quantify information about local and global node-node connections in different types of complex networks. For it, we have used MI-NODES, a new tool for the study of complex networks which is an upgrade of the software MARCH-INSIDE, classically used to study drugs and proteins.

Conflicts of Interest

The author declares no conflict of interest.

References and Notes

[1] K.S. Sandhu, G. Li, H.M. Poh, Y.L. Quek, Y.Y. Sia, S.Q. Peh, F.H. Mulawadi, J. Lim, M. Sikic, F. Menghi, A. Thalamuthu, W.K. Sung, X. Ruan, M.J. Fullwood, E. Liu, P. Csermely, Y. Ruan, Large-scale functional organization of long-range chromatin interaction networks, Cell Rep, 2 (2012) 1207-1219.

[2] M.E. Gaspar, P. Csermely, Rigidity and flexibility of biological networks, Brief Funct Genomics, 11 (2012) 443-456.

[3] P. Csermely, T. Korcsmaros, H.J. Kiss, G. London, R. Nussinov, Structure and dynamics of molecular networks: A novel paradigm of drug discovery: A comprehensive review, Pharmacol Ther, 138 (2013) 333-408.

[4] M. Vidal, M.E. Cusick, A.L. Barabasi, Interactome networks and human disease, Cell, 144 (2011) 986-998.

[5] A.L. Barabasi, N. Gulbahce, J. Loscalzo, Network medicine: a network-based approach to human disease, Nat Rev Genet, 12 (2011) 56-68.

[6] A.L. Barabasi, Z.N. Oltvai, Network biology: understanding the cell's functional organization, Nat Rev Genet, 5 (2004) 101-113.

[7] S.H. Strogatz, Exploring complex networks, Nature, 410 (2001) 268-276.

[8] J.C. Reijneveld, S.C. Ponten, H.W. Berendse, C.J. Stam, The application of graph theoretical analysis to complex networks in the brain, Clin Neurophysiol, 118 (2007) 2317-2331.

[9] E. Borenstein, M.W. Feldman, Topological signatures of species interactions in metabolic networks, J Comput Biol, 16 (2009) 191-200.

[10] R.E. Ulanowicz, Quantitative methods for ecological network analysis, Comput Biol Chem, 28 (2004) 321-339.

[11] H. Olff, D. Alonso, M.P. Berg, B.K. Eriksson, M. Loreau, T. Piersma, N. Rooney, Parallel ecological networks in ecosystems, Philos Trans R Soc Lond B Biol Sci, 364 (2009) 1755-1779.

[12] J.H. Fowler, T.R. Johnson, J.F.S. II, S. Jeon, P.J. Wahlbeck, Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court, (2007).

[13] P. Riera-Fernandez, R. Martin-Romalde, F.J. Prado-Prado, M. Escobar, C.R. Munteanu, R. Concu, A. Duardo-Sanchez, H. Gonzalez-Diaz, From QSAR models of Drugs to Complex Networks: State-of-Art Review and Introduction of New Markov-Spectral Moments Indices., Curr Top Med Chem, 12 (2012) 927-960.

[14] H. Gonzalez-Diaz, QSAR and Complex Networks in Pharmaceutical Design, Microbiology, Parasitology, Toxicology, Cancer and Neurosciences, Current Pharmaceutical Design, 16 (2010) 2598-U2524.

[15] H. González-Díaz, F. Prado-Prado, L.G. Pérez-Montoto, A. Duardo-Sánchez, A. López-Díaz, QSAR Models for Proteins of Parasitic Organisms, Plants and Human Guests: Theory, Applications, Legal Protection, Taxes, and Regulatory Issues, Curr Proteomics, 6 (2009) 214-227.

[16] A. Speck-Planche, V.V. Kleandrova, F. Luan, M.N. Cordeiro, Multi-target drug discovery in anti-cancer therapy: fragment-based approach toward the design of potent and versatile anti-prostate cancer agents., Bioorg Med Chem, 19 (2011) 6239-6244.

[17] A. Speck-Planche, V.V. Kleandrova, F. Luan, M.N. Cordeiro, Chemoinformatics in Multi-Target Drug Discovery for Anti-Cancer Therapy: In Silico Design Of Potent And Versatile Anti-Brain Tumor Agents., Anticancer Agents Med Chem, (2011).

[18] F.J. Prado-Prado, F.M. Ubeira, F. Borges, H. Gonzalez-Diaz, Unified QSAR & Network-Based Computational Chemistry Approach to Antimicrobials. II. Multiple Distance and Triadic Census Analysis of Antiparasitic Drugs Complex Networks, Journal of Computational Chemistry, 31 (2010) 164-173.

[19] C.E. Shannon, A Mathematical Theory of Communication, The Bell System Technical Journal, 27 (1948) 379-423.

[20] M. Dehmer, F. Emmert-Streib, Analysis of Complex Networks. From Biology to Linguistics, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2009.

[21] M. Dehmer, M. Grabner, K. Varmuza, Information indices with high discriminative power for graphs, PLoS One, 7 (2012) e31214.

[22] M. Dehmer, K. Varmuza, S. Borgert, F. Emmert-Streib, On entropy-based molecular descriptors: statistical analysis of real and synthetic chemical structures, Journal of chemical information and modeling, 49 (2009) 1655-1663.

[23] E. Estrada, D. Avnir, Continuous symmetry numbers and entropy., J Am Chem Soc, 125 (2003) 4368-4375.

[24] D.J. Graham, S. Grzetic, D. May, J. Zumpf, Information properties of naturally-occurring proteins: Fourier analysis and complexity phase plots, Protein J, 31 (2012) 550-563.

[25] D.J. Graham, J.L. Greminger, On the information expressed in enzyme structure: more lessons from ribonuclease A, Mol Divers, 15 (2011) 769-779.

[26] D.J. Graham, J.L. Greminger, On the information expressed in enzyme primary structure: lessons from Ribonuclease A, Mol Divers, 14 (2010) 673-686.

[27] D.J. Graham, M. Kim, Information and classical thermodynamic transformations, J Phys Chem B, 112 (2008) 10585-10593.

[28] D.J. Graham, C. Malarkey, W. Sevchuk, Experimental investigation of information processing under irreversible Brownian conditions: work/time analysis of paper chromatograms, J Phys Chem B, 112 (2008) 10594-10602.

[29] D.J. Graham, Information Content in Organic Molecules: Brownian Processing at Low Levels, Journal of chemical information and modeling, 47 (2007) 376-389.

[30] D.J. Graham, Information content in organic molecules: aggregation states and solvent effects, J Chem Inf Model, 45 (2005) 1223-1236.

[31] D.J. Graham, M.V. Schulmerich, Information Content in Organic Molecules: Reaction Pathway Analysis via Brownian Processing, J Chem Inf Comput Sci, 44 (2004).

[32] D.J. Graham, C. Malarkey, M.V. Schulmerich, Information Content in Organic Molecules: Quantification and Statistical Structure via Brownian Processing. , J. Chem. Inf. Comput. Sci., 44 (2004).

[33] D.J. Graham, Information and organic molecules: structure considerations via integer statistics, J Chem Inf Comput Sci, 42 (2002) 215-221.

[34] D.J. Graham, D.V. Schacht, Base information content in organic formulas, J Chem Inf Comput Sci, 40 (2000) 942-946.

[35] S.J. Barigye, Y. Marrero-Ponce, O.M. Santiago, Y.M. Lopez, F. Perez-Gimenez, F. Torrens, Shannon's, Mutual, Conditional and Joint Entropy Information Indices. Generalization of Global Indices Defined from Local Vertex Invariants, Curr Comput Aided Drug Des, (2013).

[36] V. Aguiar-Pulido, C.R. Munteanu, J.A. Seoane, E. Fernández-Blanco, L.G. Pérez-Montoto, H. González-Díaz, J. Dorado, Naïve Bayes QSDR classification based on spiral-graph Shannon entropies for protein biomarkers in human colon cancer., Mol Biosyst, (2012).

[37] H. Gonzalez-Diaz, A. Duardo-Sanchez, F.M. Ubeira, F. Prado-Prado, L.G. Perez-Montoto, R. Concu, G. Podda, B. Shen, Review of MARCH-INSIDE & Complex Networks Prediction of Drugs: ADMET, Anti-parasite Activity, Metabolizing Enzymes and Cardiotoxicity Proteome Biomarkers, Curr Drug Metab, 11 (2010) 379-406.

[38] C.R. Munteanu, A.L. Magalhaes, E. Uriarte, H. Gonzalez-Diaz, Multi-target QPDR classification model for human breast and colon cancer-related proteins using star graph topological indices, J. Theor. Biol., 257 (2009) 303-311.

[39] C.R. Munteanu, H. Gonzalez-Diaz, F. Borges, A.L. de Magalhaes, Natural/random protein classification models based on star network topological indices, J. Theor. Biol., 254 (2008) 775-783.

[40] C.R. Munteanu, H. Gonzalez-Diaz, A.L. Magalhaes, Enzymes/non-enzymes classification model complexity based on composition, sequence, 3D and topological indices, J. Theor. Biol., (2008) 476-482.

[41] F.J. Prado-Prado, I. Garcia, X. Garcia-Mera, H. Gonzalez-Diaz, Entropy multi-target QSAR model for prediction of antiviral drug complex networks, Chemometrics and Intelligent Laboratory Systems, 107 (2011) 227-233.

[42] H. González-Díaz, A. Pérez-Bello, M. Cruz-Monteagudo, Y. González-Díaz, L. Santana, E. Uriarte, Chemometrics for QSAR with low sequence homology: Mycobacterial promoter sequences recognition with 2D-RNA entropies, Chemom Intell Lab Systs, 85 (2007) 20-26.

[43] H. González-Díaz, L. Saíz-Urra, R. Molina, E. Uriarte, Stochastic molecular descriptors for polymers. 2. Spherical truncation of electrostatic interactions on entropy based polymers 3D-QSAR, Polymer, 46 (2005) 2791–2798.

[44] Y. Rodriguez-Soca, C.R. Munteanu, J. Dorado, J. Rabunal, A. Pazos, H. Gonzalez-Diaz, Plasmod-PPI: A web-server predicting complex biopolymer targets in plasmodium with entropy measures of protein-protein interactions, Polymer, 51 (2010) 264-273.

[45] P. Riera-Fernandez, C.R. Munteanu, M. Escobar, F. Prado-Prado, R. Martin-Romalde, D. Pereira, K. Villalba, A. Duardo-Sanchez, H. Gonzalez-Diaz, New Markov-Shannon Entropy models to assess connectivity quality in complex networks: from molecular to cellular pathway, Parasite-Host, Neural, Industry, and Legal-Social networks, J. Theor. Biol., 293 (2012) 174-188.

[46] D.C. Van Essen, K. Ugurbil, E. Auerbach, D. Barch, T.E. Behrens, R. Bucholz, A. Chang, L. Chen, M. Corbetta, S.W. Curtiss, S. Della Penna, D. Feinberg, M.F. Glasser, N. Harel, A.C. Heath, L. Larson-Prior, D. Marcus, G. Michalareas, S. Moeller, R. Oostenveld, S.E. Petersen, F. Prior, B.L. Schlaggar, S.M. Smith, A.Z. Snyder, J. Xu, E. Yacoub, The Human Connectome Project: a data acquisition perspective, Neuroimage, 62 (2012) 2222-2231.

[47] E.W. Lang, A.M. Tome, I.R. Keck, J.M. Gorriz-Saez, C.G. Puntonet, Brain connectivity analysis: a short survey, Comput Intell Neurosci, 2012 (2012) 412512.

[48] M.P. Richardson, Large scale brain models of epilepsy: dynamics meets connectomics, J Neurol Neurosurg Psychiatry, 83 (2012) 1238-1248.

[49] A. Fornito, A. Zalesky, C. Pantelis, E.T. Bullmore, Schizophrenia, neuroimaging and connectomics, Neuroimage, 62 (2012) 2296-2314.

[50] D.S. Modha, R. Singh, Network architecture of the long-distance pathways in the macaque brain, Proc. Natl. Acad. Sci. U. S. A., 107 (2010) 13485-13490.

[51] J.D. Yeakel, P.R. Guimaraes, Jr., M. Novak, K. Fox-Dobbs, P.L. Koch, Probabilistic patterns of interaction: the effects of link-strength variability on food web structure, J R Soc Interface, 9 (2012) 3219-3228.

[52] V. Gagic, S. Hanke, C. Thies, C. Scherber, Z. Tomanovic, T. Tscharntke, Agricultural intensification and cereal aphid-parasitoid-hyperparasitoid food webs: network complexity, temporal variability and parasitism rates, Oecologia, 170 (2012) 1099-1109.

[53] R.J. Williams, D.W. Purves, The probabilistic niche model reveals substantial variation in the niche structure of empirical food webs, Ecology, 92 (2011) 1849-1857.

[54] F. Jordan, Keystone species and food webs, Philos Trans R Soc Lond B Biol Sci, 364 (2009) 1733-1741.

[55] R.E. Ulanowicz, Some steps toward a central theory of ecosystem dynamics, Comput Biol Chem, 27 (2003) 523-530.

[56] H. Gonzalez-Diaz, P. Riera-Fernandez, A. Pazos, C.R. Munteanu, The Rucker-Markov invariants of complex Bio-Systems: applications in Parasitology and Neuroinformatics, Biosystems, 111 (2013) 199-207.

[57] H. Gonzalez-Diaz, P. Riera-Fernandez, New Markov-Autocorrelation Indices for Re-evaluation of Links in Chemical and Biological Complex Networks used in Metabolomics, Parasitology, Neurosciences, and Epidemiology, Journal of Chemical Information and Modeling, 52 (2012) 3331-3340.

[58] I. Riera-Fernandez, R. Martin-Romalde, F.J. Prado-Prado, M. Escobar, C.R. Munteanu, R. Concu, A. Duardo-Sanchez, H. Gonzalez-Diaz, From QSAR models of Drugs to Complex Networks: State-of-Art Review and Introduction of New Markov-Spectral Moments Indices, Current Topics in Medicinal Chemistry, (2012).

[59] P. Riera-Fernandez, C.R. Munteanu, M. Escobar, F. Prado-Prado, R. Martin-Romalde, D. Pereira, K. Villalba, A. Duardo-Sanchez, H. Gonzalez-Diaz, New Markov-Shannon Entropy models to assess connectivity quality in complex networks: From molecular to cellular pathway, Parasite-Host, Neural, Industry, and Legal-Social networks, Journal of Theoretical Biology, 293 (2012) 174-188.

[60] P. Riera-Fernandez, C.R. Munteanu, J. Dorado, R. Martin-Romalde, A. Duardo-Sanchez, H. Gonzalez-Diaz, From Chemical Graphs in Computer-Aided Drug Design to General Markov-Galvez Indices of Drug-Target, Proteome, Drug-Parasitic Disease, Technological, and Social-Legal Networks, Current Computer-Aided Drug Design, 7 (2011) 315-337.

[61] P. Riera-Fernández, C.R. Munteanu, N. Pedreira-Souto, R. Martín-Romalde, A. Duardo-Sanchez, H. González-Díaz, Definition of Markov-Harary Invariants and Review of Classic Topological Indices and Databases in Biology, Parasitology, Technology, and Social-Legal Networks, Current Bioinformatics, 6 (2011) 94-121.

[62] A. Duardo-Sanchez, G. Patlewicz, H. González-Díaz, A Review of Network Topological Indices from Chem-Bioinformatics to Legal Sciences and back, Current Bioinformatics, 6 (2011) 53-70.

[63] H. Gonzalez-Diaz, I. Bonet, C. Teran, E. De Clercq, R. Bello, M.M. Garcia, L. Santana, E. Uriarte, ANN-QSAR model for selection of anticancer leads from structurally heterogeneous series of compounds, European Journal of Medicinal Chemistry, 42 (2007) 580-585.

[64] M. Jalali-Heravi, M.H. Fatemi, Prediction of thermal conductivity detection response factors using an artificial neural network, J Chromatogr A, 897 (2000) 227-235.

[65] F.J. Prado-Prado, X. Garcia-Mera, H. Gonzalez-Diaz, Multi-target spectral moment QSAR versus ANN for antiparasitic drugs against different parasite species, Bioorganic & Medicinal Chemistry, 18 (2010) 2225-2231.

[66] E. Tenorio-Borroto, C.G. Penuelas Rivas, J.C. Vasquez Chagoyan, N. Castanedo, F.J. Prado-Prado, X. Garcia-Mera, H. Gonzalez-Diaz, ANN multiplexing model of drugs effect on macrophages; theoretical and flow cytometry study on the cytotoxicity of the anti-microbial drug G1 in spleen, Bioorganic & Medicinal Chemistry, 20 (2012) 6181-6194.

[67] H. Gonzalez-Diaz, S. Arrasate, N. Sotomayor, E. Lete, C.R. Munteanu, A. Pazos, L. Besada-Porto, J.M. Ruso, MIANN Models in Medicinal, Physical and Organic Chemistry, Curr Top Med Chem, 13 (2013) 619-641.

[68] K.Y. Sanbonmatsu, C.S. Tung, High performance computing in biology: multimillion atom simulations of nanoscale systems, J Struct Biol, 157 (2007) 470-480.

[69] J.W. Pitera, Current developments in and importance of high-performance computing in drug discovery, Curr Opin Drug Discov Devel, 12 (2009) 388-396.

[70] T.A. Maniatis, K.S. Nikita, N.K. Uzunoglu, Ultrasonic diffraction tomography: an application connecting high performance computing centers with clinical environment, Stud Health Technol Inform, 79 (2000) 214-243.

[71] W.E. Johnston, V.L. Jacobson, S.C. Loken, D.W. Robertson, B.L. Tierney, High-performance computing, high-speed networks, and configurable computing environments: progress toward fully distributed computing, Crit Rev Biomed Eng, 20 (1992) 315-354.

[72] J.J. Fernandez, High performance computing in structural determination by electron cryomicroscopy, J Struct Biol, 164 (2008) 1-6.

[73] T.H. Dunning, Jr., R.J. Harrison, D. Feller, S.S. Xantheas, Promise and challenge of high-performance computing, with examples from molecular modelling, Philos Trans A Math Phys Eng Sci, 360 (2002) 1079-1105.

[74] S. Cant, High-performance computing in computational fluid dynamics: progress and challenges, Philos Trans A Math Phys Eng Sci, 360 (2002) 1211-1225.

[75] J.H. Fowler, S. Jeon, The authority of Supreme Court precedent, Social Networks, 30 (2008) 16-30.

[76] StatSoft.Inc., STATISTICA (data analysis software system), version 6.0, www.statsoft.com.Statsoft, Inc., in, 2002.

[77] T. Hill, P. Lewicki, STATISTICS Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining, StatSoft, Tulsa, 2006

[78] K.E. Stephan, L. Kamper, A. Bozkurt, G.A. Burns, M.P. Young, R. Kotter, Advanced database methodology for the Collation of Connectivity data on the Macaque brain (CoCoMac), Philos. Trans. R. Soc. Lond. B. Biol. Sci., 356 (2001) 1159-1186.

[79] R. Kotter, Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database, Neuroinformatics, 2 (2004) 127-144.

[80] S. Wasserman, K. Faust, Social network analysis: methods and applications, Cambridge University Press, Cambridge, 1999.

[81] A. Duardo-Sánchez, Study of criminal law networks with Markov-probability centralities, in: H. González-Díaz (Ed.) Topological Indices for Medicinal Chemistry, Biology, Parasitology, Neurological and Social Networks, Bentham, Kerala, India, 2010, pp. 205-212.

[82] A. Duardo-Sánchez, Criminal law networks, markov chains, Shannon entropy and artificial neural networks, in: H. González-Díaz (Ed.) Complex Network Entropy: From Molecules to Biology, Parasitology, Technology, Social, Legal, and Neurosciences, Bentham, Kerala, India, 2011, pp. 107-114.

SciForum
Mol2Net

Chemoinformatics Profiling of Ionic Liquids Cytotoxicity—From Machine Learning to Network-Like Similarity Graphs †

Maykel Cruz-Monteagudo [1,2,*], Eduardo Tejera [2], Cesar Paz-y-Miño [2], Yunierkis Perez-Castillo [3,4], Aminael Sánchez-Rodríguez [5], Fernanda Borges [1] and M. Natália D. S. Cordeiro [6]

[1] CIQUP/Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Porto 4169-007, Portugal; E-Mail: fborges@fc.up.pt (F.B.)

[2] Instituto de Investigaciones Biomédicas (IIB), Universidad de Las Américas, 170513 Quito, Ecuador; E-Mails: eduardo.tejera@udla.edu.ec (E.T.); cesar.pazymino@udla.edu.ec (C.P.-Y.-M.)

[3] Sección Físico Química y Matemáticas, Departamento de Química, Universidad Técnica Particular de Loja, San Cayetano Alto S/N, EC1101608 Loja, Ecuador; E-Mail: yunierkis@gmail.com

[4] Molecular Simulation and Drug Design Group, Centro de Bioactivos Químicos (CBQ), Central University of Las Villas, Santa Clara, 54830, Cuba

[5] Departamento de Ciencias Naturales, Universidad Técnica Particular de Loja, Calle París S/N, EC1101608 Loja, Ecuador; E-Mail: asanchez2@utpl.edu.ec

[6] REQUIMTE, Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal; E-Mail: ncordeir@fc.up.pt

* Author to whom correspondence should be addressed; E-Mail: gmailkelcm@yahoo.es;
Tel.: +351-220-402-000; Fax: +351-220-402-009.

Received: 20 September 2015 / Accepted: 20 September 2015 / Published: 2 December 2015

Abstract: Ionic liquids (ILs) possess a unique physicochemical profile providing a wide range of applications. However, their "greenness", specifically their claimed relative non toxicity has been frequently questioned, hindering their REACH registration processes and so, their final application. In this work we introduce a reliable, predictive, simple and chemically interpretable classification and regression tree (CART) classifier enabling the prioritization of ILs with a favourable cytotoxicity profile. By inspecting the structure of the CART several moieties that can be regarded as "cytotoxicophores" were identified and used to establish a set of SAR trends specifically aimed to prioritise low cytotoxicity ILs. We also demonstrated the suitability of the joint use of the CART classifier and a group fusion similarity search as a virtual screening strategy for the automatic prioritisation of safe ILs disperse in a data set of ILs of moderate to very high cytotoxicity. Additionally, we decided to complement the quantitative results already obtained by applying the network-like similarity graphs (NSG) approach to the mining of relevant structure-cytotoxicity relationships (SCR) trends. Finally, the SCR information concurrently gathered by both, quantitative (CART classifier) and qualitative (NSG) approaches was used to design a focused combinatorial library enriched with potentially safe ILs.

Mol2Net YouTube channel*: http://bit.do/mol2net-tube*

[†] The full content of this communication can be found in Cruz-Monteagudo M, Ancede-Gallardo E, Jorge M, Dias Soeiro Cordeiro MN. Toxicol Sci. 2013; 136(2):548-65 and Cruz-Monteagudo M, Cordeiro MN. Toxicol Sci. 2014; 138(1):191-204.

1. Introduction

Ionic liquids (ILs) constitute one of the hottest areas in chemistry since they have become increasingly popular as reaction and extraction media [1]. Their almost limitless structural possibilities, as opposed to limited structural variations within molecular solvents, make ILs ‘‘designer solvents’’ [2]. They have also been widely promoted as “green solvents” [3] but such a “greenness” has been frequently questioned [1].

Despite the scarcity of reports of the prediction of the cytotoxicity of ILs by using a classification approach [4], we consider that a computational prediction system based on the use of classification methods is well justified and can offer a practical tool for the identification of new and safe ILs. So, in this work we intend to introduce a computational system allowing a fully automatic and chemically interpretable IPC-81 cytotoxicity profiling of ILs. In addition to test the predictive capabilities of the system, its potential as the core of a virtual screening (VS) strategy directed to prioritise safe ILs will be demonstrated. Additionally, we complemented these results by applying the NSG approach to the mining of SAR trends relevant for the cytotoxicity of ILs, namely, structure-cytotoxicity relationships (SCR) trends which can be used as useful tips guiding the molecular design of new and safe ILs. Finally, the SCR information concurrently gathered by both, quantitative (CART classifier) and qualitative (NSG) approaches was used to design a focused combinatorial library enriched with potentially safe ILs.

2. Results and Discussion

Cytotoxicity CART Classifier. The main goal of this work is to derive a reliable tool for the automatic prioritisation of safe (low cytotoxicity) ILs. The decision tree corresponding to the simplest best performing CART classifier found is shown in Figure 1.

In general terms, the classifier exhibits a good classification performance. The levels of accuracy (ILs correctly classified), sensitivity (Class_1 ILs correctly classified) and specificity (Class_0 ILs correctly classified) achieved by the CART were around 86%, evidencing the discrimination power and statistical significance of the pattern found. See details in Table 1.

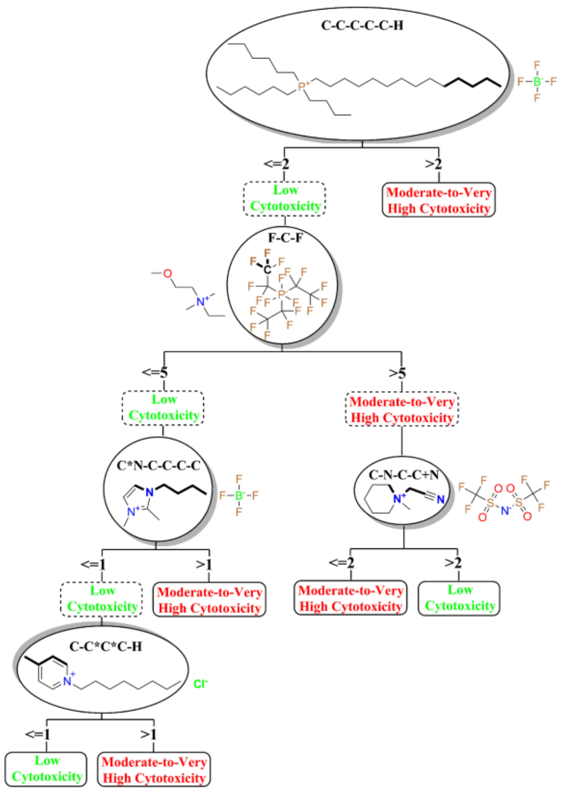

C-C-C-C-C-H
F-B-F
<=2
>2
Low Cytotoxicity
Moderate-to-Very High Cytotoxicity
F-C-F
<=5
>5
Low Cytotoxicity
Moderate-to-Very High Cytotoxicity
C*N-C-C-C-C
C-N-C-C#N
<=1
>1
Low Cytotoxicity
Moderate-to-Very High Cytotoxicity
<=2
>2
Moderate-to-Very High Cytotoxicity
Low Cytotoxicity
C-C*C*C-H
Cl⁻
<=1
>1
Low Cytotoxicity
Moderate-to-Very High Cytotoxicity

Figure 1. Chemically interpretable decision tree corresponding to the best preforming CART classifier found.

Table 1. Classification matrix and classification performance metrics of the CART classifier for the training, test and external evaluation sets.

		TRAINING SET		TEST SET		EXTERNAL EVALUATION SET	
		Observed					
		0	*1*	*0*	*1*	*0*	*1*
Predicted	*0*	**125**	8	**20**	2	**25**	3
	1	21	**51**	4	**8**	5	**9**
Acc. (%)		85.85		82.35		80.95	
Se. (%)		86.44		80.00		75.00	
Sp. (%)		85.62		83.33		83.33	
FN (%)		13.56		20.00		25.00	
FP (%)		14.38		16.67		16.67	
MCC (%)		68.34		60.39		55.90	
$F_{\text{Class 1}}$ (%)		77.86		72.73		69.23	
$F_{\text{Class 0}}$ (%)		89.6		86.96		86.21	

Acc.: Accuracy; *Se.*: Sensitivity or true positives (*TP*) rate; *Sp.*: Specificity or true negatives (*TN*) rate; *FN*: False negatives (*FN*) rate; *FP*: False positives (*FP*) rate; *MCC*: Matthews correlation coefficient; $F_{Class\ 1}$: F-measure for Class

Cytotoxicophores Identification. The influence of the SMFs must be interpreted as a function of the level occupied by the respective SMFs in the decision tree (the influence of the SMF decreases from the base to the leaf of the tree). So, considering the structure of the decision tree depicted in Figure 1 and the structural information of the SMFs it is possible to identify several moieties on ILs inducing a moderate-to-very high cytotoxicity that can be regarded as "cytotoxicophores". According to this analysis, in order of influence, the cytotoxicophores identified are:

- Cationic linear alkyl side chain of length > 5.
- Anions with highly fluorinated alkyl side chains (a fluorocarbonated side chain of length ≥ 2 or two or more trifluoromethyl groups).
- Cationic aromatic N-heterocycles with linear alkyl side chain of length ≥ 4.
- Six membered aromatic rings with a methyl substituent, which can be either the cation head group or its substituent.

Only one moiety was found to have a positive influence on the cytotoxicity profile of ILs, reducing their cytotoxicity from moderate-to-very high to low:

- Short alkyl side chains functionalized with polar nitrile groups on (essentially although not restricted to) aliphatic cation head groups containing nitrogen atoms.

It is important to highlight that the five SMFs identified can also be directly used as cytotoxicophores suitable for automatic procedures of ILs prioritisation such as expert systems, in addition to the cytotoxicity CART classifier proposed in this work.

Joint Use of CART Classifiers and Group Fusion Similarity Searches for the Automatic Prioritization of Safe ILs. The use of the cytotoxicity CART as a virtual screening tool could provide a practical solution to the automatic prioritisation of safe (poorly cytotoxic) ILs.

First, a group fusion similarity search (GFSS) approach [5] was applied. The set of reference structures consist of 20 structurally diverse ILs of lowest cytotoxicity, specially focused on the anion species. The degree of structural proximity by the corresponding values of 1 – the normalized Euclidean distance (1−ED). Finally, the set of 1−ED values between each reference IL and each database IL is combined into a fused similarity score (ε) by averaging the 20 corresponding 1−ED values. In this way, ε "captures" the structural patterns determining ILs of low cytotoxicity and thus can be used independently as a ranking criterion in a GFSS task. However, ε was derived to modify $PP_{\text{Class\_1}}$ and attain the variability required for library ranking. So, the result of using ε as a weighing factor of $PP_{\text{Class\_1}}$ is a new scoring metric that quantifies the likelihood of an IL to exhibit a favourable cytotoxicity profile based on probabilistic ($PP_{\text{Class\_1}}$) and structural similarity (ε) criteria. This new scoring metric will be denoted from now on as Π and it is defined as the geometric mean of $PP_{\text{Class\_1}}$ and ε ($\Pi = \sqrt{PP_{\text{Class\_1}} \times \varepsilon}$).

So, decided to simulate an experiment to evaluate the ability of the approach to retrieve just those 12 ILs of low cytotoxicity (Class 1) of the external evaluation set dispersed in the full set of 200 ILs of moderate-to-very high cytotoxicity (Class 0). For comparison purposes we decided to estimate also the enrichment ability of the independent use of the GFSS approach by using as ranking criterion the fused similarity score ε.

The respective values of *AUAC* and *ROC* metrics obtained from the application of the CART-GFSS approach suggest that it is able to rank a safe IL earlier than an IL of moderate-to-very high cytotoxicity with a probability > 0.85. Instead, the values of these metrics obtained for the GFSS approach show a still good overall enrichment performance (*ROC* = 0. 78), but inferior to the CART-GFSS approach by about 8%.

Table 2. Classic and early recognition enrichment metrics computed to evaluate the enrichment performance of the CART-GFSS and GFSS approaches, respectively.[a]

Metric	*CART-GFSS*	*GFSS*
Classic Enrichment Metrics		
AUAC	0.8557(±0.0014)	0.7775(±0.0013)
ROC	0.8771(±0.0015)	0.7942(±0.0014)
$EF_{1\%}$	11.7778(±0.3200)	11.7778(±0.3200)
$EF_{5\%}$	6.4242(±0.0766)	4.8182(±0.0574)
$EF_{10\%}$	5.6212(±0.0449)	3.2121(±0.0257)
$EF_{20\%}$	3.6977(±0.0197)	3.2868(±0.0175)
Early Recognition Metrics		
$RIE_{1\%}$	6.9142(±0.1692)	6.9116(±0.1691)
$RIE_{5\%}$	6.5692(±0.0717)	5.8978(±0.0643)
$RIE_{10\%}$	5.3265(±0.0397)	4.3204(±0.0322)
$RIE_{20\%}$	3.9049(±0.0190)	3.1222(±0.0152)
$BEDROC_{1\%}$	0.3914(±0.0234)	0.3913(±0.0234)
$BEDROC_{5\%}$	0.4435(±0.0053)	0.3982(±0.0047)
$BEDROC_{10\%}$	0.5042(±0.0028)	0.4089(±0.0023)
$BEDROC_{20\%}$	0.6065(±0.0014)	0.4849(±0.0012)

a: The relative error associated to each enrichment metric is reported. *AUAC*: area under the accumulation curve; *ROC*: area under the ROC curve; $EF_{1\%/5\%/10\%/20\%}$: enrichment factor at χ = 1%/5%/10%/20%, respectively; $RIE_{1\%/5\%/10\%/20\%}$: robust initial enhancement at χ = 1%/5%/10%/20%, respectively; $BEDROC_{1\%/5\%/10\%/20\%}$: Boltzmann-enhanced discrimination of ROC at χ = 1%/5%/10%/20%, respectively.

The analysis of *RIE* at the respective top 1%, 5%, 10% and 20% fractions also points to an attractive early recognition ability of both approaches, consistently favouring the CART-GFSS approach. This pattern is also observed when the metric analysed is *BEDROC*. See details in Table 2.

Network-like Similarity Graph SAR Mining. The analysis was directed to detect in the ILs NSG highly discontinuous regions (clusters of ILs) encoding minimal structural variations leading to significant cytotoxicity changes, with a special interest on those containing cytotoxicity cliffs. Figure 2 shows the NSG obtained at 95% similarity threshold.

This network is characterised by the coexistence of regions with continuous and discontinuous SARs. Regions of continuous SAR are characterized by clusters of small green nodes whereas discontinuous regions involve clusters composed of large green and red nodes (highlighted with a circle around). Clusters of ILs combining large red and green nodes connected by an edge are cytotoxicity cliff markers that can be easily identified. These types of cluster were visually inspected in order to identify the key structure-cytotoxicity relationship (SCR) trends dominating this ILs network.

The most significant cytotoxicity cliff pair in this network (see Figure 2) it is constituted by *N-Methyl-N,N-dioctyl-1-octanaminium bis(trifluoromethylsulfonyl)imide* and *N-Ethyl-N,N-dimethyl-1-butanaminium bis(trifluoromethylsulfonyl)imide* (nodes 1 and 2 in the network) which is a clear example of the influence of the alkyl side chain length over the cytotoxicity of ILs [6-10].

The cluster including the ILs represented by nodes 3, 4 and 5 clearly suggest the effect of highly fluorinated anions over cytotoxicity. A quite explicit correlation between the degree of fluorination of the anion and cytotoxicity it is observed, as reported in previous studies [11].

The last cluster analysed (nodes 6, 7, 8 and 9) suggest a weak influence of the cation head group over cytotoxicity. However, another previous finding can be confirmed in this cluster: the relatively higher cytotoxicity of aromatic cation head groups [12, 13].

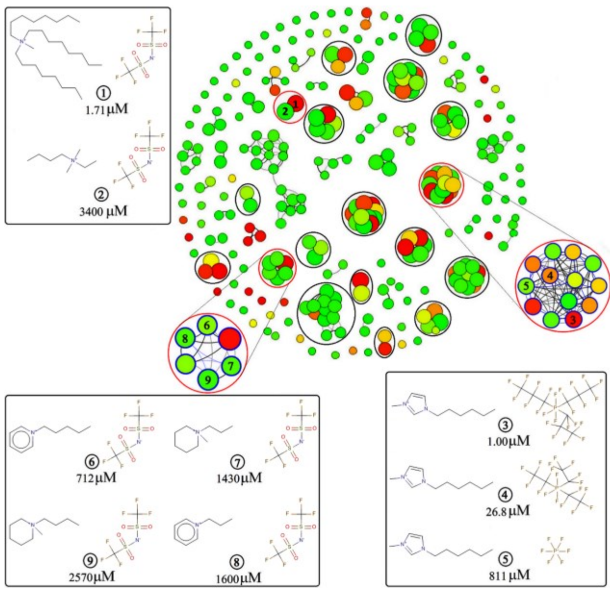


Figure 2. NSG constructed with the software SARANEA for a set of 281 ionic liquids using a Tanimoto similarity threshold of 0.95. The molecular structure and the respective EC50 values for IPC-81 leukaemia rat cell lines of the nine ionic liquids conforming the three clusters analysed are highlighted in three respective square boxes. These three clusters are highlighted in the NSG by red circles while the rest of clusters including ILs inducing a high discontinuity (connected large red and green nodes) are highlighted with black circles and further subjected to SAR pathway analysis.

Design and Assembling of a Focused Combinatorial Library Enriched with Potentially Safe ILs. Finally, the SCR trends identified were used to assemble a focused combinatorial library enriched with potentially safe ILs. The final result is a focused combinatorial library of 697748 ILs. We estimate the quality of the library assembled based on the use of a combined scoring metric (Π) that quantifies the likelihood of a IL to exhibit a favourable cytotoxicity profile. The values of Π near to 1 will be obtained for ILs with a high probability of exhibiting a favourable cytotoxicity profile. The analysis of the combinatorial library revealed that 75.57% of the ILs in the library exhibited values of Π ≥ 0.8, while just 17.72%

exhibited values of Π < 0.5. The mean value of Π obtained for the library was of 0.67. Considering these values one can expect that an IL randomly selected from the library assembled will have a probability of exhibiting a favourable cytotoxicity profile around 67%.

3. Materials and Methods

Data Collection. The UFT/Merck IL DB reports the half cytotoxic concentration (EC_{50}) values (expressed in micromolar units) towards the rat leukemia cell line IPC-81 for 309 ILs and related salts.

Structure Codification. The structural codification was conducted by using the approach proposed by Prof. Varnek´s group and depicted in [14].

Design of the Experiment. The dataset of 281 ILs was subdivided by applying an EC_{50} threshold of 5000 µM into 81 safe or low toxicity ILs (Class_1) and 200 ILs with moderate-to-very high toxicity (Class_0). Once the classes were assigned, we proceeded to split the dataset into three subsets: training, test and external evaluation sets, as part of the model validation scheme [15].

Feature selection, Modelling and Validation. The full vector of ISIDA SMFs was reduced by means of the mRMR software [16] to a minimally redundant vector of size 50 composed of 9/41 anion/cation SMFs. Once this subset was found, the definitive subset of features, and consequently the final classification model, was directly determined by using the Classification and Regression Trees (CART) approach implemented on the *Data Mining* module of STATISTICA 8.0. Both the learning and predictive ability of the classification tree model were assessed by checking overall and class-specific performance measures on training, test and external evaluation sets, respectively [17].

Enrichment Analysis. The main goal in a virtual screening effort is to select a subset from a large pool of compounds (typically a compound database or a virtual library) and try to maximise the number of known actives in this subset. That is, to select the most "enriched" subset as possible. Several enrichment metrics have been proposed in the literature to measure the enrichment ability of a VS protocol [18]. In this work, we use some of the most extended metrics.

Network-like Similarity Graphs Analysis. For this task we resort to SARANEA [9], a freely available program that implements a graphical user interface to NSGs and NSG-based data mining techniques. In SARANEA, as a criterion for edges between nodes in NSGs, connected ILs needed to exceed a predefined Tanimoto similarity threshold value. To search for highly discontinuous regions in the network containing "cytotoxicity cliffs" pairs encoding critical structure variations for cytotoxicity we used a Tanimoto similarity threshold of 0.95.

Combinatorial Library Generation. The assembling of the focused combinatorial library was based on three sets of 15 cationic head groups, 20 cationic side chains and 31 anions previously identified as favouring the cytotoxicity behaviour of ILs. A combinatorial library of 22508 unique cations was generated with the aid of the SmiLib software [19] by using as inputs the corresponding SMILES notation of the two sets of head groups and side chains. The SmiLib software generated an SDF file comprising 22508 unique cations. The SDF file comprising the 31 anions was generated by using the ChemAxon´s JChem for Excel software [20]. Both, cation's and anion's SDF files were submitted to the ISIDA Fragmentor software [21] to compute the corresponding 371/2136 SMFs used to establish the structural reference space for the similarity assessment of the initial set of 281 ILs. Finally, the corresponding SVM output files provided by the ISIDA Fragmentor were converted to a fixed format/length vector file and concatenated into a unique vector file of size 2507 (including the corresponding vector files of 371/2136 anion/cation SMFs) for each one of the 697748 ILs of the combinatorial library. The similarity assessment and the corresponding basic statistical analysis of the combinatorial library were conducted by using a MatLab implementation developed in our group.

Conclusions

In this work we have derived a reliable, predictive, simple and chemically interpretable CART classifier enabling the prioritisation of ILs with a favourable cytotoxicity profile. The analysis of the structure of the corresponding decision tree allowed us to identify several moieties that can be regarded as "cytotoxicophores. We also demonstrated the suitability of the joint use of the CART classifier and a group fusion similarity search (the CART-GFSS approach) as a virtual screening strategy for the automatic prioritisation of safe On the other hand, the NSG approach and NSG-based data mining techniques implemented on SARANEA have proved to be an efficient tool to mine relevant SCR information guiding the design of potentially safe ILs. The adaptation of the NSG approach proposed here to the particular and special case of disconnected molecular structures such as ILs also contributes to the integration of approaches like the traditional T-SAR analysis and the computational mining and visualisation of relevant SCRs of this interesting family of chemicals. Finally, the SCR information gathered from both quantitative (CART classifier) and qualitative (NSG) approaches guided the design of a focused combinatorial library of about 700000 ILs with a likelihood to exhibit a favourable cytotoxicity profile of about 80%. Such a virtual library represents a valuable decision making element for the development of ILs for various technical applications that fulfil the principles of green chemistry.

Acknowledgments

This work is supported by projects financed by FEDER through COMPETE – Programa Operacional Factores de Competitividade and by FCT – Fundação para a Ciência e a Tecnologia [PTDC/EQU-FTT/104195/2008, Pest-C/EQB/LA0006/2011, Pest-C/EQB/LA0020/2011]. M.C-M acknowledges also FCT for the grant [SFRH/BPD/90673/2012] co-financed by the European Social Fund.

Author Contributions

All the authors contributed equally.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Ranke, J.; Stolte, S.; Stormann, R.; Arning, J.; Jastorff, B., Design of sustainable chemical products--the example of ionic liquids. *Chem. Rev.* **2007**, 107, (6), 2183-206.
2. Sheldon, R. A., Green solvents for sustainable organic synthesis: state of the art. *Green Chem.* **2005,** 7, (5), 267-278.
3. Rogers, R. D.; Seddon, K. R., *Ionic Liquids As Green Solvents: Progress and Prospects*. American Chemical Society: 2003; Vol. 856, p 620.
4. Alvarez-Guerra, M.; Irabien, A., Design of ionic liquids: an ecotoxicity (Vibrio fischeri) discrimination approach. *Green Chem.* **2011,** 13, (6), 1507.
5. Willett, P., Similarity-based virtual screening using 2D fingerprints. *Drug Discov. Today* **2006,** 11, (23-24), 1046-53.
6. Stumpfe, D.; Bajorath, J., Methods for SAR visualization. *RSC Advances* **2012,** 2, 369-378.

7. Wawer, M.; Bajorath, J., Extracting SAR Information from a Large Collection of Anti-Malarial Screening Hits by NSG-SPT Analysis. *ACS Med. Chem. Lett.* **2011,** 2, 201-206.

8. Wawer, M.; Lounkine, E.; Wassermann, A. M.; Bajorath, J., Data structures and computational tools for the extraction of SAR information from large compound sets. *Drug Discov. Today* **2010,** 15, (15/16), 630-639.

9. Lounkine, E.; Wawer, M.; Wassermann, A. M.; Bajorath, J., SARANEA: a freely available program to mine structure–activity and structure–selectivity relationship information in compound data sets. *J. Chem. Inf. Model.* **2009,** 50, 68-78.

10. Wawer, M.; Peltason, L.; Weskamp, N.; Teckentrup, A.; Bajorath, J., Structure-activity relationship anatomy by network-like similarity graphs and local structure-activity relationship indices. *J. Med. Chem.* **2008,** 51, 6075-6084.

11. Stolte, S.; Arning, J.; Bottin-Weber, U.; Matzke, M.; Stock, F.; Thiele, K.; Uerdingen, M.; Welz-Biermann, U.; Jastorff, B.; Ranke, J., Anion effects on the cytotoxicity of ionic liquids. *Green Chem.* **2006,** 8, (7), 621.

12. Ranke, J.; Muller, A.; Bottin-Weber, U.; Stock, F.; Stolte, S.; Arning, J.; Stormann, R.; Jastorff, B., Lipophilicity parameters for ionic liquid cations and their correlation to in vitro cytotoxicity. *Ecotoxicol. Environ. Saf.* **2007,** 67, (3), 430-8.

13. Stolte, S.; Arning, J.; Bottin-Weber, U.; Müller, A.; Pitner, W.-R.; Welz-Biermann, U.; Jastorff, B.; Ranke, J., Effects of different head groups and functionalised side chains on the cytotoxicity of ionic liquids. *Green Chem.* **2007,** 9, (7), 760.

14. Billard, I.; Marcou, G.; Ouadi, A.; Varnek, A., In silico design of new ionic liquids based on quantitative structure-property relationship models of ionic liquid viscosity. *J. Phys. Chem. B* **2011,** 115, (1), 93-8.

15. Tropsha, A., Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Inf.* **2010,** 29, 476-488.

16. Peng, H.; Long, F.; Ding, C., Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Trans. Pattern Anal. Machine Intel.* **2005,** 27, (8), 1226-1238.

17. Witten, I. H.; Frank, E., Chapter 5: Credibility: Evaluating what's been learned. In *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed.; Gray, J., Ed. Morgan Kaufman: San Francisco, CA, 2005; pp 143-186.

18. Truchon, J. F.; Bayly, C. I., Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. *J. Chem. Inf. Model.* **2007,** 47, (2), 488-508.

19. Schüller, A.; Schneider, G.; Byvatov, E., SMILIB: Rapid Assembly of Combinatorial Libraries in SMILES Notation. *QSAR & Comb Science* **2003,** 22, 719-721.

20. ChemAxon *JChem for Excel*, version 5.10.2.725; 2012.

21. Varnek, A.; Fourches, D.; Horvath, D.; Klimchuk, O.; Gaudin, C.; Vayer, P.; Solov'ev, V.; Hoonakker, F.; Tetko, I.; G., M., ISIDA - platform for virtual screening based on fragment and pharmacophoric descriptors. *Curr. Comput.-Aided Drug Des.* **2008,** 4, 191-198.

SciForum
Mol2Net

Complex Networks of anti-HIV Drugs Activity *vs.* Prevalence of AIDS in US Counties Using Symmetry Information Indices

Diana María Herrera-Ibatá [1,*] and Ricardo Alfredo Orbegozo-Medina [2]

[1] Department of Information and Communication Technologies, University of A Coruña UDC, 15071 Ferrol, A Coruña, Spain

[2] Department of Microbiology and Parasitology, University of Santiago de Compostela (USC), 15782 Santiago de Compostela, A Coruña, Spain

* Author to whom correspondence should be addressed; E-Mail: dianamariahi@gmail.com.

Received: 6 October 2015 / Accepted: 6 October 2015 / Published: 2 December 2015

Abstract: Different aspects about the epidemiology, drugs, targets, chem-bioinformatics, and systems biology methods, related to AIDS/HIV have been reviewed. Next, we developed a new model to predict complex networks of the AIDS prevalence in U.S. counties taking into consideration the Gini coefficient (income inequality) and activity/structure data of anti-HIV drugs in preclinical assays. First, we trained different Artificial Neural Networks (ANNs) using as input Markov and Symmetry information indices of social networks and of molecular graphs, respectively. We obtained the data about AIDS prevalence and Gini coefficient from the AIDSVu database of the Rollins School of Public Health at Emory University and the data about anti-HIV compounds from ChEMBL database. To train/validate the model and predict the complex network we needed to analyze 43,249 data points including values of AIDS prevalence in 2310 US counties *vs.* ChEMBL results for 21,582 unique drugs, 9 viral or human protein targets, 4856 protocols, and 10 possible experimental measures. The best model found was a Linear Neural Network (LNN) with Accuracy, Specificity, Sensitivity, and AUROC above 0.72-0.73 in training and external validation series. The new linear equation was shown to be useful to generate complex network maps of drug activity *vs.* AIDS/HIV epidemiology in U.S. at county level.

Keywords: anti-HIV drugs, Gini coefficient, neighborhood symmetry indices; complex networks.

Mol2Net YouTube channel*:* *http://bit.do/mol2net-tube*

1. Introduction

Human immunodeficiency virus (HIV) is a retrovirus belonging to the family of lentiviruses that causes AIDS. Retroviruses[1] can use their RNA and host DNA to make viral DNA, and are known for their long incubation periods. There are two types of HIV: HIV type 1 and HIV type 2. Despite progresses, HIV[2] remains a public health challenge. After thirty years in the AIDS epidemic, there are over 34 million people living with HIV[3], and still 2.5 million new infections and 1.7 million deaths each year.

A useful chemoinformatics-pharmacoepidemiology model must be multi-level to account molecular and population structure. We need to process diverse types of input data. Initially, we need the information about the anti-HIV drugs, such as chemical structure of the drug (level i) and preclinical information, like biological targets (level ii), organisms (level iii), or assay protocols (level iv). Afterwards, we need to incorporate population structure descriptors (level v) that quantify the epidemiological and socioeconomic factors affecting the population selected for the study.

2. Results and Discussion

After analysis of the previous results, we decided to test the predictive power of these indices in a simpler model using the STATISTICA 6.0[4] software. In so doing, we trained the LNN predictors using only each family of information indices of drugs ($^qIC_{5f}$) of 5- order, their MA operators (Δ^qIC_{5fj})) and the fifth MA operator of the U.S. counties (ΔI^a5s). The LNN model based on qIC_{51} (LNN-IC_{51}) presented the higher values of Sn = 72.04/72.81 and Sp = 72.38/72.50 in training/ and external validation sets (see Table 1). LNN-IC_{51} presented also the higher values for the AUROC in train and validation series (0.73 and 0.74 respectively). Analyzing all the previous results for this dataset, we found that the IC_k index appears to be the most important to predict the drug structure-activity relationships. We can conclude it by comparison to the other indices, which have lower values of classification. The equation of LNN-IC_{51} this model is the following:

$$
\begin{aligned}
S_{aq}\overset{\cdot}{(}c_j) = {} & -25.48 \cdot {}^q IC_{51} + 1081.64 \cdot \Delta^q IC_{51}(c_1) + 29.36 \cdot \Delta^q IC_{51}(c_2) \\
& -1084.52 \cdot \Delta^q IC_{51}(c_3) - 0.7727 \cdot \Delta^q IC_{51}(c_4) \\
& -0.0792 \cdot \Delta I^a{}_5(s) - 0.5025
\end{aligned}
$$

Last, we used this LNN-ALMA model to generate/predict a complex network of the prevalence of AIDS in the United States at county level with respect to the preclinical activity of anti-HIV drugs (Figure 1). The bipartite network has two types of nodes (counties vs. drug). Thus, this is a multiscale network similar to bipartite networks of drugs vs. target proteins reported by other groups[5-7]. However, the nodes in the present network contain information about the molecules, i.e., chemical structure as well as assay conditions (target protein, organism, experimental measure, etc.). Additionally, the other set of nodes contain information about socioeconomic factors, such as the income inequality in the county.

Multiscale networks of this type have been discussed by Barabasi et al.[8] as one of the more important tools to perform trans-disciplinary research. The links of this complex network are the outputs Laq(cj)pred = 1 of our model. In Figure 1, we illustrate the sub-network of AIDS prevalence vs. Anti-HIV drug preclinical activity for the state of Florida. For instance, the model predicts a high effectively for the drug Zidovudine to treat AIDS in Nassau County.

Table 1. LNN classifier for symmetry information indices of 5-order

Type of Index	Observed	$L_{pq} = 1$	$L_{pq} = 0$	$L_{pq} = 1$	$L_{pq} = 0$	AUROC
		Training		**Validation**		
$^{q}IC_{51}$	Parameter [a]	Sn	Sp	Sn	Sp	(T / V)
	Predicted	72.04	72.38	72.81	72.50	0.73 / 0.73
	$L_{pq} = 1$	8255	5746	2775	1908	
	$L_{pq} = 0$	3203	15060	1036	5031	

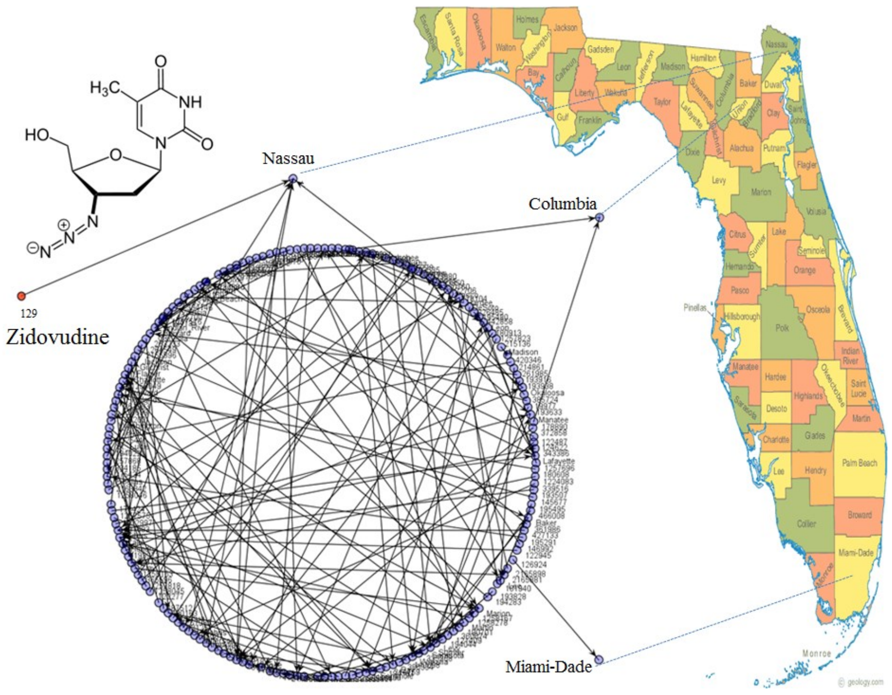


Figure 1. Sub-network of AIDS prevalence *vs.* Anti-HIV drug activity for U.S. state of Florida (FL)

3. Materials and Methods

In the present paper, we changed the Balaban information indices ($I^{q}k$) by Symmetry information content indices ($^{q}IC_{kf}$)[9]. These indices are calculated for H-included molecular graph and based on neighbor degrees and edge multiplicity.[10, 11] The symmetry information indices are calculated by partitioning graph vertices into equivalence classes; the topological equivalence of two vertices is that the corresponding neighborhoods of the k*th* order are the same. However, we used the $I^a k(s)$ indices to characterize the different populations. We used the software DRAGON[12] to calculate the ${}^qIC_{kf}$ indices for the molecules of the ChEMBL dataset of anti-HIV drugs. In this case we calculated a total of $N_{indices}$ = Nk•Nf = 6*5 = 30 values of ${}^qIC_{kf}$ indices with Nk = 6 different orders (k) that belong to Nf = 5 different families of descriptors (f). We have used Markov chains to calculate Shannon information indices of different systems including simulations of disease spreading relevant to epidemiology. [13]

The codification of the chemical structure of the compounds is the first step here. We have data about a large number of assays developed in very different conditions (c_j) for equal or different targets (molecular or not). The non-structural information here refers to different assay conditions (c_j) like concentrations, temperature, targets, organisms, *etc.* A solution may rely upon the use of the idea of Moving Average (MA) operators used in time series analysis with a similar purpose. We have developed a similar approach called ALMA (Assessing Links with Moving Averages) using also MA operators. ALMA models remember those used in ARIMA models of time series analysis[14]. They are adaptable to all molecular descriptors and/or graphs invariants or descriptors for complex networks. In consonance with the previous section, we use a similar terminology. The inputs of one ALMA model are the descriptors $D^q{}_k$ of type kth of the qth system (compound or drug d_q in this case) represented by a matrix M. On the other hand, the outputs of one ALMA model are the links (Laq = 1 or Laq = 0) of a complex network with Boolean matrix L and formed by different pairs of input systems. We developed different ANN models using all the set of parameters as well as simple models using different sub-sets of descriptors. The new ALMA model developed using these other set of indices has the following general form:

$$\begin{aligned}
S_{aqj} = {} & \sum_{k=0}^{k=5}\sum_{f=1}^{f=5} e_{kf} \cdot {}^qIC_{kf} \\
& + \sum_{k=0}^{k=5}\sum_{f=1}^{f=5}\sum_{j=1}^{j=4} e_{kfj} \cdot \Delta^q IC_{kfj} \\
& + \sum_{k=1}^{k=5} e_{ak} \cdot \Delta I^a{}_{ks} + e_0 \\
= {} & \sum_{k=0}^{k=5}\sum_{f=1}^{f=5} e_{kf} \cdot {}^qIC_{kf} \\
& + \sum_{k=0}^{k=5}\sum_{f=1}^{f=5}\sum_{j=1}^{j=4} e_{kfj} \cdot \left({}^qIC_{kf} - \left\langle {}^qIC_{kf} \right\rangle_j\right) \\
& + \sum_{k=1}^{k=5} e_k \cdot \left(I^a{}_k - \left\langle I^a{}_k \right\rangle_s\right) + e_0
\end{aligned}$$

4. Conclusions

This work presents a review of several aspects of the disease, including the epidemiology, pathophysiology, treatments, etc. We also developed a model called LNN-ALMA to generate complex networks of the prevalence of AIDS in the counties of the U.S. with respect to the preclinical activity of anti-HIV drugs. The best classifier found was the LNN-$IC_{51;}$ this classifier has only six inputs based on neighborhood information content indices, compared to the other models, the IC_k index seems to be the most important to predict the drug structure-activity relationships. The new model has similar performance but is notably simpler than a previous model based on Balaban's information indices with >20 inputs.

Acknowledgments

R.O.M acknowledges financial support of FPI fellowship associated to research project (AGL2011-30563-C03-01) funded by MECD (Spanish Ministry of Education, Culture and Sport).

Conflicts of Interest

"The authors declare no conflict of interest".

References and Notes

1. Lindemann, D.; Steffen, I.; Pohlmann, S., Cellular entry of retroviruses. *Advances in experimental medicine and biology* **2013**, 790, 128-49.
2. Moss, J. A., HIV/AIDS Review. *Radiologic technology* **2013**, 84, 247-67; quiz p.268-70.
3. Piot, P.; Quinn, T. C., Response to the AIDS pandemic--a global health model. *The New England journal of medicine* **2013**, 368, 2210-8.
4. *STATISTICA*, version 6.0; StatSoft Inc.: Tulsa, Oklahoma, 2001.
5. Prado-Prado, F.; Garcia-Mera, X.; Escobar, M.; Alonso, N.; Caamano, O.; Yanez, M.; Gonzalez-Diaz, H., 3D MI-DRAGON: new model for the reconstruction of US FDA drug- target network and theoretical-experimental studies of inhibitors of rasagiline derivatives for AChE. *Current topics in medicinal chemistry* **2012**, 12, 1843-65.
6. Prado-Prado, F.; Garcia-Mera, X.; Abeijon, P.; Alonso, N.; Caamano, O.; Yanez, M.; Garate, T.; Mezo, M.; Gonzalez-Warleta, M.; Muino, L.; Ubeira, F. M.; Gonzalez-Diaz, H., Using entropy of drug and protein graphs to predict FDA drug-target network: theoretic-experimental study of MAO inhibitors and hemoglobin peptides from Fasciola hepatica. *European journal of medicinal chemistry* **2011**, 46, 1074-94.
7. Vina, D.; Uriarte, E.; Orallo, F.; Gonzalez-Diaz, H., Alignment-free prediction of a drug-target complex network based on parameters of drug connectivity and protein sequence of receptors. *Molecular pharmaceutics* **2009**, 6, 825-35.
8. Barabasi, A. L.; Gulbahce, N.; Loscalzo, J., Network medicine: a network-based approach to human disease. *Nature reviews. Genetics* **2011**, 12, 56-68.
9. González-Díaz, H.; Herrera-Ibatá, D. M.; Duardo-Sanchez, A.; Munteanu, C. R.; Orbegozo-Medina, R. A.; Pazos, A., Model of the Multiscale Complex Network of AIDS prevalence in US at county level vs. Preclinical activity of anti-HIV drugs based on information indices of molecular graphs and social networks. *Journal of chemical information and modeling* **2014**, 54, 744-755.
10. Magnuson, V. R.; Harriss, D. K.; Basak, S. C. In *Studies in Physical and Theoretical Chemistry; King, R.B.*; Elsevier: Amsterdam (The Netherlands), 1983, pp 178-191.
11. Todeschini, R.; Consonni, V., *Handbook of Molecular Descriptors*. Wiley-VCH Verlag GmbH: Weinheim, Germany, 2000.
12. Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M., In; Talete srl: Milano, Italy, 2005.
13. Riera-Fernandez, P.; Munteanu, C. R.; Escobar, M.; Prado-Prado, F.; Martin-Romalde, R.; Pereira, D.; Villalba, K.; Duardo-Sanchez, A.; Gonzalez-Diaz, H., New Markov-Shannon Entropy models to assess connectivity quality in complex networks: from molecular to cellular pathway,

Parasite-Host, Neural, Industry, and Legal-Social networks. *Journal of theoretical biology* **2012**, 293, 174-88.

14. Langenfeld, M. C.; Cipani, E.; Borckardt, J. J., Hypnosis for the control of HIV/AIDS-related pain. *The International journal of clinical and experimental hypnosis* **2002**, 50, 170-88.

SciForum
Mol2Net

Kinetic Study of Activated Carbon Synthesis from Marabou Wood

Pedro Julio Villegas
CUBEL Consultancy, 375, Baron Bliss Street, Benque Viejo del Carmen, Cayo District, Belize, Italy.
E-Mail: pjva00@hotmail.com or pjva00@gmail.com; Tel.: +501-669-8812

Received: 28 September 2015 / Accepted: 28 September 2015 / Published: 2 December 2015

Abstract: In the last years the demand of activated carbons for environmental remediation and medical applications has been growing. This situation has stimulated the study of new precursors for the synthesis of these adsorbents. This work shows the kinetic parameters of activation process of Marabou Wood (*Leptoptilus Crumeniferus*) using a simple mathematical model. These parameters were compared with ones corresponding to other tropical biomasses studied under similar conditions. To conduct the study, a thermo-gravimetric analysis was carried out in steam water. The study was carried on from room temperature to 1000°C with a heating rate of 10°C/min, additionally; the crystallinity was determined by X-rays diffraction analysis. The characterization of the activated carbon was carried out through parameters that provide an indirect measure of the mechanical resistance. Interesting correlations for the analyzed thermal conversion processes were also obtained.

Keywords: biomass, activated carbons, thermo-gravimetric study, kinetic parameters, X-ray, mechanical resistance.

1. Introduction

Various studies about the morphological and textural characterization during the thermal conversion of biomasses resources have been reported. These works studied mainly coconut shell and olive stones as raw materials. (5, 6, 13, 16) However, other resources that are nowadays widely available in tropical areas, with less competence in other applications, have been barely studied. (13, 17, 19)

Marabou is an exotic wood that is increasingly infecting cultivated fields of Central American Countries. In the last years, important efforts have been conducted, in order to find different alternatives to reduce the negative environmental impact of these great amounts of biomass that is difficult to manage.

Moreover, the possibility to valorize them could allow the achievement of a sustainable agricultural development. The preparation of activated carbon, an expensive adsorbent highly demanded in the international market, could be an interesting alternative. (2, 4, 8)

In this work, the kinetic parameters of the activation processes of Marabou Wood were determined. These parameters were compared with those obtained for other woods studied under similar conditions. Besides, some useful correlations between the mechanical properties and the kinetic parameters were also inferred.

2. Materials and Methods

Marabou (*Leptoptilus Crumeniferus*) was the biomass precursor studied in this work. Other nine woods were also included with comparative purposes. (19)

2.1. Physical- chemical and activation studies

The physical-chemical study of the activation processes consisted in the determination of some kinetic parameters such as: activation energy, kinetic constant and reaction order. To conduct the study, non-isothermal thermo-gravimetric registers were executed using a Shimatzu-TGA 50 equipment.

The experimental conditions were similar to those used in previous studies with others biomasses what minimize possible diffusive effects, time and money. The final temperature used in this study was 1000°C and the heating rate of 10°C/min. (18)

2.2. Kinetic parameters evaluation

For the evaluation of the kinetic parameters, a simple model known as "*Transient kinetic model in non-stationary state"* was used. This model has been widely used to study chemical reactions between gas and solid products. (9) There are various works that used this model in the kinetic characterization of the heterogeneous catalysis and the carbonaceous materials activation with CO_2, O_2 and/or H_2O. (3, 11, 15)

For these reasons, this model should be adequate to study the kinetic of carbonaceous adsorbents preparation.

This model considers the thermo-chemical reactions of biomasses as processes that occur in a single global stage. This assumption allows the mathematical modeling of the experimental data with a reduced number of parameters using a single following expression:

$$\frac{dX}{dt} = k\,(1-X)^{n} \qquad [1]$$

Where: **X** is the solid conversion: $X = \frac{m_0 - m}{m_0}$ [2]

t, the time; **k**, the kinetic constant of the global reaction of activation and **n** the reaction order with respect to solid. The experimental data **X** *vs.* **t** was modeling by non-lineal regression. The characteristics parameter of the model was estimated by minimizing the objective function OF:

$$OF = \sum_{i=1}^{N}\left(\left.\frac{dX}{dt}\right|_{exp_i} - \left.\frac{dX}{dt}\right|_{cal_i}\right)^2 \qquad [3]$$

Where: **N** is the experimental data amount; $\left.\frac{dX}{dt}\right|_{exp}$ is the experimental reaction rate, obtained from the thermo-gravimetric registry and $\left.\frac{dX}{dt}\right|_{cal}$ is the reaction rate calculated by the model.

The kinetic energy, AE was calculated by Arrhenius equation: $k = k_0 \exp\left\{-\frac{AE}{RT}\right\}$ [4]

Where: **k_0** is the pre-exponential factor; **AE** is the activation energy of the global reaction and **R** the gases universal constant.

2.3. X-Ray Diffraction study

Marabou was characterized by X-ray diffraction analysis using Philip equipment with the following characteristics: radiation of Co (λ=1,78897Å), 40kV, 30mA and 1°2θ*min^{-1}. The particles size was below 62μm obtained by grinding in agate mortar and sieving and then located on a glycerin film.

The measurement of the intensities and positions of the diffracted beams in X-Ray Diffraction (XRD) spectrum, and the use of structure factor equations are necessary in order to determine the atoms distributed in the unitary cells. (1, 20)

In order to determine the crystallinity (η) of the precursors was considered that the energy involved in the diffraction keep constant. Then it can be affirmed that the sum of the dispersed radiation and diffracted one is constant allowing the calculation of the crystallinity through the following expression:

3. Results and Discussion

Non-isothermal thermo-gravimetric registry that characterizes the activation of the raw material is shown in Figure 1. In this Figure **w** is the conversion that can be calculated by: w = (1 - X)*100.

The thermo-gravimetric curve is divided in 3 sections. In section one, up to 220°C it can be clearly observed that the weight loss is minimal. This first loss can be associated with the elimination of absorbed water and the removing of some volatile compounds. In section 2, temperatures higher than 220°C and up to 380°C, a strong weight loss can be observed. This loss is attributed mainly to the pyrolysis o de-volatilization process; in this process a loss of 50% of total mass was registered.

$$\%\eta = \frac{I}{It} * 100 \qquad [5]$$

Where: **I** is the dispersed radiation and **It** is the total intensity of the radiation. In order to evaluate the crystallinity using equation 5, firstly, the area under the curve in the angular range from 10° to 45° was determined. The XRD pattern of the studied precursor presents a wide band in this interval indicating the presence of high amount of non-crystalline substances. (12)

Finally, the mechanical resistance of the activated carbon obtained from Marabou was measured. (10) This simple method used a known mass of the granular material that is impacted by six glass balls into a semispherical container of stainless steel. The perceptual relation between the fragmented mass retained in a 0.5 mm mesh and the initial mass is used to estimate the mechanical resistance of the activated carbon. (7)

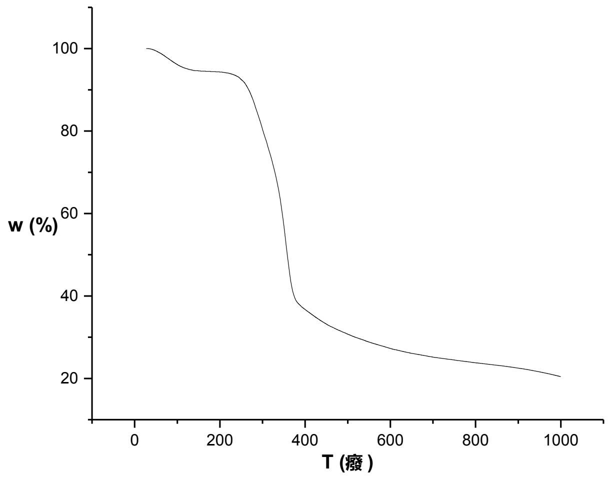


Figure 1. Thermo-gravimetric registry of Marabou.

In section 3, temperature above 380°C the carbon content increased significantly obtained a more porous solid product. This last section is considered, in the majority of revised works, as the activation process.

The carbonized product increases the porosity with the temperature, producing an excellent adsorbent. Furthermore, it is advisable to use low heating rates to avoid undesirable morphological damages during the activation process. It is also necessary to point out that temperature higher than 800°C, significantly affects the yield without notable increase of the micro-porosity. Hence to obtain a better product, it is advisable temperatures near 800°C or below. (6, 18)

Taking into account previous observations derived from Figure 1, the temperature interval from 380°C to 800°C was used to estimate the kinetics parameters of the thermo-chemical reaction studied. The previously defined mathematical model (Equation 1) was used. The results are reported in Table 1.

Table 1. Kinetics parameters of Marabou activation compared with other woods under similar conditions

Sample	AE (kJ/mol)	k(min^{-1})	n	S.D.(%)	V.C.(%)
Iron Wood (Schinosis Balansae)	101.88	9.38*10^6	1.15	0.19	0.25
Holy Wood (Bulnesia Sarmientoi)	96.18	4.82*10^6	1.00	0.20	0.23
Teak (Tectona Grandis)	89.52	9.94*10^6	1.00	0.22	0.26
River Oak (Casuarina Cunninghamiana)	90.52	1.04*10^5	1.00	0.37	0.44
Eucalyptus (Eucalyptus Robusta)	85.93	5.00*10^6	1.07	0.16	0.20
Marabou (Leptoptilus Crumeniferus)	84.42	5.43*10^6	1.00	0.12	0.13
White Carob-tree (Prosopis Alba)	76.60	8.00*10^5	0.97	0.03	0.03
Mahogany (Jacaranda Semiserrata)	72.72	6.11*10^5	0.69	0.09	0.11
Pine (Araucaria Angustifolia)	50.00	1.98*10^5	0.74	0.21	1.04
Cedar (Cedrela Balansae)	49.50	3.89*10^5	0.98	0.05	0.15

(**AE**: activation energy; **k**: kinetic constant; **n**: reaction order; **S.D.**: standard deviation; **V.C.**: variation coefficient).

From Table 1 it can be deduced that the thermal conversion of Marabou has values similar to semi-hard woods. The reaction orders obtained are, in all cases near to one. Should be noted that the higher values correspond to the *Red Quebracho or Iron Wood*, what is attributable to its higher hardness and hence it's lower reactivity. The very low values for standard deviations (< 0,40%) and variation coefficients (≤0,45%) are indicative that the method applied to estimate the kinetic parameters is adequate.

The crystallinity degree (η) of the raw materials was evaluated from the XRD registry (Figure 2) such as the mechanical resistance. This value was compared with other materials in Table 2.

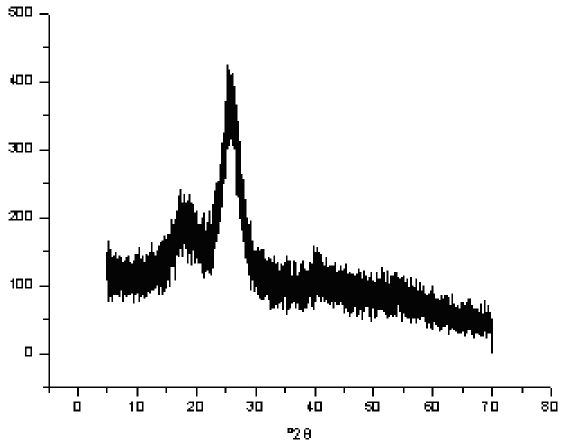


Figure 2. X-Ray Diffraction Registry of Marabou.

Table 2. Crystallinity (η), Mechanical Resistance (Rm) and Density (d) of Marabou and other Materials.

Sample	η (%)	Rm (%)	d (g/cm^3)
Iron Wood (Schinopsis Balansae)	55.36	98.97	1.200
Holy Wood (Bulnesia Sarmientoi)	53.01	96.63	1.150

Teak (Tectona Grandis)	47.32	92.72	1.100
River Oak (Casuarina Cunninghamiana)	45.53	90.15	0.900
Eucalyptus (Eucalyptus Robusta)	42.86	89.00	0.800
Marabou (Leptoptilus Crumeniferus)	39.10	86.63	0.780
White Carob Tree (Prosopis Alba)	41.84	85.76	0.764
Mahogany (Jacaranda Semiserrata)	39.72	81.04	0.850
Pine (Araucaria Angustifolia)	35.49	79.47	0.551
Cedar (Cedrela Balansae)	35.14	77.64	0.483

From Table 2 it should be noted that significant differences were appreciated in the crystallinity degree for the analyzed precursor, compared to the others. Although, in principle, it can be attributed a low reactivity to a greater crystallinity degree, it should be also considered the influence of other factors, for example the chemical composition of the precursor.

The mechanical resistance **(Rm)** of the activated carbon from *Marabou* has an appropriate value. Compared with other materials the *Iron Tree* or *Red Quebracho* has the higher value. From this table it can also be inferred that the less reactive precursors are the harder activated carbons. The differences between the **Rm** values could be due to the different chemical composition of the original products.

The mechanical resistance is very useful to evaluate differences between precursor's reactivity. It is known that the diffusion of weak oxidative reagents, in the structure of precursors of high hardness is hindered; consequently the necessary energy to favor the activation reactions would be higher. Hence, the mechanisms of thermo-chemical conversion differ from one sample to other. During this complex process, different arrangements of the carbon chains take place with the increment of temperature. The thermo-chemical conversion process will be conditioned by the operational conditions used such as the mechanical properties of the materials.

From Table 1 and 2 some correlation between physical-chemical and mechanical parameters can be obtained as show following:

$$Rm = 0.373*AE + 58.104 \ (R^2 = 0.8762) \quad [6]$$

$$d = 0122*AE - 0.1155 \ (R^2 = 0.8353) \quad [7]$$

$$\eta = 0.338*AE + 16.607 \ (R^2 = 0.784) \quad [8]$$

4. Conclusions

The simple model used in the present study, was an appropriate tool for the determination of the kinetic parameters of *Marabou Wood* activation with steam water.

The kinetic parameters that characterize the synthesis of activated carbon from *Marabou Wood* are similar to those reported to semi-hard woods. These values assure the feasibility of this precursor in the production of this adsorbent.

The statistical parameters assure the right adaptation of the experimental results of the Theoretical Model used, what means that this study is adequate and accurate.

Some useful correlation for the analyzed thermal conversion process was also obtained.

The crystallinity degree of the precursor during the thermo-chemical process studied was determined from the X-ray diffraction analysis.

Acknowledgments

The author wishes to acknowledge Maria Andrea Camerucci and Ana Lia Cavalieri from Mar del

Plata University, Argentina, they provided a crucial help in the experiments of this work.

Conflicts of Interest

The author declares no conflict of interest.

References

1. Alexander, L. E. X-ray diffraction methods in polymer science. *Ed. Wiley Interscience. John Wiley & Soc. Inc.* New York-London-Toronto, **1989**, p. 582.
2. Deiana, A. C., Granados, D. L., Petkovic, M. F., Sardela, M. F., Silva, H. Use of grape must as a binder to obtain activated carbon briquettes. *Brazilian Journal of Chemical Engineering,* **2004**, 21, 4, 585–591.
3. Di Blasi, C., Buonano, F., Branca, C. Combustion kinetics of chars derived from agricultural residues. *Proceeding of the Biomass for Energy and Industry, 10th European Conference and Technology Exhibition,* Würzburg, Alemania, **1998.**
4. Elkady, M. F., Hussein, M. M., Salama, M. M. Synthesis and Characterization of Nano-Activated Carbon from El Maghara Coal, Sinai, Egypt to be Utilized for Wastewater Purification. *American Journal of Applied Chemistry*. **2015,** 3, 3, 1-7.
5. González, M. T., Molina Sabio, M. Rodríguez Reinoso, F. Steam activation of olives stone chars, development of porosity. *Carbon*, **1994**, 32, 8, 1407-1413.
6. González, M. T., Rodríguez Reinoso, F., García, A. N., Marcilla, A. Activation of olive stones carbonized under different experimental conditions. *Carbon,* **1997,** 35, 159-162.
7. Heschel, W., Klose, E. On the suitability of agricultural by-product for the manufacture of granular activated carbon. *Fuel*, **1995,** 74, 12, 1787-1791.
8. Jaguaribe, E. F., Medeiros, L. L., Barreto, M. C. S., Araujo, L. P. The performance of activated carbons from sugarcane bagasse and coconut shells in removing residual chlorine. *Brazilian Journal of Chemical Engineering*, **2004,** 22, 1, 41-47.
9. Lizzio, A. A., Jiang, H., Radovic, L. R. On the kinetics of carbon (char) Gasification: reconciling models with experiments. *Carbon*, **1990,** 28, 1, 7-19.
10. Lovera, R. G. Activated Carbons. *Proceeding of III Iberia-American Workshop "Adsorbents for Environmental Protection"*, La Plata, Argentina, **2003,** 79-90
11. Luo, M., Stanmore, B. The combustion characteristics of char from pulverized bagasse. *Fuel*, **1992,** 71, 1074-1076.
12. Magnaterra, M., Fusco, J. R. Ochoa, J., Cukierman, A. L. Kinetic study of the reaction of different hardwood sawdust chars with oxygen, chemical and structural characterization of the samples. *Proceeding of the International Conference on Advances in Thermochemical Biomass Conversion*, Ed. A. V. Bridwater, Blackie A & P, **1994**. 116-130.
13. Olontsev, V. Pyrolysis of Coconut Shells for the Manufacture of Carbon Sorbents. *Solid Fuel Chemistry* **2011,** 45, 1, 47-52.
14. Rashid, K.; Reddy,S. K.; Al Shoaibi, A.; and Srinivasakannan, C. Process optimization of porous carbon preparation from date palm pits and adsorption kinetics of methylene blue. *The Canadian Journal of Chemical Engineering,* **2014,** 92, 426-434.
15. Roberts, D. G., Harris, D. J. Char gasification with O_2, CO_2, and H_2O: effects of pressure on intrinsic reaction kinetics. *Energy & Fuels*, **2000,** 14, 2, 483-489.

16. Satya Sai, P. M., Ahmed, J., Krishnaiah, K. Production of activated carbon from coconut shell char in a fluidized bed reactor. *Industrial & Engineering Chemistry Research*, **1997,** 36, 3625-3630.
17. Shawabkeh, R. A.; Al-Harthi, M. and Al-Ghamdi, S. M. The Synthesis and Characterization of Microporous, High Surface Area Activated Carbon from Palm Seeds. *Energy Sources,* **2014,** 36, 1, 93-103.
18. Villegas Aguilar, P. J. Optimal use of sugar cane mill fibrous wastes by thermal conversion processes. *Doctoral Thesis,* Central University of Las Villas, Santa Clara, Cuba. **2000**.
19. Villegas, P. J.; Camerucci M. A. and Quintana-Puchol R. Kinetic of the Thermal Conversion Processes of Tropical Biomasses. *Handbook on Emerging Trends in Scientific Research*, **2014**, 37-43.
20. Voinshtein, B. K. Diffraction of X ray by chain molecules. Ed. Elsevier Publishing Co. Amsterdam-London- New York, **1986**, p.414.

SciForum
Mol2Net

Computational Study of Mycobacterial Promoters with Low Sequence Homology

Alcides Pérez-Bello

Veterinary Medicine Department, Central University of 'Las Villas', 54830, Cuba. E-Mail: alcidopb@yahoo.com.

Received: 29 September 2015 / Accepted: 29 September 2015 / Published: 2 December 2015

Abstract: This communication shows a classification model for prediction of mycobacterial promoter sequences (mps), which constitute a very low sequence homology problem. The model developed (mps = –4.664·$^{0}\xi_{M}$ + 0.991·$^{1}\xi_{M}$ – 2.432) was intended to predict whether a naturally occurring sequence is an mps or not on the basis of the calculated $^{k}\xi_{M}$ value for the corresponding RNA secondary structure. The model predicted 115/135 mps (85.2%) and 100% of control sequences (cs). The detailed results have been published in detail in: Bioorg Med Chem Lett. 2006 Feb;16(3):547-53, the present is a short communications.

Keywords: Mycobacteria; Protein synthesis promoters; RNA secondary structure

1. Introduction

Harshey and Ramkrishnan stated that *Mycobacteria* have a low transcription rate and a low RNA content per unit DNA and that their genomes are rich in Guanine and Cytosine (g + c) content. Given that the g + c content of a genome affects the codon usage and the promoter recognition sites in an organism, Nakayama *et al.*, and Ohama *et al.* predicted that the transcription and translation signals in *Mycobacteria* may be different from those in other bacteria such as *E. coli*. Therefore, understanding the factors responsible for the low level of transcription and the possible mechanisms of regulation of gene expression in *Mycobacteria* requires examination of the structure of mycobacterial promoter sequences (mps) and their transcription machinery, including information concerning the RNA macromolecules involved. Unfortunately, mps present a very low sequence homology and mathematical methods to assign biological activity based on sequence alignment are not of practical use in this case. Different mathematical methods have been used for the analysis of genome information. The group of Professor Grau has reported results on genome algebras. Markov models are also well-known tools for analyzing biological sequence data. However, advances have not been reported concerning the treatment of this macromolecular structure-activity problem from the point of view of the corresponding RNA structure.

A real possibility to address this problem involves the analysis of structure-activity relationships for naturally occurring RNA macromolecules, synthetic polymers and small molecules in general with Markov molecular descriptors. For this reason, one may expect higher success for classical molecular indices in branched biomacromolecules. However, it must be remembered that the more commonly known branched biomacromolecule is the RNA secondary structure as described by Mathews and Zukker.

Researchers worldwide have reported increasing interest in the characterization of biomacromolecules, particularly the RNA macromolecular structure, by computational techniques. In this context, we propose here that 2D-RNA-QSAR is a promising field within biomacromolecules research. New analogues of our stochastic molecular descriptors will be introduced for the RNA secondary structure and these descriptors have been largely applied to small molecules and biomacromolecules. Two preliminary studies into secondary QSAR of RNA macromolecules have also been published, but these focus only on local properties of a single RNA molecule. As a consequence, the main aim of the present paper is to introduce in RNA-QSAR studies the Markov electrostatic potentials (${}^{k}\xi_{M}$) previously used for proteins QSAR. In this sense, we intend to predict whether a naturally occurring DNA sequence is an mps or not on the basis of the ${}^{k}\xi_{M}$ calculated for the macromolecular secondary structure of its putative RNA. Consequently, a more specific but still important aim of this work is to introduce a novel approach to predict mps. This work has led to the first 2D-RNA-QSAR to discriminate between two groups comprising several RNA macromolecules, including 135 mycobacterial promoters and 450 control sequences.

2. Results and Discussion

Several authors have studied the mycobacterial promoter sequence problem from the point of view of DNA. Mulder *et al.* listed −35 and −10 DNA regions of a few mycobacterial promoters. *Mycobacteriophage I3* and *M. paratuberculosis* promoter sequences and their similarity with the *E. coli* promoters have been studied by Ramesh and Gopinathan and Bannantine *et al.*, respectively. Kremer *et al.* studied the DNA sequences essential for transcription in promoters like *M. tuberculosis 85A*. It is possible that DNA promoters with a high GC content in the −10 region[52] are the true representatives of the mycobacterial type. An analysis of *M. smegmatis* and *M. tuberculosis* promoters by Bashyam *et al.* showed that there are similarities to *E. coli* 70 promoters; however, in this case the −35 regions showed greater sequence variability. Strohl studied DNA promoter sequences for *Streptomyces* promoters.

O'Neill and Chiafari have also made efforts to develop statistical algorithms for sequence analysis and motif prediction by searching for homologous regions or by comparing the sequence information with a consensus sequence. Two studies by Mulligan and McClure and Mulligan *et al.* pointed out that the variations that exist within individual promoter sequences are primarily responsible for the unsatisfactory results yielded by the promoter-site-searching algorithms, which in essence perform statistical analysis. It can therefore be inferred that recognition of mycobacterial promoter sequences requires a powerful technique that is capable of unravelling those hidden pattern(s) in the biomacromolecule structure – patterns that are difficult to identify visually.

Linear Discriminant Analysis was used to classify RNA macromolecules as mycobacterial promoter sequence (mps) or control group sequence (cs). In the development of the LDA the output was a dummy variable, mps, which codifies whether a sequence lies within the mps class (mps = 1) or belongs to the cs group (mps = 0). In this problem the inputs were the Markov electrostatic potentials (${}^{k}\xi_{M}$) of interaction between nucleotides located with respect to each other at a topologic distance k within the 2D-RNA backbone, with *k* it is in the range [0, 5]. The ${}^{k}\xi_{M}$ are parameters derived by means of a Markov chain model and are used here as molecular descriptors to encode RNA secondary structure (see methods section for details). The best discriminat equation found to discriminate between mps and the control group was:

$$\text{mps} = -4.664 \cdot {}^{0}\xi_{M} + 0.991 \cdot {}^{1}\xi_{M} - 2.432 \qquad (1)$$

$$N = 585 \quad \lambda = 0.41 \quad F = 38.8 \quad p < 0.001$$

Where λ is Wilk's statistic, *N* is the number of RNA sequences studied, *F* is Fisher's statistics and *p* is the *p*-level (probability of error) <0.001. This latter factor means that the hypothesis of groups overlapping with a 5% error can be rejected. A high Matthews' regression coefficient (C = 0.903) was observed and this high C value indicates a strong linear relationship between the structural descriptors of the biomacromolecules and the classification of the RNA sequences. The significance of the two variables (${}^{0}\xi_{M}$ and ${}^{1}\xi_{M}$) in the model was demonstrated with the stepwise analysis (see original work). Conversely, the second order potential ${}^{2}\xi_{M}$ does not have a significant relationship with the mps characteristic or RNA sequences. In physical terms the above results show that, as in other studies, there is a relationship between the electrostatic potential of the RNA molecule and its biological activity. However, in this case not all the electrostatic interactions affect the activity in the same way. The RNA-QSAR predicts that the possibility of a sequence acting as an mps decreases by a factor of 4.664 per unit of electrostatic potential on considering isolated nucleotides (${}^{0}\xi_{M}$). Conversely, the variations of global electrostatic potential (${}^{1}\xi_{M}$) due to secondary structure folding[65] as a result of direct covalent and/or hydrogen bonds between nucleotides increase by a factor of only 0.991 with respect to the possibility of RNA being encoded as an mps. Finally, long-term electrostatic interaction potentials between nucleotides at distances longer than 1 (${}^{2}\xi_{M}$, ${}^{3}\xi_{M}$, ${}^{4}\xi_{M}$) do not correlate with the mps activity. The detailed results of the forward stepwise analysis are given in the original work.

Jack-knife cross validation (cv) experiments were performed by the re-substitution technique, leaving out four different groups selected at random and containing 25% of the RNA molecules. The cross validation accuracies and the average cross validation accuracy (cv-average) were cv1 = 95.9%, cv2 = 96.6%, cv3 = 96.6% and cv4 = 96.5%, respectively, with the average Cv-average = 85.7.

The testing of the model fit to data and its robustness – although very important – is not the only characteristic of an acceptable QSAR.

The data for mps name, sequences, training and cross-validation probabilities for all the RNAs used in this work are given in Table 2SM and Table 3SM of the supplementary material of the original work. Finally, as far as the quality of the model is concerned, we would like to point out that the present linear QSAR model compares very favourably to a previous non-linear model reported by Kalate *et al.* in terms of simplicity (two variables: ${}^{0}\xi_{M}$ and ${}^{1}\xi_{M}$). This non-linear model presented only slightly higher accuracy (97%) but makes use of very large space parameters to describe DNA sequences rather than RNA structure. The success of our RNA-QSAR model, which uses only two variables, can be explained by considering that RNA structure molecular descriptors encode not only sequences (as is the case for DNA linear sequence descriptors) but also molecular branching.

The present paper introduces the simplest up-to-date reported method to predict mycobacterial promoters. With this ultimate aim in mind, we changed the classical point of view and used RNA 2D-macromolecular descriptors instead of DNA sequence analysis. In this sense, this work opens a new way for the application of classical QSAR approaches to biomacromolecules.

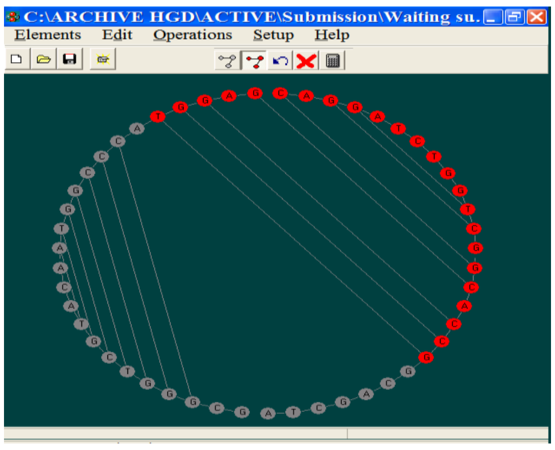


Figure 1. Circular representation for a folded RNA macromolecule of mps T3 from *M. tuberculosis*, note main stem highlighted in red.

4. Conclusions

In accordance with the aims of the work presented here, two main conclusions can be drawn from the results and discussion. Firstly, the 2D structure of RNA can be encoded with ${}^{k}\xi_M$ to develop QSAR studies in the presence of low sequence homology, as in the mps problem. Secondly, there is a very simple linear QSAR model for mps prediction that involves the first two members of the ${}^{k}\xi_M$ series (${}^{0}\xi_M$, ${}^{1}\xi_M$).

Conflicts of Interest

State any potential conflicts of interest here or "The authors declare no conflict of interest".

References and Notes

1. Harshey, R.M.; Ramkrishnan, T. *J. Bacteriol.* **1977**, *129*, 616.
2. Nakayama, M.; Fujita, N.; Ohama, T.; Osawa, S.; Ishihama, A. *Mol. Gen. Genet.* **1989**, *218*, 384.
3. Ohama, T.; Yamao, F.; Muto, A.; Osawa, S. *J. Bacteriol.* **1987**, *169*, 4770.
4. Sanchez, R.; Morgado, E.; Grau, R. *WSEAS Trans. Biol. Biomed.* **2004**, *1*, 190.
5. Sanchez, R.; Morgado, E.; Grau, R. *MATCH* **2004**, *52*, 29.
6. Chou, K.C. *Biopolymers* **1997**, *42*, 837.
7. Di Francesco, V.; Munson, P.J.; Garnier J. *Bioinformatics* **1999**, *15*,131.
8. Vorodovsky, M.; Macininch, J.D.; Koonin, E.V.; Rudd, K.E.; Médigue, C.; Danchin, A. *Nucleic Acid Res.* **1995**, *23*, 3554.
9. Hughey, R.; Krogh, A. *CABIOS*, **1996**, *12*, 95.
10. Yuan, Z. *FEBS Lett.* **1999**, *451*, 23.
11. Kubinyi, H.; Taylor, J.; Ramdsen, C. Quantitative Drug Design, in *Comprehensive Medicinal Chemistry*, Ed. C. Hansch. Pergamon. 1990, vol. 4, p. 589-643.
12. Todeschini, R.; Consonni, V. 2000. *Handbook of molecular descriptors*, Weinheim, Germany, Wiley VCH.

13. Mathews, D.H.; Zuker, M. RNA secondary structure prediction. In *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics* P. Clote ed., John Wiley & Sons, NY. 2004.
14. Ruan, J.; Stormo, G.D.; Zhang, W. *Bioinformatics* **2004**, *20*, 58.
15. Ieong, S., Kao, M.-Y, Lam, T.-W., Sung, W.-K. and Yiu, S.-M. *J. Comp. Biol.* **2003**, *10*, 981–995.
16. González-Díaz, H.; Molina, R.; Uriarte, E. *Polymer* **2004**, *45*, 3845.
17. González-Díaz, H.; Molina, R.; Uriarte. E. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4691.
18. González-Díaz, H.; Bastida, I.; Castañedo, N.; Nasco, O.; Olazabal, E.; Morales, A.; Serrano, H.S.; Ramos de A, R. *Bull. Math. Biol.* **2004**, *66*, 1285.
19. González-Díaz, H.; Gia, O.; Uriarte, E.; Hernádez, I.; Ramos, R.; Chaviano, M.; Seijo, S.; Castillo, J.A.; Morales, L.; Santana, L.; Akpaloo, D.; Molina, E.; Cruz, M.; Torres, L.A.; Cabrera, M.A. *J. Mol. Mod.* **2003**, *9*, 395.
20. González-Díaz, H.; Hernández, S.I.; Uriarte, E.; Santana, L. *Comput. Biol. Chem.* **2003**, *27*, 217.
21. González-Díaz, H.; Olazábal, E.; Castañedo, N.; Hernádez, S.I.; Morales, A.; Serrano, H.S.; González, J.; Ramos de A, R. *J. Mol. Mod.* **2002**, *8*, 237.
22. González-Díaz, H.; Uriarte, E.; Ramos de A, R. *Bioorg. Med. Chem*. **2005**, *13*, 323.
23. Gia, O.; Marciani-Magno, S.; González-Díaz, H.; Quezada, E.; Santana, L.; Uriarte, E.; DallaVia, L. *Bioorg. Med. Chem*. **2005**,*13*, 809.
24. Ramos de A, R.; González-Díaz, H.; Molina, R.; Uriarte, E. *Prot. Struct. Func. Bioinf.* **2004**, *56*, 715.
25. González-Díaz, H.; Marrero, Y.; Hernández, I.; Bastida, I.; Tenorio, I.; Nasco, O.; Uriarte, E.; Castañedo, N.; Cabrera-Pérez, M.A.; Aguila, E.; Marrero, O.; Morales, A.; González, M.P. *Chem. Res. Tox.* **2003**, *16*, 1318.
26. González-Díaz, H.; Ramos de A, R.; Molina, R. *Bioinformatics* **2003**, *19*, 2079.
27. González-Díaz, H.; Ramos de A, R.; Molina, R. *Bull. Math. Biol.* **2003**, *65*, 991.
28. Saíz-Urra, L.; González-Díaz, H.; Uriarte, E. *Bioorg. Med. Chem.* **2005**, *13*, 3641.
29. Norberg, J.; Nilsson, L. *Acc. Chem. Res.* **2002**, *35*, 465.
30. González-Díaz, H.; Molina, R.; Sanchez, I. BIOMARKS ©, **2004**, *version 1.0.*
31. Mathews, D.H.; Zuker, M.; Turner, D.H. RNAStructure ©, **2002**, *version 4.0*.
32. Marrero-Ponce, Y.; González-Díaz, H.; Romero-Zaldivar, V.; Torrens, F.; Castro, E.A. *Bioorg. Med. Chem.* **2004**, *12*, 5331.
33. Marrero-Ponce, Y. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2010.
34. Marrero-Ponce, Y.; Montero-Torres, A.; Romero-Zaldivar, C.; Iyarreta-Veitía, M.; Mayón-Peréz, M.; García-Sánchez, R. *Bioorg. Med. Chem.* **2005**, *13*, 1293.
35. González-Díaz, H.; Cruz-Monteagudo, M.; Molina, R.; Tenorio, E.; Uriarte, E. *Bioorg. Med. Chem*. **2005**, *13*, 1119.
36. González-Díaz, H.; Agüero, G.; Cabrera, M.A.; Molina, R.; Santana, L.; Uriarte, E.; Delogu, G.; Castañedo, N. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 551.
37. Mulder, M.A.; Zappe, H.; Steyn, L.M., *Tuber. Lung Dis.* **1997**, *78*, 211.
38. Ramesh, G.; Gopinathan, K.P., *Indian J. Biochem. Biophys.* **1995**, *32*, 361.
39. Bannantine, J.P.; Barletta, R.G.; Thoen, C.O.; Andrews, R.E., Jr., *Microbiol.* **1997**, *143*, 921.

40. Kremer, L.; Baulard, A.; Estaquier, J.; Content, J., Capron, A.; Locht, C. *J. Bacteriol.* **1995**, *177*, 642.

SciForum
Mol2Net

Pairwise Ortholog Detection in Related Yeast Species by Using Big Data Supervised Classifications

Deborah Galpert Cañizares [1], Sara del Río García [2], Francisco Herrera [2], Evys Ancede Gallardo [3], Agostinho Antunes [4,5], Guillermin Agüero-Chapin [4,*]

[1] Departamento de Ciencias de la Computación, Universidad Central ¨Marta Abreu¨ de Las Villas (UCLV), Santa Clara, 54830, Cuba; E-Mail: deborah@uclv.edu.cu

[2] Dept. of Computer Science and Artificial Intelligence, CITIC-UGR, University of Granada, Granada, Spain; E-Mails: srio@decsai.ugr.es (S.R.G.); herrera@decsai.ugr.es (F.F.)

[3] Centro de Bioactivos Químicos, Universidad Central ¨Marta Abreu¨ de Las Villas (UCLV), Santa Clara, 54830, Cuba; E-Mail: eancedeg@uclv.edu.cu

[4] CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Rua dos Bragas, 177, 4050-123 Porto, Portugal; E-Mail: aantunes@ciimar.up.pt

[5] Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal

[*] Author to whom correspondence should be addressed; E-Mail: gchapin@ciimar.up.pt.

Received: 1 October 2015 / Accepted: 1 October 2015 / Published: 2 December 2015

Abstract: Orthology detection still requires more effective scaling algorithms. Combinations of alignment, synteny, evolutionary distances and protein interactions have been used in different unsupervised algorithms to improve effectiveness while many available databases are concerned with the scaling problem. In this paper, a set of gene pair features based on similarity measures, such as alignment scores, sequence length, gene membership to conserved regions and physicochemical profiles are combined in a supervised Pairwise Ortholog Detection (POD) approach to improve effectiveness considering low ortholog ratios in relation to all possible pairwise comparisons between two genomes. In this POD scenario, big data supervised classifiers managing imbalance between ortholog and non-ortholog pair classes allow for an effective scaling solution built from two genomes and extended to other genome pairs. The supervised approach for POD was compared with Reciprocal Best Hits (RBH), Reciprocal Smallest Distance (RSD) and a Comprehensive, Automated Project for the Identification of Orthologs from Complete Genome Data (OMA) algorithms by using (i) *Saccharomyces cerevisiae - Kluyveromcyes lactis*, (ii) *Saccharomyces cerevisiae - Candida glabrata* and (iii) *Saccharomyces cerevisiae - Schizosaccharomyces pombe* yeast genome pairs as benchmark datasets. Four datasets derived from each genome pair comparison with different alignment settings were used. Because of the large amount of instances (gene pairs) and the data imbalance, the building and testing of the supervised model was only possible by using big data supervised classifiers managing imbalance. Evaluation metrics taking low ortholog ratios into account were applied. From the effectiveness perspective, MapReduce Random Oversampling combined with Spark Support Vector Machines outperformed RBH, RSD and OMA, probably, because of the consideration of gene pair features beyond alignment similarities combined with the advances in big data supervised classification.

Keywords: ortholog detection; big data supervised classification; similarity measures

1. Introduction

Ortholog detection (OD) algorithms should distinguish orthologous genes from other types of homologs such as paralogs evolving from a common ancestor through a duplication event. A great deal of unsupervised graph-based approaches has been developed to identify orthologs resulting in corresponding repositories for pre-computed orthology relationships.

When OD is based only on sequence similarity, it has been limited by evolutionary processes such as recent paralogy events, horizontal gene transfers, gene fusions and fissions, domain recombinations or different genetic events [1-2]. In fact, the identification of homologs is a difficult task in the presence of short sequences, those that evolved in a convergent way, and the ones that share less than 30% of amino acid identities (twilight zone). Algorithm failures have been particularly shown in benchmark datasets from *Saccharomycete* yeast species that underwent whole genome duplications (WGD) presenting rampant paralogies and differential gene losses [3]. To tackle these shortcomings, some OD solutions merge sequence similarity with synteny genome rearrangements, protein interactions, domain architectures and evolutionary distances.

On the other hand, the integration of different gene or protein information and the massive increase in complete proteomes highly increase the dimensionality of the OD problem and the total number of proteins to be classified. In a thorough paper from the Quest for Orthologs consortium [4], the authors emphasize the idea that this increase in proteome data brings out the need to work out not only efficient but effective OD algorithms. As they mention, the increase in computational demands in sequence analyses is not easily met by an increase in computational capacities but rather calls for new approaches or algorithmic implementations [4]. They summarized some methodological shortcuts implemented by the existing orthology databases to deal with the scaling problem.

In this paper, we propose a new supervised approach for pairwise OD (POD) that combines several gene pairwise features (alignment-based and synteny measures with others derived from the pairwise comparison of the physicochemical properties of amino acids) to address big data problems [4]. Our big data supervised POD approach allows scaling to related species and data imbalance management (low ortholog ratio found in two or more genomes) for an effective OD. The methodology consists of three steps: (1) the calculation of gene pair features to be combined, (2) the building of the classification model using machine learning algorithms to deal with big data from a pairwise dataset, and (3) the classification of related gene pairs.

Since traditional supervised classifiers cannot scale large datasets, the supervised classification for the POD problem should be addressed as a big data classification problem according to [5-7] and big data solutions should be applied for binary classification in imbalanced data such as the ones presented in [8].

Finally, we evaluate the application of several big data supervised techniques that manage imbalanced datasets [8-9] such as cost-sensitive Random Forest (RF-BDCS), Random Oversampling with Random Forest (ROS+RF-BD) and the Apache Spark Support Vector Machines (SVM-BD) [9] combined with MapReduce ROS (ROS+SVM-BD). The effectiveness of the supervised approach is compared to RBH, RSD and OMA algorithms, taking data imbalance into account. All the algorithms were evaluated on benchmark datasets derived from the following yeast genome pairs: *S. cerevisiae* and *K. lactis, S. cerevisiae* and *C. glabrata* [3] and *S. cerevisiae* and *S. pombe* [10]. The *S. cerevisiae* and *C. glabrata* pair is particularly complex for OD since both species had undergone WGD. We found that our supervised approach outperformed traditional methods, mainly when we applied ROS combined with SVM-BD.

2. Results and Discussion

For the evaluation of POD algorithms, we compare the supervised solutions and the unsupervised ones following the evaluation scheme in Figure 1. The process separates the pairs into train and test sets and calculates pairwise similarity measures (average of local and global alignment similarity measures, length of sequences, gene membership to conserved regions (synteny), and physicochemical profiles within 3, 5 and 7 window sizes) for the pairs of both sets. The sequences of the test sets should be used to run the unsupervised reference algorithms. The train set should be used for building the supervised models to be tested only with the test set.

The performance quality evaluation involves the calculation of the Geometric Mean (*G-Mean*) [11], seeking to maximize the accuracy of the two classes (orthologs and non-orthologs) by achieving a good balance between sensitivity and specificity that consider misclassification costs; and the Under the ROC Curve (*AUC*) [12] to show the classifier performance over a range of data distributions [13].

In Experiment 1, we evaluated the algorithms inside a genome by partitioning at random 75% of the complete set of pairs for training and 25% for testing, while in Experiment 2 we built the model from a genome pair and tested it in two different pairs. Specifically, in Experiment 1 we divided the *S. cerevisiae* - *K. lactis* set into 16.986.996 pairs for training and 5.662.332 pairs for testing. The four datasets (BLOSUM50, BLOSUM62_1, BLOSUM 62_2 and PAM250) of each genome pair were built from combinations of alignment parameter settings. On the other hand, in Experiment 2, we built the classification model from 22.649.328 pairs of *S. cerevisiae* and *K. lactis* genomes and tested it in 29.887.416 pairs of *S. cerevisiae* and *C. glabrata*, and 8.095.907 pairs of *S. cerevisiae* and *S. pombe* genomes.

Comparison of big data supervised classifiers

The *G-Mean* values of the supervised algorithms change only slightly with the selection of different alignment parameters (Table 1). These results may be either caused by the aggregation of global and local alignment scores in a single similarity measure or by the appropriate combination of scoring matrices and gap penalties in relation to the sequence diversity between the two yeast genomes [14].

The average results of *AUC* and *G-Mean* obtained in experiments 1 and 2 for the supervised algorithms with different parameter values are shown in Table 1. The average TP_{Rate} and TN_{Rate} are also depicted in Figure 2. SVM-BD has been left out from the table due to its very poor performance in *G-Mean* caused by its imbalance between TP_{Rate} and TN_{Rate}. Both Table 2 and Figure 2 prove that big data supervised classifiers managing imbalance outdo their corresponding big data supervised versions.

The ROS pre-processing method for big data makes SVM-BD useful for POD and improves the performance of RF-BD even more with a higher value for the resampling size parameter of 130% [15]. In contrast, both experiments show that the variation in this parameter value from 100% to 130% does not significantly influence on the performance of the SVM-BD classifier with different regulation values.

Specifically, RF-BDCS shows the best performance in *S. cerevisiae - C. glabrata* and *S . cerevisiae - K. lactis* when the classification quality is measured by *G-Mean* and *AUC* metrics, because it enhances the learning of the minority class. The criterion used to select the best tree split is based on the weighting of the instances according to their misclassification costs, and such costs are also considered to calculate the class associated with a leaf [8]. This cost treatment does not explicitly change the sample distribution and avoids the possible overtraining, that it is present in the ROS solutions due to replicated cases. The election of the cost values ($C(+|-) = IR$ and $C(-|+) = 1$) may also define the success of the algorithm.

In the case of SVM-BD, the fixed regularization parameter defines the trade-off between the goal of minimizing the training error (i.e., the loss) and minimizing the model complexity to avoid overfitting. The higher is its value, the simpler the model. Nonetheless, setting an intermediate value, or one close to cero may produce a better performance in classification [16]. This is the case of the ROS (RS: 100%) + SVM-BD (regParam: 0.5) classifier that exhibits the best *AUC* and *G-Mean* values in *S. cerevisiae - S. pombe*, and the best balance between TP_{Rate} and TN_{Rate} in the three datasets (Figure 2).

In order to balance time with classification quality, time consumption is another aspect to have in mind when comparing big data solutions. Table 3 contains run time in seconds for all big data solutions in each dataset and the faster algorithms are highlighted in bold face. These results allow us to prove that the time required is directly related to the operations needed for each method, as well as to the size of the datasets used to build the model. The fastest algorithm considering the average run time is SVM-BD followed by SVM-BD combined with ROS. Thus, the fastest algorithms coincide with the ones with better performance. In general, the ROS (RS: 100%) + SVM-BD (regParam: 0.5) classifier can be considered the best supervised solution considering both performance and time.

Comparison of supervised vs. unsupervised classifiers

The average results of *AUC* and *G-Mean* obtained for the best supervised algorithms and the unsupervised algorithms with different parameter values are shown in Table 4 for experiments 1 and 2. The supervised classifiers outperform the unsupervised ones. Among the unsupervised algorithms, RSD reaches the highest *G-Measure* value by setting E-value = 1e-05 and α = 0.8 (recommended values in [17]) in *S. cerevisiae - C. glabrata* where similar results can also be seen for *AUC* and TP_{Rate} values. On the contrary, OMA was the best among the unsupervised algorithms in *S. Cerevisiae - S. pombe* datasets (Table 4).

In general, the performance of all classifiers declined in *S. Cerevisiae - S. pombe* datasets due to the fact that *S. pombe* is a distant relative of *S. cerevisiae* [18]. The supervised classifiers performance is affected for the same reason and also, by the difference in data distribution between the train and test sets [19]. On the contrary, ROS (RS: 100%) + SVM-BD (regParam: 0.5) remained stable in *S. Cerevisiae*

- *C. glabrata* and *S. Cerevisiae* - *S. pombe* datasets when considering the balance between TP_{Rate} and TN_{Rate}. Superior results in *S. cerevisiae* - *C. glabrata* are outstanding, since both genomes underwent a WGD and a subsequent differential loss of gene duplicates, so that algorithms are prone to produce false positives. Thus, this dataset contains “traps” for OD algorithms [3].

The reduced quality shown by RBH, RSD and OMA, mainly in the case of RBH, could be caused by their initial assumption that the sequences of orthologous genes/proteins are more similar to each other than they are to any other genes from the compared organisms. This assumption may produce classification errors [1], in spite of the fact that BLAST parameters can be tuned as has been recommended in [20]. Conversely, RSD not only compares the sequence similarity, but it relies on maximum likelihood estimation of evolutionary distances to detect orthologs between two genomes, and as a result, it finds many putative orthologs missed by RBH because it is less likely than RBH to be misled by existing close paralogs.

The OMA algorithm also displays advantages over RBH. It uses evolutionary distances instead of alignment scores. This algorithm allows the inclusion of one-to-many and many-to-many orthologs. It also considers the uncertainty in distance estimations and detects potential differential gene losses.

From the point of view of the intrinsic information managed by the algorithms, the success of big data supervised classifiers managing imbalance over RSD and OMA may be explained by feature combinations calculated for the datasets together with the learning from curated classifications. With the aggregation of global and local alignment scores we are combining protein structural and functional relationships between sequence pairs, respectively. Besides, we incorporate other gene pair features: (i) the periodicity of the physicochemical properties of amino acids that allows us to detect similarity among protein pairs in their spectral dimension [21]; (ii) the conserved neighbourhood information, which considers that genes belonging to the same conserved segment in genomes of different species will probably be orthologs; and (iii) the length of sequences

Table 1. Geometric mean results of the best supervised classifiers in each dataset.

Dataset	**ROS (RS: 100%) + RF-BD (Scer-Klac)**	**ROS (RS: 130%) + RF-BD (Scer-Klac)**	**RF-BDCS (Scer-Klac)**	**ROS (RS: 100%) + RF-BD (Scer-Cgla)**	**ROS (RS: 130%) + RF-BD (Scer-Cgla)**	**RF-BDCS (Scer-Cgla)**	**ROS (RS: 100%) + SVM-BD (regParam: 1.0) (Scer-Spombe)**	**ROS (RS: 100%) + SVM-BD (regParam: 0.5) (Scer-Spombe)**
Blosum50	0.9818	0.9818	**0.9896**	0.9889	0.9885	**0.9934**	0.8393	**0.8673**
Blosum621	0.9801	0.9818	**0.9855**	0.9891	0.9903	**0.9932**	0.8707	**0.8959**
Blosum622	0.9793	0.9793	**0.9905**	0.9910	0.9910	**0.9929**	0.8536	**0.8694**
Pam250	0.9818	0.9818	**0.9899**	0.9912	0.9905	**0.9941**	0.8495	**0.8839**

Table 2. *AUC* and *G-Mean* results of supervised classifiers in experiments 1 and 2.

	S.cerevisiae-S.Klactis		***S.cerevisiae-C.glabrata***		***S.cerevisiae-S.pombe***	
Algorithm	***AUC***	***G-Mean***	***AUC***	***G-Mean***	***AUC***	***G-Mean***
RF-BD	0.6979	0.6291	0.7455	0.7005	0.5172	0.1851
ROS (RS: 100%)+RF-BD	0.9809	0.9807	0.9901	0.9900	0.6096	0.4527
ROS (RS: 130%)+RF-BD	0.9813	0.9812	0.9901	0.9901	0.6121	0.4581
RF-BDCS	**0.9889**	**0.9889**	**0.9934**	**0.9934**	0.7294	0.6745
ROS (RS: 100%) + SVM-BD (regParam: 1.0)	0.9477	0.9477	0.9542	0.9542	0.8632	0.8533
ROS (RS: 100%) + SVM-BD (regParam: 0.5)	0.8845	0.8791	0.9540	0.9539	**0.8845**	**0.8791**
ROS (RS: 100%) + SVM-BD (regParam: 0.0)	0.6135	0.4961	0.9432	0.9431	0.6135	0.4961
ROS (RS: 130%) + SVM-BD (regParam: 1.0)	0.8164	0.7956	0.9523	0.9522	0.8164	0.7956
ROS (RS: 130%) + SVM-BD (regParam: 0.5)	0.8629	0.8528	0.9539	0.9539	0.8629	0.8528
ROS (RS: 130%) + SVM-BD (regParam: 0.0)	0.6248	0.5147	0.9429	0.9428	0.6248	0.5147

Table 3. Run time results in seconds of the big data solutions in experiments 1 and 2.

Algorithm	***S.cerevisiae-S.Klactis***	***S.cerevisiae-C.glabrata***	***S.cerevisiae-S.pombe***
RF-BD	1201.59	2174.90	2060.99
ROS (RS: 100%)+RF-BD	2983.75	4562.38	4440.03
ROS (RS: 130%)+RF-BD	3345.04	4805.50	4681.51
RF-BDCS	1302.41	2362.04	2025.15
SVM-BD	**461.87**	**482.85**	**480.45**
ROS (RS: 100%) + SVM-BD (regParam: 1.0)	**867.38**	**1011.59**	**1012.46**
ROS (RS: 100%) + SVM-BD (regParam: 0.5)	**874.62**	**1008.77**	**1013.32**
ROS (RS: 100%) + SVM-BD (regParam: 0.0)	**859.17**	**1008.24**	**999.31**
ROS (RS: 130%) + SVM-BD (regParam: 1.0)	927.14	1079.19	1079.58
ROS (RS: 130%) + SVM-BD (regParam: 0.5)	929.17	1084.19	1076.33
ROS (RS: 130%) + SVM-BD (regParam: 0.0)	924.42	1076.37	1077.21

Table 4. *AUC* and *G-Mean* of the unsupervised and the best supervised classifiers.

	S. cerevisiae-.K. lactis		***S. cerevisiae-C .glabrata***		***S. cerevisiae-S. pombe***	
Algorithm	***AUC***	***G-Mean***	***AUC***	***G-Mean***	***AUC***	***G-Mean***
RBH	0.1497	0.0062	0.8196	0.7995	0.4697	0.4525
RSD 0.2 1e-20	0.5862	0.4862	0.9238	0.9206	0.4874	0.4438
RSD 0.5 1e-10	0.5926	0.4643	0.9340	0.9316	0.4980	0.4063
RSD 0.8 1e-05	0.5886	0.4518	0.9382	0.9362	0.5009	0.3899
OMA	0.5765	0.4904	0.9287	0.9259	0.5151	0.4644
RF-BDCS	**0.9889**	**0.9889**	**0.9934**	**0.9934**	0.7294	0.6745
ROS (RS: 100%) + SVM-BD (regParam: 1.0)	0.9477	0.9477	0.9542	0.9542	0.8632	0.8533
ROS (RS: 100%) + SVM-BD (regParam: 0.5)	0.8845	0.8791	0.9540	0.9539	**0.8845**	**0.8791**

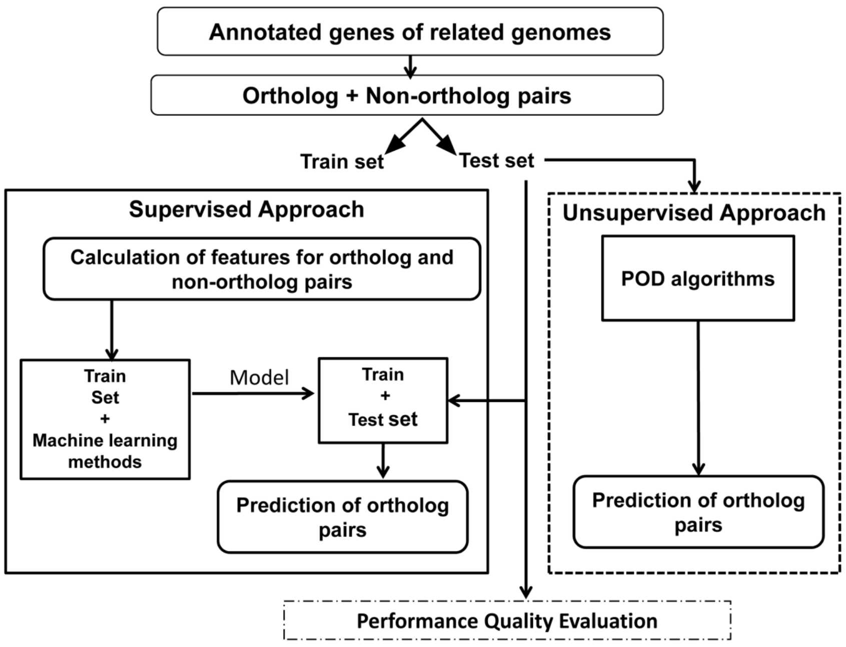

Annotated genes of related genomes
Ortholog + Non-ortholog pairs
Train set
Test set
Supervised Approach
Calculation of features for ortholog and non-ortholog pairs
Train Set + Machine learning methods
Model
Train + Test set
Prediction of ortholog pairs
Unsupervised Approach
POD algorithms
Prediction of ortholog pairs
Performance Quality Evaluation

Figure 1. Workflow of the evaluation of supervised *vs*. unsupervised POD algorithms.

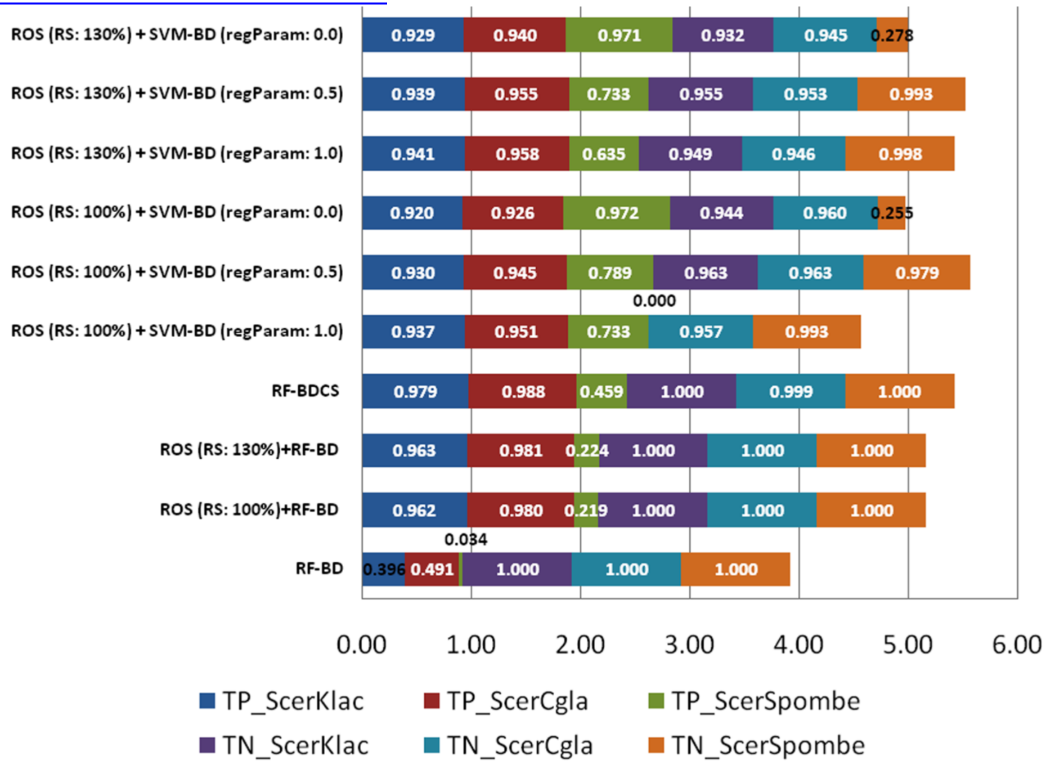


Figure 2. Average true positive and true negative rate values of supervised classifiers obtained in experiments 1 and 2.

3. Materials and Methods

Datasets

The characteristics of the datasets are summarized in Table 5 where the label #Atts represents the number of attributes or gene pair features, and #Class (maj; min), the number of pairs in both classes. *S. cerevisiae - S. pombe* dataset contains ortholog pairs representing 95.18% of the union of the Inparanoid7.0 and GeneDB classifications described in [10]. On the other hand, *S. cerevisiae - K. lactis* and *S. cerevisiae - C. glabrata* datasets contain all ortholog pairs in the gold groups reported in [3]. When we built the set of instances with all possible pairs, we excluded some genes since we didn't find their genome physical location data in the YGOB database [22], required for the conserved membership feature calculation.

Big data supervised classification managing data imbalance

We use the open-source project Hadoop [23] with its highly scalable and fault-tolerant Hadoop Distributed File System (HDFS). We also utilize the scalable Mahout data mining and machine learning library [24] with machine learning algorithms adapted according to the MapReduce scheme as the MapReduce implementation of the (Random Forest (RF) algorithm [25]. Finally, we use the Apache Spark framework [9] interacting with HDFS, when the implementation of SVM-BD in the scalable MLLib machine learning library [16] is combined with the MapReduce ROS implementation [8].

Table 5. Characteristics of the datasets.

Genome pair	#Atts	#Class (maj; min)	Imbalance ratio (*IR*)	Excluded genes
S. cerevisiae - K. lactis 1	6	(22.646.914; 2414)	9381.489	89 de 5861 genes de *S. cerevisiae* 37 de 5215 genes de *C. glabrata* 1403 de 5327 genes de *K. lactis*
S. cerevisiae - C. glabrata 1	6	(29.884.575; 2841)	10519.034	
S. cerevisiae - S. pombe 2	6	(8.090.950; 4.957)	1632.227	

4. Conclusions

The development of effective supervised algorithms for POD in a big data scenario was made possible by: (i) the availability of curated databases (authentic orthologs), (ii) the combination of traditional alignment measures with other gene pair features (sequence length, gene membership to conserved regions and physicochemical profiles) to complement homology detection, and (iii) the treatment of the low ratio of orthologs to the total possible gene pairs between two genomes. By applying evaluation metrics such as *G-mean*, AUC and the balance between TP_{Rate} and TN_{Rate}, our results show that gene pairwise feature combinations provide excellent POD in a big data supervised scenario that consider data imbalance. The SVM-BD classifier combined with the ROS (RS: 100%) pre-processing with regulation parameter 0.5 outdid the rest of the big data supervised solutions and the popular unsupervised (RBH, RSD and OMA) algorithms even when the supervised model was extended to datasets containing "traps" for OD algorithms. The classification performance of the supervised algorithms measured by *G-Mean* and AUC metrics did not significantly change in the four test sets obtained with different alignment parameter settings. When the balance between time and classification quality is considered, ROS (RS: 100%) + SVM-BD (regParam: 0.5) also proves to be the algorithm of choice. In future research, the introduction of new gene pair features might improve the effectiveness and efficiency of the supervised algorithms for POD.

Acknowledgments

GACh acknowledge the Portuguese Fundação para a Ciência e a Tecnologia (FCT) for financial support with reference (SFRH/BPD/92978/2013). AA was partially supported by the European Regional Development Fund (ERDF) through the COMPETE - Operational Competitiveness Programme and national funds through FCT under the projects PEst-C/MAR/LA0015/2013 and PTDC/AAC-AMB/121301/2010 (FCOMP-01-0124-FEDER-019490). This work was also partially supported by the Spanish Ministry of Science and Technology under project TIN2014-57251-P and the Regional Andalusian Research Projects P11-TIC-7765 and P10-TIC-6858.

Author Contributions

Conceived and designed the experiments: DGC and GACh. Performed the experiments: DGC, SRG and EAG. Analyzed the data: DGC, SRG, FH and GACh, Contributed reagents/materials/analysis tools: FH, EAG, and AA. Wrote the paper: DGC, SRG and GACh. Critically revised the manuscript: GACh, FH and AA.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Kristensen, D.M.; Wolf, Y.I.; Mushegian, A.R.; Koonin, E.V. Computational methods for gene orthology inference. *Briefings in bioinformatics* **2011**, *12*, 379-391.
2. Kuzniar, A.; Ham, R.C.H.J.v.; Pongor, S.; Leunissen, J.A.M. The quest for orthologs: Finding the corresponding gene across genomes. *Trends in Genetics* **2008**, *30*, 1-13.

3. Salichos, L.; Rokas, A. Evaluating ortholog prediction algorithms in a yeast model clade. *PLoS ONE* **2011**, *6*, 1-11.
4. Sonnhammer, E.L.L.; Gabaldón, T.; Sousa da Silva, A.W.; Martin, M.; Robinson-Rechavi, M.; Boeckmann, B.; Thomas, P.D.; Dessimoz, C.; Orthologs, c.Q.f. Big data and other challenges in the quest for orthologs. *Bioinformatics Editorial* **2014**, 1-6.
5. Fernández, A.; Río, S.d.; López, V.; Bawakid, A.; Jesus, M.J.d.; Benítez, J.M.; Herrera, F. Big data with cloud computing: An insight on the computing environment, mapreduce, and programming frameworks. In *WIREs Data Mining Knowl Discov*, 2014.
6. Beyer, M.; Laney, D. 3d data management: Controlling data volume, velocity and variety. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf
7. Chen, C.L.P.; Zhang, C.Y. Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences* **2014**, *275*, 314–347.
8. Río, S.d.; López, V.; Benítez, J.M.; Herrera, F. On the use of mapreduce for imbalanced big data using random forest. *Information Sciences* **2014**, *285*, p. 112-137.
9. Zaharia, M.; Chowdhury, M.; Das, T.; Dave, A.; Ma, J.; McCauley, M.; Franklin, M.; Shenker, S.; Stoica, I. Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing. In *9th USENIX Conference on Networked Systems Design and Implementation*, San Jose, CA, 2012; pp 1-14.
10. Koch, E.N.; Costanzo, M.; Bellay, J.; Deshpande, R.; Chatfield-Reed, K.; Chua, G.; D'Urso, G.; Andrews, B.J.; Boone, C.; Myers, C.L. Conserved rules govern genetic interaction degree across species. *Genome Biology* **2012**, *13*.
11. Barandela, R.; Sánchez, J.S.; García, V.; Rangel, E. Strategies for learning in class imbalance problems. *Pattern Recognit.* **2003**, *36*, 849–851.
12. Bradley, A.P. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition* **1997**, *30*, 1145–1159.
13. He, H.; Garcia, E.A. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering* **2009**, *21*, 1263-1284.
14. Pearson, W.R. Selecting the right similarity-scoring matrix. *Current Protocols in Bioinformatics* **2013**, *43*, 3.5.1-3.5.9.
15. Triguero, I.; Río, S.d.; López, V.; Bacardit, J.; Benítez, J.M.; Herrera, F. Rosefw-rf: The winner algorithm for the ecbdl'14 big data competition: An extremely imbalanced big data bioinformatics problema. *Knowledge-Based Systems* **2015**.
16. Krishnan, S.; Smith, V. Linear support vector machines (svms). <https://spark.apache.org/docs/latest/mllib-linear-methods.html#linear-support-vector-machines-svms> (January 2015),
17. DeLuca, T.F.; Wu, I.-H.; Pu, J.; Monaghan, T.; Peshkin, L.; Singh, S.; Wall, D.P. Roundup: A multi-genome repository of orthologs and evolutionary distance. *Bioinformatics* **2006**, *22*, 2044-2046.
18. Wood, V.; Piskur, P.J. Schizosaccharomyces pombe comparative genomics; from sequence to systems. In *Topics in Current Genetics*, P. Sunnerhagen, J. Piškur (Eds.): Comparative Genomics ed.; Springer-Verlag Berlin Heidelberg 2005: 2005; Vol. 15

19. Moreno-Torres, J.G.; Llorà, X.; Goldberg, D.E.; Bhargava, R. Repairing fractures between data using genetic programming-based feature extraction: A case study in cancer diagnosis. . *Information Sciences* **2013**, 805-823.
20. Hagelsieb, G.M.; Latimer, K. Choosing blast options for better detection of orthologs as reciprocal best hits. *Bioinformatics* **2008**, *24*, 319-324.
21. Carpio-Muñoz, C.A.D.; Carbajal, J.C. Folding pattern recognition in proteins using spectral analysis methods. *Genome Informatics* **2002**, *13*, 163-172
22. Byrne, K.P.; Wolfe, K.H. The yeast gene order browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Research* **2005**, *15*, 1456–1461.
23. White, T. Hadoop, the definitive guide.
24. Owen, S.; Anil, R.; Dunning, T.; Friedman, E. Mahout in action. Manning Publications Co.: 2011.
25. Hakim, D.A. Partial data mapreduce random forests. https://mahout.apache.org/users/classification/partial-implementation.html

SciForum
Mol2Net

Fatty Acids Distribution Networks in Ruminal Membrane by Computational and Experimental Studies

Yong Liu [1,2,3]*, Zhiliang Tan [2], Claudia Giovanna Peñuelas-Rivas [1] and Esvieta Tenorio-Borroto [1]

[1] Faculty of Veterinary Medicine and Animal Science, Autonomous University of the State of Mexico, Toluca, 50090, México

[2] Key Laboratory of Subtropical Agro-ecological Engineering, Institute of Subtropical Agriculture, the Chinese Academy of Sciences, Changsha, Hunan, 410125, P. R. China

[3] Computer Science Faculty, University of A Coruña, Campus de Elviña s/n, A Coruña, 15071, Spain

* Yong Liu, e-mail: y.liu@udc.es. Published in *Mol. BioSyst.*, 2015, Aug. DOI: 10.1039/C5MB00325C

Received: 6 October 2015 / Accepted: 8 October 2015 / Published: 2 December 2015

Abstract: The present communication introduces a new classification model for fatty acids (FA) distribution networks in ruminal microbe membrane based on experimental and computational studies. In the experimental part, long chain fatty acids and volatile fatty acids in ruminal microbe membrane or liquid phase were investigated by supplementation of different ratios of Omega-6 / Omega-3 and in the processes of base- / acid- methylation. In the computational part, Perturbation Theory (PT) and Linear Free-Energy Relationships (LFER), combined with corresponding Box-Jenkins (ΔV_{kj}) and PT Operators ($\Delta\Delta V_{kj}$) were applied into the calculation of physicochemical parameters (V_k) of fatty acids. The best PT-LFER model found to predict the effects of perturbations over the FA distribution network with Sensitivity, Specificity, and Accuracy > 80% for 407,655 cases. In final, PT-LFER model based on LDA was used to reconstruct the complex networks of perturbations in the FA distribution and compared with random Erdős–Rényi network models. The detail results have been published in *Mol. BioSyst.*, 2015, Aug., the present is a short communications.

Keywords: Fatty acids, Distribution networks, Computational, Experimental, Ruminal membrane

1. Introduction

The ω-6/ω-3 ratio plays an important role not only in the pathogenesis of cardiovascular diseases, but also in inflammatory, cancer and autoimmune diseases[1-3]. It is an efficient way to benefits the fatty acids composition in the diets by enrichment of ruminant meat or milk with ω-3 polyunsaturated fatty acids (PUFAs). However, the ruminal complex biohydrogenation process limits their bioavailability of Omega-3 [4]. On the other hand, methylation methods are directed effected the experimental values of PUFAs or volatile fatty acids (VFAs) distribution[5-7]. In the view of biology, the structure properties of long chain fatty acids (LCFAs, especially the number, location or topology structure of double bonds) are highly related to the chronic disease. To address this problem, it was postulated that the LCFAs in ruminal microbial membranes change with the supply of ω-6/ω-3 ratios.

Chemoinformatics is related to Chemometrics, data mining, Machine Learning [8] accompany with Chemistry, Information Science, and the areas of topology, chemical graph theory. A new Corwin Hansch model is based on lipophilicity-activity relationship [9].

$$f(\varepsilon_i) = a_0 + a_1 \cdot \log P_i + a_2 \cdot pK_a + a_3 \cdot MR - a_4 \cdot (\log P_i)^2$$

Steric, electrostatic, and hydrophobicity factors combined with water/n-octanol partition coefficients (P_i), molecular refractivity (MR), acidity constants logarithmic (pKa) [10] *etc.* might be biologically relevant were set as input variables of model [11, 12]. A molecular property (ε_i) or a function of this property $f(\varepsilon_i)$ for a given chemical compound or molecular entity (m_i) was set as output of the model.

Hansch model is an extra-thermodynamic approach closely related to LFER [13, 14]. The input variables (${}^iV_k$) can be calculated as physicochemical parameters or molecular descriptors. In fact, the basic assumption for Hansch's analysis is based on the theory of similar molecules with similar activities.[15-17] In addition, the "*small*" variations or perturbations at the molecular structural level need to be quantified.

The Perturbation Theory (PT)[18] can be used to account for "*small*" problem by the view of Chemoinformatics. In this work, PT and LFER ideas were used to formulate a new PT-LFER approach for complex networks of FA distribution in Lipidomics. In this work, the first experiments were carried out to determine LCFA compositions in the rumen microbiome by addition of different ratios of omega-3 / omega-6. Then, Chemoinformatics study was provided, and the validation of new PT-LFER classification models. Artificial Neural Networks (ANNs) were used to test PT-NLFER models (compared to Non-Linear). Next, the best PT-LFER model found was used to predict the effect of perturbations on initial boundary conditions over a large complex network of FA distribution/uptake in the ruminal microbiome. After that, the observed complex network was constructed and compared to the predicted network and random networks model with similar scale for the first time.

In a word, the present work paves the way to combine the perturbation theory with complex fatty acids molecular systems under the consideration of chemical structures and various boundary experimental conditions.

2. Results and Discussion

The imbalance intake of ω-6/ω-3 has the potential to induce some chronic diseases, such as inflammation, asthma, vascular disease [19]. In this study, long chain fatty acids (LCFAs) in microbial membrane, volatile fatty acids (VFAs) in media were differentiated on the conditions of exogenous ω-6/ω-3 ratios. First, *cis*-FA increased and *trans*-FAs decreased with exogenous ω-3 PUFA in bacteria phase. This reflects that the ratio of *cis*/*trans*-FAs increased with the exogenous ω-3 PUFA ratios. That means exogenous PUFAs are degraded by rumen microorganisms to some extent, or have more complex metabolism processes to intermediary metabolism in both of *cis*- and *trans*- unsaturated FAs formulation. This study showed that ω-3 PUFA (α-linolenic acid) could increase the *cis*-FAs content compared to ω-6 PUFA (linoleic acid) on both of bacteria and protozoa phases. The biohydrogenation of linoleic acid (*cis* 9, *cis* 12- C18:2) is first isomerized to conjugated linoleic acid (CLA, *cis* 9, *trans* 11- C18:2 isomer) in rumen environment, then conversion to vaccenic acid (*trans* 11- C18:1), or further reduction to stearic acid (saturated C18:0) [20]. Whereas the bio-hydrogenation of α-linolenic acid (ALA) in rumen is first characterized by isomerization to isomer (9, 11, 15- *cis*, *trans*, *cis*- C18:3) and subsequent reduction by isomerase and/or reductase via *cis*, *trans*- isomers of C18:2, C18:1 and in final to stearic acid [21].

A new model was developed powerful to predict FAs distribution networks in various phases of ruminal microbiome in addition of/out of exogenous PUFAs after perturbation dealing within chemical molecular descriptors (V_k) and initial experimental boundary conditions (c_j). Each $'f(L_{nr})_{new}$ represents a corresponding coefficient in the new model for predicting $IPA(\%)_{new}$. This model can classify as high (L_{nr} = 1)/ low (L_{nr} = 0) of the expected values of FAs (LCFAs/VFAs) between the *new* and *reference* states after changed the boundary conditions (c_j). The parameter n(L_{nr} = 1) represents the number of cases in the sub-set with L_{nr} = 1 (links in complex network), or means the $IPA(\%)_{new}$ of *new* sub-set is higher than that of *reference* ($IPA(\%)_{ref}$). Meanwhile, n(L_{nr} = 0) represents the number of cases observed and predicted in the sub-set with L_{nr} = 0 (without connected nodes in complex networks) or implied that $IPA(\%)_{new}$ is lower than $IPA(\%)_{ref}$. The best PT-LFER model found using the LDA algorithm theory has only 12 independent variables and presented as following algorithm.

$$
\begin{aligned}
'f(L_{nr})_{new} = & -0.021\cdot f(\varepsilon_{ij})_{ref} + 0.0026\cdot\langle f(\varepsilon_{ij})\rangle_{ref} + 0.3713\cdot {}^{new}V_6 + 1.0709\cdot {}^{new}\omega 6 - 1.1264\cdot {}^{new}\omega 3 \\
& + 0.0237\cdot\Delta\Delta V_5(c_1) - 0.0063\cdot\Delta\Delta V_6(c_2) + 0.0044\cdot\Delta\Delta V_1(c_4) - 0.0037\cdot\Delta\Delta V_1(c_5) \\
& - 0.0036\cdot\Delta\Delta V_4(c_6) - 0.1682\cdot({}^{new}V_6)^2 + 0.0182\cdot(\Delta\Delta V_6(c_3))^2 - 13.7236 \\
& N = 407{,}655; \quad \chi^2 = 244{,}532.9; \qquad p < 0.005
\end{aligned}
\tag{10}
$$

The output function $'f(L_{nr})_{new}$ is useful to classify the pairs of states (pairs of nodes). Some input terms were expanded as follows. Like, $\Delta\Delta V_k(c_j) = p(c_j)_{new}\cdot\Delta V_k(c_j)_{new} - p(c_j)_{ref}\cdot\Delta V_k(c_j)_{ref}$. This can be further expanded in turn as $\Delta\Delta V_k(c_j) = p(c_j)_{new}\ ({}^{new}V_k - \langle V_k(c_j)\rangle_{new}) - p(c_j)_{ref}\ ({}^{ref}V_k - \langle V_k(c_j)\rangle_{ref})$, $\langle V_k(c_j)\rangle$ is the average of V_k for each c_j. The statistical parameters, such as specificity (Sp), sensitivity (Sn), and accuracy (Ac) are always used to evaluate a new model. For the present study, the best new model found predicted the effects of perturbations under the initial conditions (c_j) over FA distribution with Sn, Sp, and Ac greater than 80% for a total of = 407,655 cases in training and external validation series.

Additional machine learning techniques were used to do some artificial neural networks - linear and non-linear ANNs (LNNs and MLPs) for comparing them with LDA model found. The best 11 ANN models were found to compare with the best LDA classification. In present study, the LNN models are based on 8 to 12 variables and MLP models have 5 to 12 input variables. The best MLP model (MLP 12:12-11-1:1) has 12 input variables and only one hidden layer with 11 neurons with 93.73% of accuracy in training set and 92.54% in validation set. The PT-NLFER model obtained with MLP number 6 classified our dataset better than the LDA PT-LFER model. However, PT-LFER is notably simpler and shows a direct relationship between the input variables and the output. Thus, the LDA model has a better prediction capacity than all LNNs but less than MLPs. In addition, LDA model has the lower training and validation errors compared to all ANNs. In a word, MLP models were better problem solving capacity, but notably more complicated.

Network biology [22] is a very useful approach to shed light on the functional organization of the cell. Thus, the observed complex networks were built for perturbations in FA metabolism/distribution between ruminal media and microbial membrane. Two states are connected (L_{nr} = 1) for both IPA(%)$_{obs}$ ($f(\varepsilon_{ij})_{new}$) and IPA(%)$_{ref}$ ($f(\varepsilon_{ij})_{ref}$), if IPA(%)$_{obs}$ - IPA(%)$_{ref}$ > 0, and L_{nr} = 0 otherwise. It was considered that L_{nr} = 1 (lined nodes) if both values of '$f(\varepsilon_{ij})_{new}$ and '$f(\varepsilon_{ij})_{ref}$ predicted by new model have the probability $p(c_{ij}) > 0.5$ with $f(\varepsilon_{ij})_{ref}$ = IPA(%)$_{obs}$ - IPA(%)$_{ref}$ > 0.

Last, two random networks models (random network **1** and **2**) were built. To set each model with a number of nodes and links as similar as possible to the observed and predicted networks, respectively. The results showed that the average values of the topological distance, node degree and closeness are similar between the observed and predicted networks (1.83 *vs.* 1.77, 72.75 *vs.* 80.29, and 0.000755 *vs.* 0.000836, respectively).

4. Conclusions

Combined with experimental and computational methodology are useful to study the effect of multiple conditional factors over fatty acids distribution networks on ruminal microbiome and liquid phase. Meanwhile, PT and LFER ideas can be combined to develop a PT-LFER model on fatty acid distribution network. PT Operators and Box-Jenkins of physicochemical parameters are useful to define some inputs. ANN algorithms are also valuable to test the performance of alternative PT-NLFER; Non-Linear models. In final, ER random network models can be carried out the comparative studies with the observed and predicted networks referred to the effect of perturbations on the fatty acid distribution processes.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. S. L. Kronberg, E. J. Scholljegerdes, G. Barcelo-Coblijn and E. J. Murphy, *Lipids*, 2007, **42**, 1105-1111.
2. K. i. Ichihara and Y. Fukubayashi, *Journal of Lipid Research*, 2010, **51**, 635-640.
3. J. G. Kramer, V. Fellner, M. R. Dugan, F. Sauer, M. Mossoba and M. Yurawecz, *Lipids*, 1997, **32**, 1219-1228.
4. C. Hansch, A. R. Steward and J. Iwasa, *Molecular Pharmacology*, 1965, **1**, 87-92.
5. C. Hansch, W. E. Steinmetz, A. J. Leo, S. B. Mekapati, A. Kurup and D. Hoekman, *J Chem Inf Comput Sci*, 2003, **43**, 120-125.
6. H. Gonzalez-Diaz, S. Arrasate, A. Gomez-SanJuan, N. Sotomayor, E. Lete, L. Besada-Porto and J. M. Ruso, *Current topics in medicinal chemistry*, 2013, **13**, 1713-1741.

SciForum
Mol2Net

In Silico Design of New Drugs for Myeloid Leukemia Treatment

Washington Pereira and Ihosvany Camps *

Laboratório de Modelagem Computacional - LaModel, Instituto de Ciências Exatas - ICEx, Universidade Federal de Alfenas - UNIFAL-MG, Av. Jovino Fernandes Sales s/n, Bairro Santa Clara 37130-000. Alfenas. MG. Brazil

* Author to whom correspondence should be addressed; E-Mail: icamps@unifal-mg.edu.br;
Tel.: +55-35-3701-1963; Fax: +55-35-3299-1063.

Received: 5 October 2015 / Accepted: 5 October 2015 / Published: 2 December 2015

Abstract: In this work we use *in silico* tools like *de novo* drug design, molecular docking and absorption, distribution, metabolism and excretion (ADME) studies in order to develop new inhibitors for tyrosine-kinase protein (including its mutate forms) involved in myeloid leukemia disease. This disease is the first cancer directly associated with a genetic abnormality and is associated with hematopoietic stem cells that are manifested primarily with expansion myelopoiesis. Starting from a family of fragment and seeds from known reference drugs, a set of more than 6k molecules were generated. This first set was filtered using the Tanimoto similarity coefficient as criterion. The second set of more dissimilar molecules were then used in the docking and ADME studies. As a result, we obtain a group of molecule that inhibit the tyrosine-kinase family and have ADME properties better than the reference drugs used in the treatment of myeloid leukemia.

Keywords: tyrosine-kinase; de novo design; fragment docking; myeloid leukemia

1. Introduction

The Chronic Myeloid Leukemia (CML) was initially reported in 1825 in France. French literature describes an autopsy done on a 63 years old woman where a large increase in the spleen and liver size was found. The spleen, due to the disease, had been increased 20 times compared to the healthy spleen [1,2].

Another important step on the understanding of CML was given in 1960, with the identification of a chromosomal abnormality as the source of the disease. This anomaly is the subtle exchange of the ends of the chromosomes 9 and 22, respectively (referred to as ABL and BCR gene), thereby creating a new hybrid chromosome which has been assigned the name Philadelphia chromosome. This hybrid chromosome has the ability to start the disease by encouraging disordered cell division of some blood cells [3-7]. The tyrosine-kinase mechanism of BCR-ABL works by binding ATP (adenosine triphosphate) and transferring a phosphate group from ATP to tyrosine residues on various substrates [8,9]. Activation of these pathways can lead to lack of control of cell proliferation and apoptosis. The major drugs used to fight the disease use these mechanisms to inhibit tyrosine-kinase, choosing tyrosine-kinase as a perfect target for the study and design of new drugs [10].

2. Materials and Methods

The structure of the tyrosine-kinase in its wild form (without any mutation) was downloaded from the Protein Data Bank (PDB) with code 1OPJ and resolution equal to 1.75 Å [11]. To obtain the mutate forms, we apply the mutations following the codes used in the literature for the protein with code 3QRI [12].

Prior to docking studies, the proteins structure was prepared using the Protein Preparation Wizard protocol as implemented in the Schrödinger Suite [13] using the Maestro interface [14]. This protocol adds hydrogen atoms, corrects bonds, complete chains, etc.

The computational modelling of new inhibitors for the tyrosine-kinase was divided into three steps. In the first step, the *de novo* design was carried out using the LigBuilder software [15]. In this work we used the growing/linking modes and the explore mode. In the second step, the generated molecules were used as ligands in the docking studies with all the protein family. The docking studies were carried out using the Glide software [16] from the Schrödinger Suite [13]. To evaluate each pose, it uses an scoring function called GlideScore that consider the van der Waals energy, the Coulomb energy, a lipophilic contact term, an hydrogen-bonding term among other terms [17, 18]. The GlideScore (or GScore) has units of kcal/mol and the lower it is, the better the interaction is. All the docking simulations were performed considering the protein as a rigid structure and the ligand as flexible. In the last step, some physicochemical descriptors related to the Absorption, Distribution, Metabolism and Excretion (ADME) properties were calculated. In this work, we used the QikProp software [19].

3. Results and Discussion

Using the *de novo* design technic, a universe of more than 6000 molecules was obtained. To validate the structural diversity of the generated library we calculated a 2D linear hashed fingerprint with a 64-bit address space. Then, we used the Tanimoto metric to compute the similarity among all the molecules (if the Tanimoto coefficient of two structures is greater than 0.85, the structures are considered similar) [20-22]. In the second step of our methodology, we dock all the molecules into the active site of all the protein family (with and without mutations). To provide a comparative of the potential of the generated molecules, we did the molecular docking of the 4 reference drugs must used as tyrosine-kinase inhibitor: imatinib, dasatinib, nilotinib and ponatinib. Comparing the score of the reference drugs with the generated molecules presented in table 1, we can see that in all the cases (including the mutated proteins) there is more than one molecule with better scores, suggesting potentials tyrosine-kinase inhibitor stronger than the reference drugs.

From the results in table 1, we can see that among the whole population of molecules, the compounds **680**, **781** and **723** repeatedly appear as well ranked ligands. Especial attention for the structure **781** that have a higher score than the reference drugs in all cases.

A comparison of the docking poses of **781** and imatinib in the binding site of the T315I protein is shown in figure 1. The imatinib is interacting with T315I through 4 hydrogen bonds with amino acids Asp400, Ile379 and Met337 and Glu305, a π–cation interaction with His380 and 2 π–π interactions between Phe401 and Tyr272 residues and the imidazole ring of imatinib. These interactions are dispersed over the whole imatinib molecule. In the case of **781**, it mades 3 hydrogen bonds with residues Glu305, Tyr272 and Asn341 and a π–π interaction between the Tyr272 and the benzofuran ring. In this case the interactions are distributed over the whole molecule structure also.

A widely used descriptor to study the drugability of molecules is the Lipinski's rule of five [24]. It predicts that a molecule will have poor absorption if its molecular weight (MW) is greater than 500Da, the average estimated number of hydrogen bonds that would be accepted by the solute from water molecules (HBAcceptor) is greater than 10, the average estimated number of hydrogen bonds that would be donated by the solute to water molecules (HBDonor) is greater than 5 and its octanol/water partition coefficient (QPlogPo/w) is greater than 5 [24]. Another descriptor that it is important is the QPlogHERG that simulate the blockage of human ether-a-go-go hERG K+ channels.

From table 2 we can see that the reference drugs nilotinib and ponatinib violate the Lipinski's rule of five. Both have the molecular weight over 500Da and the ponatinib also have a partition coefficient out of the recommended values. On the other hand, our best molecules do not have any violations.

A special attention is needed for the predicted QPlogHERG descriptor. Recently, it has been found that several non-cardiac drugs inhibit the hERG K+ channel causing cardiac side effects. Among them we can mention sudden cardiac death, significant QT prolongation (period between the start of ventricular depolarization and repolarization) and life-threatening ventricular arrhythmia. These undesirable drug interactions make the drugs withdrawn from the market owing to cardiovascular toxicity associated to them [25]. The values of QPlogHERG are concerning when bellow −5 [19]. From our simulations, all the molecules have values lower than −5 but the reference grugs imatinib, nilotinib, dasatinib and ponatinib have the more high-risk values. As it can be found elsewhere, heart problems is a recurrent side-effect of these drugs.

The other compounds (**680**, **723** and **781**) also have the QPlogHERG bellow the recommended value but in a lower extend. In the case of compound **781**, it have the higher value (less risk) among all the studied molecules.

Table 1. Docking score (GScore) for the best molecules and for the references drugs.

Protein Mutation	Molecule: GScore	Imatinib	Dasatinib	Nilotinib	Ponatinib
1OPJ	**680**: -15.34	-13.96	-9.08	-13.63	-12.96
T315I	**781**: -13.57	-13.31	-7.22	-4.89	-11.92
T315A	**781**: -14.16	-13.05	-9.90	-13.49	-13.09
M244V	**723**: -14.95	-13.16	-10.40	-13.51	-13.19
E355G	**781**: -16.13	-10.22	-11,01	-13.58	-12.98
H396A	**781**: -15.82	-13.02	-9.69	-14.12	-13.68

Table 2. ADME properties (quantities out of recommendation values are underlined).

Compound	MW	QPlogPo/w	HBDonor[1]	HBAcceptor[1]	QPlogHERG
Imatinib	493.610	3.476	2	10.00	-9.280
Dasatinib	488.006	2.509	3	10.00	-6.672
Nilotinib	529.523	5.870	2	8.00	-8.246
Ponatinib	532.567	4.602	1	9.50	-9.243

680	487.511	1.856	5	10.00	-6.307
723	430.502	4.471	3	6.25	-8.392
781	459.498	4.96	3	6.75	-5.837

[1] As they are average values, they can be non-integers.

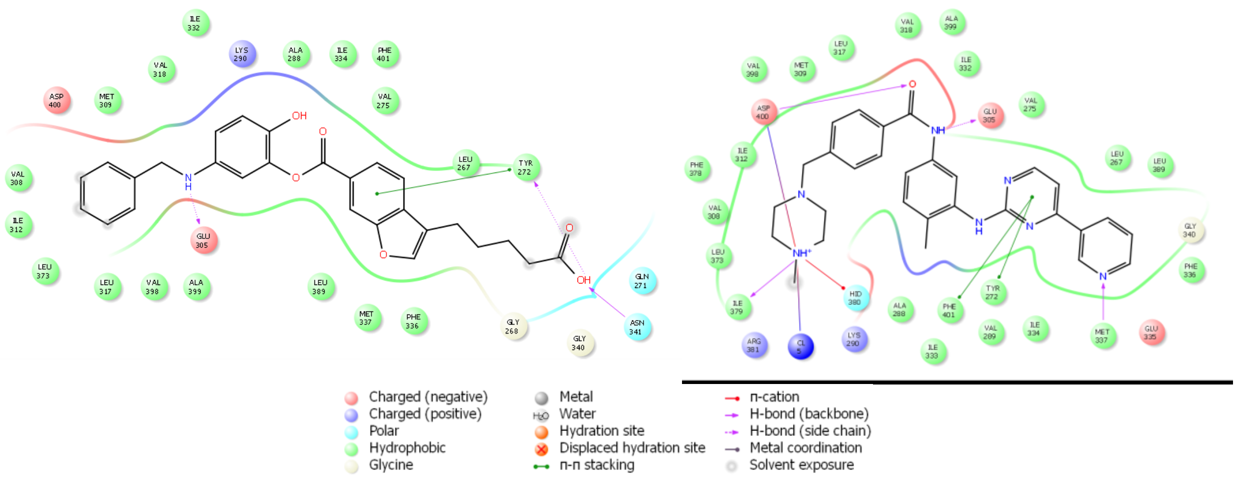


Figure 1. 2D interaction diagram between **781** (left) and imatinib (right) with mutation T315I.

4. Conclusions

The myeloid leukemia is a fatal disease, so it is of great importance to keep the patients in chronic phase where they stay asymptomatic. The fragment based drug design method used in this work turns to be a good alternative to create drugs that can control this neoplasm. Based on the calculated GScore, the de novo designed molecules have better inhibitor capacity than the tyrosine-kinase inhibitors most used in the market. These molecules shown strong potential to become drugs capable to inhibit all mutations, mainly the T315I mutation, now the leading cause of deaths due to the difficulty of inhibitors to control it.

Acknowledgments

We would like to acknowledge financial support from the Brazilian agency Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).Main text paragraph.

Author Contributions

IC idealize the experiments and prepare the images and the final manuscript. WP obtain the protein wild structure, produce the mutations and run the simulations.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. C. G. Geary, Br. J. Haematol. 110, 2 (2000).
2. A.-H. Maehle, Notes Rec. R. Soc. 65, 359 (2011).

3. T. Hunter, J. Clin. Invest. 117, 20362043 (2007).
4. C. M. Verfaillie, R. Bhatia, M. Steinbuch, T. DeFor, B. Hirsch, J. S. Miller, D. Weisdorf, and P. B. McGlave, Blood 92, 1820 (1998).
5. Y. Maru and O. N. Witte, Cell 67, 459 (1991).
6. O. Hantschel, Genes Cancer 3, 436 (2012).
7. B. Clarkson, A. Strife, D. Wisniewski, C. L. Lambek, and C. Liu, Leukemia 17, 1211 (2003).
8. H. long Xu, Z. jie Wang, X. meng Liang, X. Li, Z. Shi, N. Zhoua, and J. ku Bao, Mol. BioSyst. 10, 1524 (2014).
9. P. Coppo, I. Dusanter-Fourt, G. Millot, M. M. Nogueira, A. Dugray, M. L. Bonnet, M. T. Mitjavila-Garcia, D. L. Pesteur, F. Guilhot, W. Vainchenker, F. Sainteny and A. G. Turhan, Oncogene 22, 4102 (2003).
10. Y. Zhu and S.-X. Qian, Onco Targets Ther. 7, 395 (2014).
11. PDB ID: 1OPJ. B. Nagar, O. Hantschel, M. A. Young, K. Scheffzek, D. Veach, W. Bornmann, B. Clarkson, G. Superti-Furga, and J. Kuriyan, Cell 112, 859 (2003).
12. PDB ID: 3QRI. Wayne W. Chan, S. C. Wise, M. D. Kaufman, Y. M. Ahn, C. L. Ensinger, T. Haack, M. M. Hood, J. Jones, J. W. Lord, W. P. Lu, D. Miller, W. C. Patt, B. D. Smith, P. A. Petillo, T. J. Rutkoski, H. Telikepalli, L. Vogeti, T. Yao, L. Chun, R. Clark, P. Evangelista, L. C. Gavrilescu, K. Lazarides, V. M. Zaleskas, L. J. Stewart, R. A. V. Etten, and D. L. Flynn, Cancer Cell 19, 556 (2011).
13. Schrödinger suite: http://www.schrodinger.com/
14. Maestro, version 10.1, Schrödinger, LLC, New York, NY, 2015.
15. Y. Yuan, J. Pei, and L. Lai, J. Chem. Inf. Model. 51, 1083 (2011).
16. Glide, version 5.0, Schrödinger, LLC, New York, NY, 2008.
17. W. Sherman, T. Day, M. P. Jacobson, R. A. Friesner, and R. Farid, J. Med. Chem. 49, 534 (2006).
18. R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin, and D. T. Mainz, J. Med. Chem. 49, 6177 (2006).
19. QikProp, version 3.2, Schrödinger, LLC, New York, NY, 2009.
20. N. Nikolova and J. Jaworska, QSAR Comb. Sci. 22, 1006 (2003).
21. J. Bajorath, ed., Chemoinformatics. Concepts, Methods, and Tools for Drug Dis- covery (Humana Press, Totowa, New Jersey, 2004).
22. G. Maggiora, M. Vogt, D. Stumpfe, and J. Bajorath, J. Med. Chem. 57, 3186 (2014).
23. M. T. Delamain and M. Conchon, Rev. Bras. Hematol. Hemoter. 30, 37 (2008).
24. C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, Adv. Drug Delivery Rev. 46, 3 (2001).
25. T. Langer and R. D. Hoffmann, Pharmacophores and Pharmacophore Searches (Wiley-VCH, 2006)

SciForum
Mol2Net

Interdependence of Influenza HA and NA and Possibilities of New Reassortments

Ashesh Nandy [1,*] and Subhas Basak [2]

[1] Centre for Interdisciplinary Research and Education, 404B Jodhpur Park, Kolkata 700068, INDIA

[2] University of Minnesota Duluth-Natural Resources Research Institute (UMD-NRRI) and Department of Chemistry & Biochemistry, University of Minnesota Duluth, Duluth, MN 55811, USA

* Author to whom correspondence should be addressed; E-Mail: anandy43@yahoo.com;
Tel.: +91-94335-79452.

Received: 28 October 2015 / Accepted: 30 October 2015 / Published: 2 December 2015

Abstract: The influenza virion is characterized by two surface proteins, hemagglutinin (HA) and neuraminidase (NA). The changes in their surface antigenic sites have given rise to several subtypes – H1 to H16 for the hemagglutinin and N1 to N9 for the neuraminidase, and each influenza strain is identified with these subtypes such as the H5N1, H7N9, etc. Of the 16 x 9 combinations possible, only certain combinations are observed to proliferate in the wild, such as the H1N1, H3N2, H5N1, etc. This interdependence of the HA and NA on certain subtypes have been noticed, and experimentally demonstrated, but the underlying cause or its systematics have been unknown. We have hypothesized that the base distribution characteristics of the HA and NA constitute a coupling between them. We estimate the coupling strength by measuring the distance in graph radii between the two genes in a graphical representation scheme. We found that this distance was characteristic of each subtype combination and forced combinations with a different HA or NA subtype led to widely different values, which by our hypothesis, and the experimental findings of Zhang et al, implied unstable combinations. This hypothesis implies that given a stable subtype of pathogenic influenza, we can estimate using the coupling constants which other subtype combinations could emerge through reassortment. Thus in the case of the H5N2 strain which had an epidemic form in North America in 2015, we have calculated the consequences of altering the NA component. We found that only H5N4, H5N6 and H5N9 combinations could match the coupling strength of the H5N2, thus implying that the next epidemics could arise from these combinations rather than other subtypes of H5. This allows for more focused monitoring of emerging flu strains for epidemic potential.

.

Keywords: influenza HA-NA coupling strengths; bird flu, neuraminidase;
interdependence; new subtypes; HA-NA hemagglutinin

Mol2Net YouTube channel:
http://bit.do/mol2net-tube
YouTube link:
https://www.youtube.com/watch?v=vOyKJxv-IVo

1. Introduction

Influenza is a widely prevalent seasonal viral infection that afflicts several million people annually with fatalities that range into tens of thousands. The influenza virus is of three main types, Influenza A, B and C, of which Influenza A is the most prevalent and affects birds, mammals and humans. The influenza A virus genome is a segmented negative strand RNA virus with eight distinct pieces of the nucleic acid coding for 11 proteins. Two of these - hemagglutinin (HA) and neuraminidase (NA) - are surface situated glycoproteins responsible for virus endocytosis and progeny elution. On the basis of their antigenic segments several subtypes of these proteins have been identified – 16 for the HA and 9 for the NA, enabling the viral strains to be characterized as H1N1, H3N2, H7N9, etc.

It has been observed that while 16 x 9 combinations of the H*x*N*y* variety of viral strains are possible, only certain combinations seem to predominate in the wild [1], as can be seen from the database population of the various flu strains [2]. The reason for this interdependence of one subtype of HA for a specific subtype of NA and vice versa is not known, but that such an interdependence exists has been verified by the experiments of Zhang et al [3] who swapped different subtypes to show that outside the preferred subtype combinations the strains were not stable or adequately infective; this has also been observed in other subtype combinations [4].

This issue is quite critical. The evolution of different strains of influenza leads to the possibilities of emergence of epidemic and pandemic strains that can affect very large number of host species as had happened with the Spanish Flu (subsequently identified as a H1N1 subtype) of 1918 where the human death toll exceeded 20 million, the Swine Flu (H1N1) of 2009 where several million people were infected and over 25,000 died within one year, the H5N1 bird flu that apparently surfaced around 1997 and the H7N9 flu of 2013 whose containment necessitated culling of millions of chicken, and the recent H5N2 avian flu epidemic in North America (2015) that has led to culling of millions of poultry and farm birds. Such strains arise through genetic shift and drift, of which reassortments among the various genes, especially the glycoproteins, which can occur when two flu strains infect a single cell, are among the most important processes [5,6]. Such possibilities require incessant monitoring of evolving subtypes and combinations worldwide, a rather daunting, but inescapable task. Taking interdependence of HA and NA into account can help to reduce this task to more manageable proportions since monitoring one of them, say HA, automatically accounts for the associated NA that can retain the infectivity. Quantifying the interdependence of HA and NA therefore can be instrumental in this enterprise.

Our study of this HA-NA interdependence [2] led us to hypothesize a coupling between the two glycoproteins arising out of their base distribution and composition characteristics which are best visualized in a graphical representation. Our research showed that the major influenza subtype combinations had very distinct coupling strengths and interchange between different subtype components compromised such strengths, an outcome in keeping with Zhang et al's experimental results [3]. As a consequence of this model, we considered the current H5N2 avian epidemic in the USA and could forecast two possible reassortment products that could conceivably fuel new epidemics [7]. This report very briefly summarizes our methodology and these results and observations.

.

2. Results and Discussion

Graphical representation of the HA and NA sequences of H1N1, as described in the Methods section and shown in Fig.1, depicts the base distributions of the two genes. Our hypothesis of a coupling between the two genes is predicated upon the assumption that the base distributions are mutually dependent; indeed, Hu [8] has shown that mutational changes in one lead to a coupled mutational change in the other. We quantify this inter-dependence by the distance between the end-points of the graph radii of the two plots as defined in the Methods section.

The point is that if the coupling between the HA and the NA were to be characteristic of the specific subtype, implying co-ordinated change of some kind, then irrespective of the genetic shifts in the two gene sequences, the coupling factor as measured by the distances of the graph radii should remain constant within a reasonable limit. As shown in Table 1, taking a large sample of the H1N1 strains we found that the coupling factor worked out to 39.49±8.19. Similar analyses for other strains of recent interest showed similar trends. Table 1 lists the results of our analysis with all viral strains in our database showing the coupling actors for each individual viral strain and also an average for each flu subtype where adequate number of strains was available. We notice that this average is different for each subtype with a reasonably low standard deviation implying that each flu subtype has a specific coupling strength, which we may refer to as its characteristic value.

For the coupling to be characteristic of each flu subtype, replacing one gene with another variety should produce quite different result for the coupling factor. In fact, as we replaced the H1 sequence of A/South Dakota/01/2011(H1N1) with a H5 sequence from A/duck/France/05066b/2005(H5N1), the coupling factor changed from 33.97 to 12.03; exchanging only the NA between the two strains changed the coupling factor to 69.72 implying gross lack of compatibility between the HAs and NAs of the two strains. This effect we found in a wide variety of samples tested as shown in Table 2. We note that the HA exchange produced less dramatic or insignificant effects than NA since the HA sequences are more homologous across all HA subtypes compared to the NA subtypes: taking typical examples each of all subtypes of HA and NA, we find that in terms of the composition of the four bases a, c, g, t the standard deviation from the average composition values is <3% for a and t (a: 2.6%, t: 2.8%) and <4% for g and c (g:3.7%, c:3.8%) for the HA, whereas these figures are >3% for a,t (a:3.7%, t:9.2%) and >6% (g:6.7%, c:6.2%) for the NA; the wide variation of the NA and the comparatively lesser variation of the HA subtypes is evident too in the graphical representations shown in Ref 2.

These results of our analyses, summarized in Tables 1 and 2, show that forced exchanges between the HA and NA of the flu strains often lead to coupling factors widely different from the characteristic values. Our observations tie in neatly with the experimental results of Zhang et al [3] who found from HA, NA exchanges within a set of 1918 pandemic H1N1, a 2009 pandemic H1N1 and a HPAI H5N1 that the NA exchanges led to significant decline in influenza infectivity whereas the effect was comparatively much less when the HAs were exchanged. This he attributed to lack of "matching patterns" between the NA of the H5N1 with the HAs of the H1N1s in the experiment. From the observation noted earlier that the availability in the wild of only a few wild subtypes of flu may imply low stability of other subtypes, and the observations of wide variation in the computed coupling value and Zhang et al's results, we may infer that such "forced" subtypes will not yield stable or efficient infective strains.

This leads us to an interesting prognosis. This year has witnessed a sudden epidemic of highly pathogenic avian influenza (HPAI) H5N2 infection among poultry and farm raised birds like chicken and turkeys in the North American west and mid-west leading to death through infection or culling of millions of birds [9]. While the virus has not affected humans yet, strict monitoring is being done to ensure adequate warning in case the virus develops human-to-human transmission ability [10]. At the same time a watch has to be kept on the possibility that the virus could undergo reassortments and give rise to new subtypes, though one does not know which of the possible subtypes could be highly pathogenic too.

Our analysis provides a guideline here. Once we know that the H5N2 is a HPAI virus, we can forecast which of the possible reassortants have the potential to be stable and possess pathogenic ability [7]; it is pertinent to note here that according to the USDA, the current H5N2 subtype is itself a combination of Asian HPAI H5 and North American N2 [11,12]. We accessed all North American H5N2 gene sequences available at the time, i.e., around mid-May 2015, and determined that the magnitude of the coupling for these strains as measured through the delta-g_R was 38.58±1.46. To assess possible reassortments from these strains we are mindful of the fact that Asian H5 is a highly pathogenic virus that in its H5N1 bird flu form had caused high level of human fatalities at a mortality rate of 1 per 2 infections [13], and a continuing fear that the virus may mutate to a form causing human-to-human transmission leading to a new pandemic [14]. Taking such a HA as one component of the possible reassortants, we tried combinations with all subtypes of the NA available from typical flu sequences (Table 3). Taking a cue from the results given in Table 1 that the standard deviations of the coupling values between the various strains of the flu subtypes is 18.85% (range: 7.65% – 29.12%), we can look for those HA-NA combinations that lie within this range. The results as shown in Table 3 indicate that only combinations of the H5 with a possible N4, N6 or N9 fall within this coupling value range and therefore could be the new HPAI to evolve from reassortments of the H5N2 with other flu subtypes. Our research showed that flu subtypes with these varieties of the neuraminidase have been reported in various places in North America, indicating that it is possible for reassortments of the HPAI H5N2 with these subtypes to take place. While monitoring for genetic shifts and drifts of the H5N2 in North America, close attention, therefore, may be given to development of H5N4, H5N6 and H5N9, if any, of which H5N9 may bear extra scrutiny since a hitherto benign to human H7N9 strain in China suddenly developed a mutation in 2013 that led to human fatalities. Such focused scrutiny might reduce the monitoring overhead to some extent to concentrate on the more potent possibilities. The same exercise can be done with other HA subtypes, but as we have seen, the flu subtypes are more sensitive to the changes in NA.

Table 1. Coupling factors of HA-NA interdependence. The last two columns provide summary data for each major flu type indicated in the first column

Flu Type	Locus ID		Virus Strain Description	Year	gR values		Coupling	Group	
	HA	**NA**			NA	HA	factor η	Averages	Std Dev
H1N1	CY016699	CY016701	A/South Australia/58/2005(H1N1)	2005	113.6024	100.9997	25.37991		
	GQ150342	CY039988	A/Nonthaburi/102/2009(H1N1)	2009	95.58367	118.4997	32.45575		
	KC881952	KC881951	A/South Dakota/01/2011(H1N1)	2011	95.53349	120.0206	33.97239		
	CY039893	CY039895	A/New York/1669/2009(H1N1)	2009	95.73236	120.473	35.51551		
	FJ998208	FJ998214	A/Mexico/InDRE4487/2009(H1N1)	2009	94.97363	120.473	36.33877		
	CY039999	CY040001	A/New York/3008/2009(H1N1)	2009	95.49054	120.9794	36.44679		
	CY039901	CY039903	A/New York/1682/2009(H1N1)	2009	95.45042	120.8896	36.68635		
	KF648252	KF648260	A/Washington/05/2013(H1N1)	2013	93.2853	118.2515	36.95131		
	CY040007	CY040009	A/New York/3012/2009(H1N1)	2009	95.49054	121.9571	37.39781		
	GQ338364	GQ117071	A/Minnesota/02/2009(H1N1)	2009	94.78558	122.4494	38.14657		
	CY148235	CY039528	A/Netherlands/602/2009(H1N1)	2009	92.11382	120.1784	38.49428		
	FJ966952	FJ966956	A/California/05/2009(H1N1)	2009	93.60996	121.7123	38.42935		
	KC781785	GQ377078	A/California/07/2009(H1N1)	2009	94.87404	121.6822	38.93651		
	CY134351	CY134353	A/green-winged teal/California/123/2012(H1N1)	2012	91.47518	83.26524	40.19993		
	CY134359	CY134361	A/northern shoveler/California/138/2012(H1N1)	2012	93.76871	77.66515	33.03013		
	CY134367	CY134369	A/northern pintail/California/183/2012(H1N1)	2012	94.10496	76.34555	34.85523		
	CY077076	CY077078	A/mallard/Sanjiang/390/2007(H1N1)	2007	85.39725	96.98963	55.31523		
	EU026037	EU026039	A/mallard/Maryland/170/2002(H1N1)	2002	89.38879	76.97491	36.1104		
	EU743306	EU743308	A/blue winged teal/LA/B228/1986(H1N1)	1986	87.14094	83.45928	33.5418		
	AB546149	AB546151	A/pintail/Aomori/422/2007(H1N1)	2007	80.85941	88.08622	47.88056		
	AB670330	AB472014	A/duck/Tsukuba/718/2005(H1N1	2005	83.23725	91.84633	54.27835		
	HM193551	HM193628	A/mallard/Alaska/44430-088/2008(H1N1)	2008	91.17319	78.89723	41.29855		
	FJ432778	FJ432780	A/goose/Italy/296426/2003(H1N1)	2003	84.48628	95.62911	54.02458		
	CY014627	CY005686	A/duck/AUS/749/1980(H1N1)	1980	81.72422	92.90199	21.36802		
	FJ536818	FJ536824	A/swine/Shandong/443/2008(H1N1)	2008	88.04054	113.9492	50.04761		
	AB741039	AB741041	A/swine/Narita/aq21/2011(H1N1)	2011	95.94873	122.2932	36.65832		
	GQ229357	GQ229362	A/swine/Hong Kong/9656/2001(H1N1)	2001	109.2822	128.8602	42.5942		
	EU004444	EU004442	A/swine/Tianjin/01/04(H1N1)	2004	105.3392	106.7129	46.18326		
	CY085774	CY085776	A/swine/Hong Kong/434/2006(H1N1)	2006	99.01507	119.5231	52.64454	39.48903	8.187229
H1N2	JX069105	JX069107	A/ostrich/South Africa/AI2887/2011(H1N2)	2011	76.81041	86.06502	17.21649		
	AB741007	AB741009	A/swine/Tochigi/2/2011(H1N2)	2011	91.94878	113.8709	22.75411		
	CY133290	CY133292	A/mallard/Mississippi/10OS4593/2010(H1N2)	2010	82.69537	79.35445	33.5879		
H3N1	CY005943	CY004711	A/mallard duck/ALB/26/1976(H3N1)	1976	87.08971	76.95153	63.14924		
H3N2	CY006467	CY006469	A/New York/516/1997(H3N2)	1997	86.58321	91.76747	21.82011		
	CY091261	CY091263	A/Singapore/NHRC0007/2003(H3N2)	2003	83.2928	96.42491	22.33844		
	AB295605	AB295606	A/Aichi/2/1968(H3N2)	1968	77.86263	63.87713	27.37842		
	CY092233	CY092235	A/Australia/NHRC0010/2005(H3N2)	2005	85.46052	101.1324	27.38328		
	CY116638	CY116639	A/Tbilisi/GNCDC0557/2012(H3N2)	2012	88.55144	103.2946	28.25711		
	CY105870	CY105872	A/KhanhHoa/KH475/2008(H3N2)	2008	84.56058	106.1538	30.82751		

	CY105886	CY105888	A/HaNoi/311/2005(H3N2)	2005	87.52267	105.8014	31.34143		
	AB741031	AB741033	A/swine/Yokohama/aq138/2011(H3N2)	2011	90.78071	97.3364	13.3909		
	CY131029	CY131031	A/swine/Ohio/10SW136/2010(H3N2)	2010	84.88113	98.66828	16.12007		
	AB277754	AB277755	A/duck/Hokkaido/5/1977(H3N2)	1977	79.99952	67.8499	31.90895		
	KC422461	KC422462	A/feline/Guangdong/1/2011(H3N2)	2011	81.79576	94.96184	37.39615		
	KF155145	KF155147	A/canine/Korea/MV1/2012(H3N2)	2012	81.43785	95.31512	37.58191		
	KC422456	KC422458	A/feline/Korea/01/2010(H3N2)	2010	83.5886	92.22999	41.06222	28.21588	8.21681
H5N1	AB212054	AB212056	A/Hong Kong/213/2003(H5N1)	2003	82.66458	110.4352	48.87837		
	GU052518	GU052520	A/chicken/Scotland/1959(H5N1)	1959	86.35189	92.52505	33.31591		
	CY053325	JF758823	A/wood duck/Ohio/623/2004(H5N1)	2001	94.39676	112.5891	36.4933		
	AJ971297	AJ972921	A/duck/France/05066b/2005(H5N1)	2005	94.821	85.81665	43.58987		
	AF144305	AF144304	A/goose/Guangdong/1/1996(H5N1)	1996	76.56076	102.1669	44.23744		
	EF607855	EF607897	A/mute swan/MI/451072-2/2006(H5N1)	2006	87.95686	115.3977	45.63799		
	AF364334	AF364335	A/Goose/Guangdong/3/97(H5N1)	1997	74.52595	101.9541	47.47952		
	KC815853	KC815851	A/mallard/Italy/3401/2005(H5N1)	2005	83.11546	99.35829	52.03776		
	AY585373	AY585404	A/duck/Guangdong/07/2000(H5N1)	2000	82.8014	105.5828	52.64266		
	GU052073	GU052075	A/Goose/Hong Kong/3014.5/2000(H5N1)	2000	92.07489	107.368	53.95241		
	AY747617	AY747618	A/swine/Fujian/F1/2001(H5N1)	2001	86.5242	103.6867	56.02981	46.7541	7.134044
H5N8	CY134101	CY134103	A/mallard/California/2559P/2011(H5N8)	2011	76.93981	112.062	56.45905		
H6N1	AB298279	AB298280	A/duck/Hokkaido/W159/2006(H6N1)	2006	87.5451	102.213	24.3545		
	HM144392	HM144562	A/duck/Hunan/177/2005(H6N1)	2005	85.58079	90.42478	24.70176		
	HM144388	HM144558	A/mallard/Jiangxi/227/2003(H6N1)	2003	83.2108	90.20036	30.04321		
	EF681878	EF681880	A/chicken/Taiwan/2838V/00(H6N1)	2000	89.26956	92.06014	31.88554		
	AB294215	AB294216	A/duck/Hong Kong/716/1979(H6N1)	1979	78.10185	95.91201	33.8106		
	GQ414872	GQ414903	A/spot-billed duck/Korea/545/2008(H6N1)	2008	89.70904	92.77493	40.21473		
	GQ414864	GQ414904	A/mallard/Korea/L08-8/2008(H6N1)	2008	89.54615	94.47328	42.40703	32.4882	6.987585
H7N3	CY125730	CY125728	A/Mexico/InDRE7218/2012(H7N3)	2012	86.18209	73.54771	33.07789		
H7N9	KC853228	KC853231	A/Shanghai/4664T/2013(H7N9)	2013	103.7236	91.82757	38.80746		
	KC885956	KC885958	A/Zhejiang/DTID-ZJU01/2013(H7N9)	2013	106.4451	93.09385	40.50441		
	KC853766	KC853765	A/Hangzhou/1/2013(H7N9)	2013	105.9556	93.48328	41.09839		
	KC994453	KC994454	A/Fujian/1/2013(H7N9)	2013	108.0189	93.09385	41.31705		
	KF420298	KF420296	A/Changsha/1/2013(H7N9)	2013	106.6576	93.71526	42.77973		
	KF278746	KF226113	A/Jiangsu/1/2013(H7N9)	2013	106.8237	92.96191	43.35736		
	KF469231	KF261988	A/Nanchang/1/2013(H7N9)	2013	106.6445	93.71526	42.52465		
	KC899669	KC899671	A/chicken/Zhejiang/DTID-ZJU01/2013(H7N9)	2013	106.2791	85.76458	38.11029		
	AY999981	KF695256	A/Mallard/Sweden/91/02(H7N9)	2002	104.6166	72.43222	40.03544		
	GU060482	GU060484	A/goose/Czech Republic/1848-K9/2009(H7N9)	2009	98.69629	74.68026	43.06948		
	CY133649	CY133651	A/northern shoverl/Mississippi/11OS145/2011(H7N9)	2011	111.8206	69.69097	63.20246		
	CY067678	CY067680	A/blue-winged teal/Guatemala/CIP049-02/2008(H7N9)	2008	113.4246	72.32552	64.78273		
	AB481213	AB481212	A/duck/Mongolia/119/2008(H7N9)	2008	120.3336	79.60081	50.0517		
	KJ508892	KJ508890	A/tree sparrow/Shanghai/01/2013(H7N9)	2013	106.631	91.28917	41.58104	45.0873	8.488588

H9N2	JQ440373	JQ916910	A/chicken/Egypt/114940v/2011(H9N2)	2011	80.29504	113.7064	40.8703		
mixed	CY103245	CY103247	A/mallard/Alberta/115/2007(mixed)	2007	89.43659	82.58275	21.97967		
H5N1 short stalk strains:									
H5N1	AB239125	AB239126	A/Hanoi/30408/2005(H5N1)	2005	68.38075	112.7598	59.9053		
	DQ371928	EU128239	A/Anhui/1/2005(H5N1)	2005	75.77015	112.8584	69.38676		
	CY098681	CY098683	A/Anhui/1/2007(H5N1)	2007	73.62005	112.5678	66.89482		
	DQ835313	DQ835315	A/China/GD01/2006(H5N1)	2006	73.4652	115.0801	72.79483		
	AB478025	AB478033	A/quail/Thanatpin/2283/2007(H5N1)	2007	73.76582	113.7522	63.76267		
	AB780494	AB780496	A/duck/Vietnam/OIE-2212/2012(H5N1)	2012	80.98051	118.7173	70.41153		
	KC261463	KC261473	A/duck/Eastern China/057/2007(H5N1)	2007	76.04691	116.5277	71.06516		
	FR687258	FR687259	A/chicken/Egypt/Q1182/2010(H5N1)	2010	69.47943	113.5217	76.67715	68.86228	5.266735

(Reproduced by permission from Current Comput Aided Drug Design, Ref 2)

Table 2. Coupling factors for forced matches between HA and NA for different strains and subtype combinations

Flu SubTypes	Virus strain description	Natural coupling factor*	Computed coupling factor for first strain		Notes
			HA2+NA1	HA1+NA2	
H1N1 H3N2	A/Mexico/InDRE4487/2009(H1N1) A/Aichi/2/1968(H3N2)	36.338769 27.378422	39.4250615	49.7303551	
H1N1 H5N1	A/New York/1669/2009(H1N1) A/chicken/Egypt/Q1182/2010(H5N1)	30.403674 76.677153	31.0935325	82.399158	1
H1N1 H5N1	A/California/05/2009(H1N1) A/Anhui/1/2005(H5N1)	38.429353 69.386759	34.7546642	73.6664097	1 3
H1N1 H5N1	A/South Dakota/01/2011(H1N1) A/duck/France/05066b/2005(H5N1)	33.972387 43.589874	12.0340407	69.723639	2
H1N1 H6N1	A/Netherlands/602/2009(H1N1) A/mallard/Jiangxi/227/2003(H6N1)	38.494277 30.043214	2.74553914	67.1634074	
H1N1 H7N9	(A/New York/3008/2009(H1N1)) A/Fujian/1/2013(H7N9)	36.446792 41.317053	18.7330372	40.9006368	
H3N2 H5N1	A/Tbilisi/GNCDC0557/2012(H3N2) A/duck/France/05066b/2005(H5N1)	28.257107 43.589874	8.37696138	62.7612626	2
H3N2 H6N1	A/Australia/NHRC0010/2005(H3N2) A/duck/Hunan/177/2005(H6N1)	27.383275 24.701758	11.0170769	59.7319878	
H3N2 H7N9	A/Singapore/NHRC0007/2003(H3N2) A/blue-winged teal/Guatemala/CIP049-02/2008(H7N9)	22.338436 64.782725	12.262058	34.451943	
H5N1 H6N1	A/goose/Guangdong/1/1996(H5N1) A/chicken/Taiwan/2838V/00(H6N1)	44.237442 31.885536	20.1662511	58.710981	2
H5N1 H7N9	A/swine/Fujian/F1/2001(H5N1) A/Shanghai/4664T/2013(H7N9)	56.029805 38.807458	43.2469843	30.6110415	2
H6N1 H7N9	A/duck/Hunan/177/2005(H6N1) A/Zhejiang/DTID-ZJU01/2013(H7N9)	24.701758 40.504412	44.7498079	61.0380129	

Notes: * Values as per Table 1
1 H5N1 with short stalk neuraminidase
2 H5N1 with long stalk neuraminidase
3 Strains used in Zhang et al's (2010) experiments

(Reproduced by permission from Current Comput Aided Drug Design, Ref 2)

Table 3. Hypothetical combinations from 2015 North American H5N2: Coupling a H5 with different NA subtypes.

Gene	Locus ID – typical examples	Typical subtype / averages for avian ‘flus	Average Seq Length	Average Mu X	Average Mu Y	Graph radius g_R	Delta g_R for sample HA with other NAs	Diff from Average in Table 1 (in %)
Sample HA of H5N2	KP739389	A/domestic duck/Washington/61-16/2014(H5N2)	1704	-97.9695	-29.4883	102.31118		
NA N1 H1N1	EU743308	Average for H1N1 subtype	1410	-54.4152	-65.8155	85.39725	56.71546	47.02
NA N1 H6N1	HM144562	Average for H6N1 subtype	1410	-55.4681	-65.6722	85.962449	55.81799	44.69
NA N1 ls H5N1	KC815851	Average for H5N1 long stalk subtype	1410	-58.4532	-60.6701	84.247486	50.33731	30.48
NA N1 ss H5N1	KC261473	Average for H5N1 short stalk subtype	1350	-45.6463	-59.3848	74.900875	60.26209	56.21
NA N2 H3N2	AB277755	Average for N2 data from H1N2 and H3N2	1413	-63.7702	-47.3914	79.451709	38.60201	0.06
NA N3 H7N3	CY125730	A/Mexico/InDRE7218/2012(H7N3)	1410	-86.1333	-2.89858	86.182091	29.10508	24.55
NA N4 H4N4	AY207533	A/gray teal/Australia/2/79(H4N4)	1413	-56.1826	-30.2852	63.825366	41.79449	8.34
NA N5 H6N5	FLASHEAU	A/shearwater/Australia/1/1972(H6N5)	1407	-56.2971	-57.3163	80.340011	50.10975	29.89
NA N6 H4N6	JX454731	A/wild bird/Korea/GS26/2006(H4N6)	1413	-87.0035	2.248408	87.032586	33.57779	12.96
NA N7 H10N7	EU747332	A/quail/California/1022/1999(H10N7)	1347	-78.66	-15.3846	80.150346	23.91174	38.02
NA N8 H2N8	KC899750	A/wild duck/SH38-26/2010(H2N8)	1413	-42.2151	-48.0602	63.967938	58.76616	52.33
NA N9 H7N9	KC899669	Average for H7N9 subtype	1409	-108.696	3.491663	108.75254	34.6806	10.10

Note: ss - short stalk; ls - long stalk; Delta-g_R is the coupling factor

(Reproduced by permission from Current Comput Aided Drug Design, Ref 7)

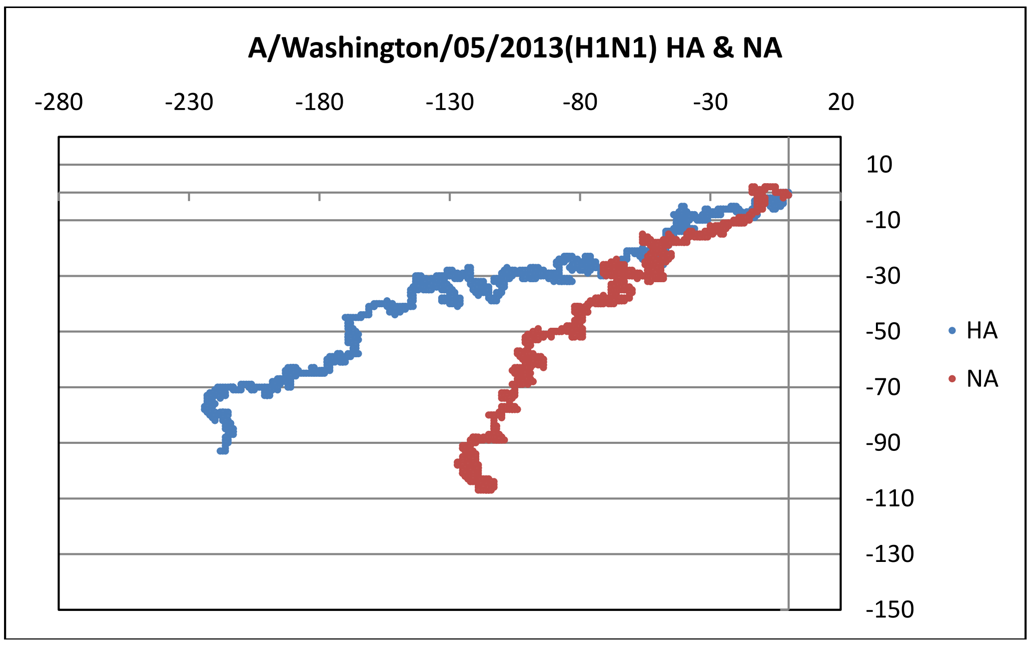


Figure 1. 2D graphical representation of HA (blue) and NA (red) sequences (cds) of A/Washington/05/2013(H1N1)

3. Materials and Methods

All sequences were downloaded from the NCBI GenBank database within the past year [15]. The IDs of the various sequences are given in the tables in Refs 2 and 7 and the details can be accessed from the NCBI website, Ref.11.

The 2D graphical representation method used here [16] is a simple device to visualize the base distribution of any DNA/RNA sequence. On a 2-dimensioanl Cartesian cor-ordinate system, the four cardinal directions are identified with the four bases which preferentially are: adenine with the –ve x-direction, cytosine with the positive y-direction, guanine with the +ve x-direction and thymine with the –ve y-direction. The query sequence is plotted starting from the origin and moving one step in the direction indicated for the base sequentially. This traces out a curve that reflects the distribution of bases along the sequence. Fig.1 is an example of two gene sequences, of the HA and NA, on the same graph.

Quantitative assessments of the different sequences, e.g., descriptors of the two sequences in Fig.1, can be made as a first approximation by defining weighted centre of mass (μ_x,μ_y) of a sequence as

$$\mu_x = {\sum_{i=1}^{N} x_i}/_{N} \;,\; \mu_y = {\sum_{i=1}^{N} y_i}/_{N}.$$

$$g_R = \sqrt{\mu_x^2 + \mu_y^2}$$

where (x_i,y_i) represent the co=ordinates of the *i*th base, *N* is the total number of bases and we define the distance from the origin to the centre of mass as g_R. Then the difference between two sequences can be represented by the distance between the end points of the graph radii of the two sequences as

$$\Delta g_R = \sqrt{(\mu_x + \mu_{x'})^2 + (\mu_y + \mu_{y'})^2}$$

We use the Δg_R as an indicator of the coupling between the two sequences as explained earlier. More discussions on the properties of g_R and Δg_R can be found in the earlier papers and related documents.

4. Conclusions

In this brief report we have discussed the observation of interdependence of hemagglutinin and neuraminidase in influenza A subtypes that appear to restrict the proliferation of influenza subtypes to a few combinations, although theoretically a much larger number should be possible. We believe the origins of this phenomenon must lie in the base distribution and composition of the related sequences; observations of Hu [8] on HA, NA mutations show that mutations in one sequence appear to regulate mutational changes in the other. To quantify this phenomenon we have hypothesized a coupling of the two sequences with a coupling factor that is characteristic of the related subtypes. Our investigations into real sequences of several different subtypes using graphical representation techniques yielded specific numbers for each flu subtype within a reasonable tolerance level; forced replacements of one gene with another subtype led to different coupling strengths, which was more dramatic in the case of NA exchanges [2].

Since the influenza genome is known to undergo genetic drift to new reassortants quite frequently due to its inherent segmented structure, this interdependence of the HA and NA serves to restrict such reassortants to a reduced subset of possible stable pathogenic varieties. Our methodology described above, and reported previously in Ref 2, allows us to compute possible such subtypes of the influenza. In response to the recent epidemic of H5N2 influenza among North American poultry, we have made a prognosis on this basis of possible new pathogenic reassortants that may arise out of the current epidemic [7]. This has important consequences on monitoring of influenza strains and mutations that allows opportunity to focus on possible more pathogenic subtypes. On a larger scale, our approach provides an opportunity worldwide to compute and monitor evolution of highly pathogenic influenza viruses.

.

Acknowledgments

The authors would like to acknowledge with thanks the opportunity provided by the organizers of the MOL2NET-1 conference to present our findings to a general audience.

Author Contributions

AN conceived the problem and hypothesis and wrote the paper; SB contributed several suggestions, reviewed the manuscript and made critical comments to improve the presentation.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1 Baigent, S.J.; McCauley, J.W. Glycosylation of haemagglutinin and stalk-length of neuraminidase combine to regulate the growth of avian influenza viruses in tissue culture. Virus Res., 2001, 79, 177-185.

2 Nandy, A; Sarkar, T; Basak, S C; Nandy, P; Das, S. Characteristics of Influenza HA-NA Interdependence Determined Through a Graphical Technique. Curr Comput Aided Drug Design 2014, 10 (4), 285-302.

3 Zhang, Y.; Lin, X.; Wang, G.; Zhou, J.; Lu, J.; Zhao, H.; Zhang, F.; Wu, J.; Xu, C.; Du, N.; Li, Z.; Zhang, Y.; Wang, X.; Bi, S.; Shu, Y.; Zhou, H.; Tan, W.; Wu, X.; Chen, Z.; Wang, Y. Neuraminidase and hemagglutinin matching patterns of a highly pathogenic avain and two pandemic H1N1 influenza A Virus, PLoS One, 2010, 5, e9167

4 Xu, R.; Zhu, X.; McBride, R.; Nycholat, C.M.; Yu, W.; Paulson, J.C.; Wilson, I.A. Functional balance of the hemagglutinin and neuraminidase activities accompanies the emergence of the 2009 H1N1 influenza pandemic. J. Virol., 2012, 86, 9221-9232.

5 Schnitzler, S.U.; Schnitzler, P. An update on swine-origin influenza virus A/H1N1: a review. Virus Genes, 2009, 39, 279-292.

6 CDC Report: Avian influenza A (H7N9) virus. http://www.cdc.gov/flu/avianflu/h7n9-virus.htm. (Accessed April 30, 2014).

7 Nandy, A; Basak, S C. Prognosis of possible reassortments in recent H5N2 epidemic influenza in USA: Implications for computer-assisted surveillance as well as drug/vaccine design. Curr Comput Aided Drug Design 2015, 11(2), 110-116.

8 Wei, Hu. The interaction between the 2009 H1N1 influenza a hemagglutinin and neuraminidase: mutations, co-mutations and the NA stalk motif. J. Biomed. Sc. Engg., 2010, 3, 1-12.

9
http://www.aphis.usda.gov/wps/portal/aphis/newsroom/news/sa_stakeholder_announcements/sa_by_date/sa_2015/sa_04/ct_hpai_sd_m n_wi (Accessed 12 May 2015)

10 Huffstutter, P.J.; Steenhuysen, J. (2015). Increased human protections offered as H5N2 outbreak spreads. Reuters 27 April 2015. http://www.reuters.com/article/2015/04/27/us-health-birdflubiosecurity-insight-idUSKBN0NI0AU20150427 (Accessed 26 May 2015)

11 http://www.aphis.usda.gov/stakeholders/downloads/2015/saees_hpai_minnesota.pdf (Accessed 12 May 2015)

12 Hall, J.S.; Dusek, R.J.; Spackman, E. Rapidly expanding range of highly pathogenic avian influenza viruses. Emerg Infect Dis. 2015 Jul. http://dx.doi.org/10.3201/eid2107.150403 [Accessed 22 May 2015]

13
http://www.who.int/influenza/human_animal_interface/EN_GIP_201503031cumulativeNumberH5N1cases.pdf?ua=1 (Accessed 16 May 2015)

14 World Health Organization. "Avian Influenza: Assessing the Pandemic Threat." January, 2005. WHO/CDS/2005.29
http://apps.who.int/iris/bitstream/10665/68985/1/WHO_CDS_2005.29_eng.pdf?ua=1
(Accessed 22 May 2015)

15 http://www.ncbi.nlm.nih.gov/

16 Nandy, A. A new graphical representation and analysis of DNA sequence structure: I. Methodology and application to Globin genes. Curr. Sci., 1994, 66, 309-314.

SciForum
Mol2Net

QSPR-Perturbation Models for the Prediction of B-Epitopes from Immune Epitope Database: An Interesting Route for Predicting *"in silico"* New Optimal Peptide Sequences and/or Boundary Conditions for Vaccine Development

Severo Vázquez-Prieto *, Esperanza Paniagua and Florencio M. Ubeira

Laboratorio de Parasitología, Departamento de Microbiología y Parasitología, Facultad de Farmacia, Universidad de Santiago de Compostela, Campus Vida, Santiago de Compostela 15782, Spain

* Author to whom correspondence should be addressed; E-Mail: severovazquezprieto@gmail.com; Tel.: +34-881-815004; Fax: +34-981-593316.

Received: 28 October 2015 / Accepted: 28 October 2015 / Published: 2 December 2015

Abstract: In the present study, three different physicochemical molecular properties for peptides were calculated using the program MARCH-INSIDE: atomic polarizability, partition coefficient, and polarity. These measures were used as input parameters of a Linear Discriminant Analysis (LDA) in order to develop three different quantitative structure-property relationship (QSPR)-perturbation models for the prediction of B-epitopes reported in the immune epitope database (IEDB) given perturbations in peptide sequence, in vivo process, experimental techniques, and source or host organisms. The accuracy, sensitivity and specificity of the models were >90% for both training and cross-validation series. The statistical parameters of the models were compared to the results achieved with the electronegativity QSPR-perturbation model previously reported. The results indicate that this type of approach may constitute an interesting route for predicting *"in silico"* new optimal peptide sequences and/or boundary conditions for vaccine development.

Keywords: Epitopes; Vaccine design; Perturbation theory; QSAR/QSPR models; Markov Chains

1. Introduction

The immune epitope database (IEDB) contains data related to antibody and T cell epitopes for humans, non-human primates, rodents, and other animal species (1). This system registers an important amount of information about the molecular structure and the experimental conditions (c_{ij}) in which different *i*-

th molecules were determined to be immune epitopes or not.

Quantitative structure-activity/property relationship (QSAR/QSPR) methods let transform molecular structures into numeric molecular descriptors (λ_i) and find relationships between these structures and their biological activity. On the other hand, perturbation theory comprises methods that add “small” variation terms to the mathematical description of problems with known solutions in order to find an appropriate solution for related problems with no known solutions.

In a recent work, González-Díaz *et al.* (2) have developed an electronegativity QSPR-perturbation model for B-epitopes reported in IEBD able to predict the probability of occurrence of an epitope after a perturbation in the peptide sequence (m_i), source organism (*so*), host organism (*ho*), immunological process (*ip*), and experimental technique (*tq*) used. In principle, there are more than 1,600 different molecular descriptors (λ_i) that may be generalized and used to solve QSPR problems in chemical structures (3). In the present study, three different physicochemical molecular properties for peptide sequences reported in IEDB were calculated in order to develop three different QSPR models able to predict the efficiency of a new peptide as B-epitope given perturbations in m_i, *so*, *ho*, *ip*, and *tq*.

2. Results and Discussion

In the present work, three different QSPR-perturbation models were developed, one for each class of molecular descriptor calculated with the software MARCH-INSIDE (Table 1). In these equations, N is the number of cases used to train the models, R_C is the canonical correlation coefficient, and U is the Wilk’s lambda or U-statistic. In line with González-Díaz *et al.* (2), the output of the models $\lambda(\varepsilon_{ij})_{\text{new}}$ is a real value function that scores the propensity with which a new peptide obtained after perturbation of the initial conditions acts as B-epitope. On the other side, the first input term $\lambda(\varepsilon_{ij})_{\text{ref}}$ is the scoring function λ of the efficiency of the initial process ε_{ij}. The function $\lambda(\varepsilon_{ij})_{\text{ref}} = 1$, if the *i*-th peptide could be experimentally demonstrated to be a B-epitope in the assay of reference (ref) carried out in the conditions c_j. $\lambda(\varepsilon_{ij})_{\text{ref}} = 0$ if otherwise. The perturbation terms $\Delta\lambda_{\text{cj}} = \lambda(m_q)_{\text{ref}} - \lambda(m_i)_{\text{new}}$ are the difference in the mean value of the molecular property in question for all amino acids in the sequence of the peptide of reference. The independent variables $\Delta\Delta\lambda_{cj} = \Delta\lambda_{cj\text{-ref}} - \Delta\lambda_{cj\text{-new}} = [\lambda(m_q)_{\text{ref}} - {}^*\lambda(c_{qr})_{\text{ref}}] - [\lambda(m_i)_{\text{new}} - {}^*\lambda(c_{ij})_{\text{new}}]$ quantify values of the conditions of the new assay *cj*-new that represent perturbations with respect to the initial conditions c_{ij}-ref of the assay of reference. The quantities ${}^*\lambda(c_{ij})$ and ${}^*\lambda(c_{qr})$ are the average values of the mean values $\lambda(m_i)$ and $\lambda(m_q)$ of the molecular property in question for all new and reference peptides in IEDB that are epitopes under the *j*-th or *r*-th boundary condition.

The models obtained here are very stable and robust, yielding values of accuracy, sensitivity and specificity > 90% for both training and cross-validation series. These models are not able to improve the model developed by González-Díaz *et al.* (2). However, the results obtained are very similar and the values of different statistical parameters demonstrate the high significance of the models, validating the consistency of the method. Thus, the information obtained from the four different types of QSPR-perturbation models developed to date may be combined to increase the likelihood of a correct prediction of new epitopes or the optimization of known peptides towards computational vaccine design.

Table 1. The best QSPR-perturbation models found in this work.

Atomic polarizability (*α*)	$\lambda(\varepsilon_{\mathrm{ij}})_{\mathrm{new}} = -4.683 \cdot \lambda(\varepsilon_{\mathrm{ij}})_{\mathrm{ref}}$ - $\mathit{44.099} \cdot \Delta\alpha_{\mathrm{seq}} + \mathit{2.667} \cdot \Delta\Delta\alpha_{\mathrm{ho}} + 16.482 \cdot \Delta\Delta\alpha_{\mathrm{so}}$ $-21.668 \cdot \Delta\Delta\alpha_{\mathrm{ip}} + 47.096 \cdot \Delta\Delta\alpha_{\mathrm{tq}} + 2.0103$ $\mathrm{N} = 155169 \quad \mathrm{Rc} = 0.91 \quad \mathrm{U} = 0.18 \quad \mathrm{p} < 0.01$
Partition coefficient (*P*)	$\lambda(\varepsilon_{ij})_{new} = -4.345 \cdot \lambda(\varepsilon_{ij})_{ref}$ - $98.689 \cdot \Delta P_{seq} + 7.741 \cdot \Delta\Delta P_{ho} + 30.378 \cdot \Delta\Delta P_{so}$ $-7.073 \cdot \Delta\Delta P_{ip} + 69.851 \cdot \Delta\Delta P_{tq} + 1.851$ $N = 155169 \quad Rc = 0.89 \quad U = 0.21 \quad p < 0.01$
Polarity (*Pol*)	$\lambda(\varepsilon_{ij})_{new} = -4.846 \cdot \lambda(\varepsilon_{ij})_{ref}$ - $708.845 \cdot \Delta Pol_{seq} + 37.565 \cdot \Delta\Delta pol_{ho} + 206.803 \cdot \Delta\Delta Pol_{so}$ $-204.545 \cdot \Delta\Delta Pol_{ip} + 661.274 \cdot \Delta\Delta Pol_{tq} + 2.084$ $N = 155169 \quad Rc = 0.92 \quad U = 0.16 \quad p < 0.01$

3. Materials and Methods

The same database recently utilized by González-Díaz *et al.* (2) was used in the present study. The calculation of the molecular descriptors was implemented in the in-house program MARCH-INSIDE (4), which makes use of a Markov Chain method to calculate the *k*-th mean values of different physicochemical molecular properties ${}^{k}\lambda(m_i)$ for *i*-th molecules (m_i) (5). In the present work, three new QSPR-perturbation models for prediction of B-epitopes reported in IEDB were developed using different types of molecular descriptors $\lambda(m_i)$ to codify structural information: atomic polarizability (*α*), partition coefficient (*P*), and polarity (*Pol*). The construction of this type of models has been explained in detail before (2); therefore, only the general equation is presented:

$$\lambda(\varepsilon_{ij})_{\mathrm{new}} = {}^{'}c_0 \cdot \lambda(\varepsilon_{qr})_{\mathrm{ref}} + \sum_{\mathrm{j=1}}^{4} {}^{'}\mathrm{d}_{\mathrm{ij}} \cdot \Delta\Delta\lambda_{ijqr} + {}^{'}e_0$$

Here, $\lambda(\varepsilon_{ij})_{\mathrm{new}}$ is the efficiency function as epitope of a new peptide obtained after a change in the structure and/or the boundary conditions $c_j \equiv (c_0, c_1, c_2, c_3 \ldots c_n)$ of a peptide of reference. The set of boundary conditions used here are the same reported in IEDB: c_0 = the specific peptide; c_1 = the organism that expresses the peptide (so_{j}); c_2 = the host organism exposed to the peptide (ho_{j}); c_3 = the immunological process (ip_{j}); and c_4 = the experimental technique (tq_{j}). The variable $\lambda(\varepsilon_{qr})_{\mathrm{ref}}$ refers to a known efficiency function as epitope of a peptide of reference experimentally determined under a set of c_j boundary conditions. The function $\lambda(\varepsilon_{ij})$ was defined as a discrete value function for classification purpose: $\lambda(\varepsilon_{ij}) = 1$ for epitopes reported in the conditions c_j and $\lambda(\varepsilon_{ij}) = 0$, when otherwise. The values c_0 and d_{ij} are the coefficients obtained for the Linear Discriminant Analysis (LDA) classification functions. The variational perturbation terms $\Delta\Delta\lambda_{ijqr}$ account both for the deviation of the molecular descriptors of all amino acids in the sequence of the new peptide with respect to the peptide of reference and with respect to all boundary conditions. The constant e_0 represents the independent term of the model.

An LDA was carried out using the STATISTICA 6.0 software (6). A forward stepwise strategy was used for variable selection, and the statistical significance of the models was determined by calculating the canonical correlation coefficient (R_c) and U-statistic. The accuracy, specificity, and sensitivity for the training and cross-validation series were also examined (7).

4. Conclusions

This work has demonstrated that atomic polarizability, partition coefficient, and polarity values calculated with MARCH-INSIDE seem to also be good molecular descriptors for finding QSPR-perturbation models which are able to predict the results of variations in peptide sequences and experimental assay boundary conditions reported in IEBD. Consequently, this type of approach may constitute an interesting route for predicting *"in silico*" new optimal peptide sequences and/or boundary conditions for vaccine development. In addition, this study may serve as a basis for building better and more reliable models in the future (e.g., consensus QSPR models). This computational technique is by no means aimed at replacing experimentation but rather helps us to somewhat rationalize this process, while at the same time reducing costs in terms of material resources and time.

Acknowledgments

This study was supported by grants AGL2011-30563-C03 (Ministerio de Ciencia e Innovación, Spain) and GPC2014/058 (Xunta de Galicia, Spain).

Author Contributions

S.V.P. conceived and designed the study, analysed and interpreted the data and wrote the paper. All authors discussed the results and implications and commented on the manuscript at all stages.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Vita, R.; Zarebski, L.; Greenbaum, J.A.; Emami, H.; Hoof, I.; Salimi, N.; Damle, R.; Sette, A.; Peters, B. 2010. The immune epitope database 2.0.. Nucleic Acids Res. 38 (Database issue), D854-862.
2. González-Díaz, H.; Pérez-Montoto, L.G.; Ubeira, F.M. 2014. Model for Vaccine Design by Prediction of B-Epitopes of IEDB Given Perturbations in Peptide Sequence, In Vivo Process, Experimental Techniques, and Source or Host Organisms. J. Immunol. Res. doi:10.1155/2014/768515.
3. Todeschini, R.; Consonni, V. 2008. Handbook of Molecular Descriptors. Wiley-VCH, Weinheim.
4. González-Díaz, H.; Molina-Ruiz, R.; Hernández, I. 2007. MARCH-INSIDE version 3.0 (MARkov CHains INvariants for SImulation & DEsign). Windows supported version under request to the main author contact email: gonzalezdiazh@yahoo.es.

5. González-Díaz, H.; Arrasate, S.; Sotomayor, N.; Lete, E.; Munteanu, C.R.; Pazos, A.; Besada-Porto, L.; Ruso, J.M. 2013b. MIANN models in medicinal, physical and organic chemistry. Curr. Top. in Med. Chem. 13, 619-641.
6. StatSoft.Inc. 2002. STATISTICA (data analysis software system), version 6.0. www.statsoft.com.
7. Hill, T.; Lewicki, P. 2006. STATISTICS: Methods and Applications: A Comprehensive Reference for Science, Industry and Data Mining. StatSoft, Tulsa.

Mol2Net, **2015**, 1(*Section B*), pages 1-11, *Proceeding*
http://sciforum.net/conference/mol2net-1

SciForum
Mol2Net

Towards Computational Prediction of Biopharmaceutics Classification System: A QSPR Approach *

Hai Pham-The [1],*, Huong Le-Thi-Thu [2], Teresa Garrigues [3], Marival Bermejo [4], Isabel González-Álvarez [4] and Miguel Ángel Cabrera-Pérez [3,4,5]

[1] Hanoi University of Pharmacy, 13-15 Le Thanh Tong, Hoan Kiem, Hanoi, Vietnam
[2] School of Medicine and Pharmacy, Vietnam National University, 144-Xuan Thuy, Cau Giay, Hanoi, Vietnam; E-Mail: ltthuong1017@gmail.com
[3] Department of Pharmacy and Pharmaceutical Technology, University of Valencia, Burjassot 46100, Valencia, Spain; E-Mails: Teresa.Garrigues@uv.es (T.G.); macabreraster@gmail.com (M.A.C.-P.)
[4] Department of Engineering, Area of Pharmacy and Pharmaceutical Technology, Miguel Hernández University, 03550 Sant Joan d'Alacant, Alicante, Spain; E-Mails: mbermejo@umh.es(M.B.)
[5] Unit of Modeling and Experimental Biopharmaceutics, Chemical Bioactive Center, Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba
* Author to whom correspondence should be addressed; E-Mail: thehai84@yahoo.com; Tel.: +84-996-888-868; Fax: +84-4-39710550.

Received: 29 October 2015 / Accepted: 29 October 2015 / Published: 2 December 2015

Abstract: Today classification of drug candidates on the Biopharmaceutics Classification System (BCS) has become an important issue in pharmaceutical researches. In this work, we provide a potential *in silico* approach to predict this system using two separately classification models of Dose number and Caco-2 cell permeability. 18 statistical linear and nonlinear models have been constructed based on 803 0-2D Dragon and 126 Volsurf+ molecular descriptors to classify the solubility and permeability properties. The voting consensus model of solubility (VoteS) showed a high accuracy of 88.7% in training and 92.3% in test set. Likewise, for the permeability model (VoteP), accuracy was 85.3% in training and 96.9% in test set. A combination of VoteS and VoteP appropriately predicts the BCS class of drugs (overall 73% with class I precision of 77.2%). This consensus system predicts the BCS allocations of 57 drugs appeared in the WHO Model List of Essential Medicines with 87.5% of accuracy. A simulation of a biopharmaceutical screening assay has been proved in a large data set of 37,377 compounds in different drug development phases (1, 2, 3 and launched), and NMEs. Distributions of BCS forecasts illustrate the current status in drug discovery and development. It is anticipated that developed QSPR models could offer the best estimation of BCS for NMEs in early stages of drug discovery.

Keywords: Biopharmaceutics Classification System (BCS); Dose Number; Caco-2 cell permeability; Quantitative Structure Activity/Property Relationship (QSAR/QSPR)

Mol2Net YouTube channel: *http://bit.do/mol2net-tube*

*The full content of this communication can be partly found in Pham-The H *et al*. Mol. Pharmaceutics, 2013; 10(6): 2445-61 and Pham-The H *et al*. Mol. Inf. 2013; 32(5-6): 459-79.

1. Introduction

After almost 20 years of the introduction and exploration of the Biopharmaceutics Classification System (BCS), it has gained a major impact on the regulation and development of immediate release (IR) solid oral drug products [1,2]. Based on the principal factors that determine the rate and extent of drug absorption, the BCS provides a scientific framework for classifying drug substances into one of four categories. According to BCS, IR solid oral dosage forms are categorized as having either rapid or slow *in vitro* dissolution, and then classified based on aqueous solubility and intestinal permeability of the active pharmaceutical ingredient (API) [1]. This system has been formally adopted by the US FDA [3], the European agency EMEA [4] and the World Health Organization (WHO) [5] as a technical standard for waiving BE test requirements for oral drugs. A recent study of the economic impact of granting biowaivers for class I and III BCS demonstrated an impressive saving annual expenditure on running BE studies, being more than 120 million dollars between the two classes [6]. Because it avoids unnecessary drug exposures to healthy subjects, while maintaining the high public health standard for therapeutic equivalence, the BCS is, without doubt, a potential tool for speeding up and reducing the cost of drug development.

There is a continuing effort worldwide to detect, in the early discovery, the possible BCS-based biowaiver candidates, e.g. BCS class I drugs [7]. One of the common strategies is based on BCS provisional classification in which the drugs are classified by two sources: dose related solubility data (Dose number, Do) and estimated human absorption data, i.e. *in vitro* permeability (usually determined by the Caco-2 cell cultured method) [3,8], or simple *in silico* partition coefficient calculation [9]. In this regard, *in silico* approach presents the two most important advantages: (i) provides a flexible approach that can be applied in different stages of drug development with different purposes, and (ii) allows estimating the BCS classes of new molecular entities (NMEs) without knowledge of therapeutic dosage. Definitely, with respect to experimental methods, computational approaches are cost-saving and no sample requirement methods.

However, up to now, robust *in silico* approach, i.e. Quantitative Structure-Activity/Property Relationships (QSAR/QSPR) modeling, has not been explored sufficiently in the BCS studies. Based on published findings [10], and to respond to the rising need of early identification of possible biowaiver drugs, in this work, we attempt to develop robust QSPR models to classify the solubility and permeability terms that compose the BCS (Figure 1). These models were rigorously validated on various published BCS class drug sets [5,9,11-13] and the feasibility of performing PBC prediction in early drug discovery is discussed.

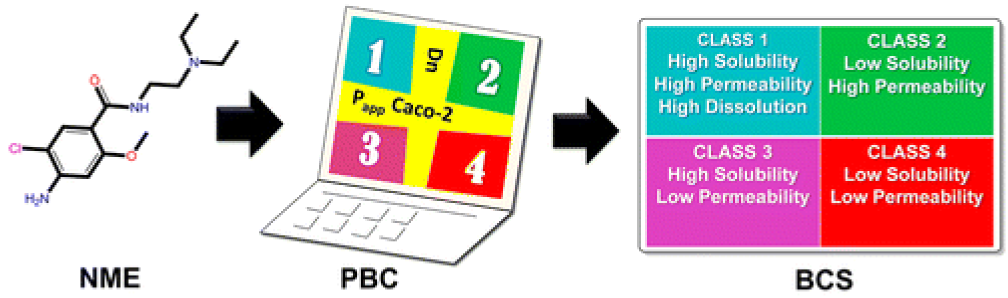


Figure 1. Summary scheme of current in silico study

2. Results and Discussion

In 2004, a number of 123 orally administered drugs on the World Health Organization (WHO) Essential Medicine List (EML) were initially classified into BCS [9,11]. Later, 200 oral drug products in the United States, Great Britain, Spain, and Japan were classified based on published solubility data and permeability data estimated by calculated log *P* [12]. Recently, increasing attention has been turned out for determining the Provisional Biopharmaceutical location of orally administered immediate-release (IR) drug products using different estimated gastrointestinal permeability, such as partition coefficients (log *D* and log *P*), molecular surface area (PSA) or other *in vitro* permeability.[14-17] It has been emphasized that the distribution of BCS class I, II, III, and IV in each classification are quite different. In this report, taking advantage of the availability of experimental *in vitro* Caco-2 cell data a Provisional Biopharmaceutical Classification (PBC) of 322 oral drug products gathered from literature was performed. To our knowledge, it is the largest data set for such classification. Classifications of current data are described in [7].

Physicochemical profiling of PBC. It is very useful to analyze the *similarity* between physicochemical spaces characterized by PBC classes, especially for developing computational predictions of current PBC and further BCS. Thus, six commonly used physicochemical parameters were calculated by Dragon and Volsurf+ for this analysis:[18,19] molecular weight (MW), polar surface area (PSA), Mlog *P*, log D_6, log $D_{7.5}$, total number of hydrogen bond donors and acceptors (nHA+B), number of free rotatable bonds (RBN), and estimated ionization states. The average and median values of maximum dose strength (D_{max}) as well as Caco-2 P_{app} were also analyzed for each class.

Unsurprisingly, class II drugs display the highest lipophilicity, while class III and IV are more hydrophilic. Class I drugs represent a balanced physicochemical profile even though they tend to be more lipophilic. In general, only the hydrogen bonding term is fairly different from one class to another. There is certain *physicochemical similarity* between class I and II (Mlog *P*, log *D* at basic medium), class III and IV (nHA+B, PSA), or class II and III (MW), etc. Values of D_{max} do not present any trend. It is demonstrated that poor bioavailability is more likely when the compounds violate two or more of the Lipinski's rules (Ro5): (i) log *P* <5, (ii) MW< 500, (iii) HBD (hydrogen bond donors) < 5, and (iv) HBA (hydrogen bond acceptors) < 10.[20] Current data was collected mostly among successful drugs. Then, it is easy to understand that many of them (>95%) passed the Ro5.

Computational models to predict PBC class from chemical structures. Solubility and Caco-2 permeability were modeled independently. The final computational PBC classification was achieved using two voting consensus (permeability and solubility) systems. QSPR models obtained by different statistical techniques for each property are described below.

Solubility modeling. Three model series were obtained using LDA, QDA and BLR. Different molecular descriptors (MDs) were used for building QSPR models. From every model series constructed with every technique, the best one was selected (detailed comparisons are described in supplement documents). Table 1 summarizes the mathematical equations and performances of the three best models for PBC solubility prediction.

Permeability modeling. The same procedure was carried out to select the best classifiers for PBC permeability class. Table 2 displays the relevant information of permeability models.

Classifications of four PBC classes. The two obtained voting models were finally combined to estimate the four PBC classes of the data (322 compounds). Table 3 displays the confusion matrix of this consensus system. A good overall accuracy of 73.0 % was obtained by this system..

Analysis of molecular descriptors (MDs). Interestingly, the PBC solubility and permeability terms are well described using a small set of MDs.

Table 1. Performances of the three best models for PBC solubility classification

Technique	Descriptor family	MCC	Accuracy	Specificity	Sensitivity	Precision	AUC (Ts)[b]
		% (Tr/Ts)[a]					
LDA (S1)	0-2D Dragon *plus* Volsurf+	0.66/0.54	83.3/76.9	82.2/79.3	84.0/75.0	86.9/81.8	0.88±0.04
QDA (S2)	0-2D Dragon	0.63/0.75	81.7/87.7	82.2/82.8	81.3/91.7	86.5/86.8	0.97±0.04
BLR (S3)	0-2D Dragon	0.60/0.69	80.5/84.6	75.5/82.1	84.1/86.5	83.0/86.5	0.96±0.03
VoteS	All	**0.68/0.87**	**84.4/93.9**	**85.0/89.3**	**84.0/97.2**	**88.7/92.3**	–

Mathematical equations

$$\mathit{CLASS}_{Do}(+/-) = -1.59 - 0.54\times \mathit{P\_VSA\_v\_3} + 0.80\times \mathit{nArC{=}N} + 0.65\times \mathit{C\text{-}005} - 0.84\times \mathit{CATS2D\_04\_AL} + 0.79\times \mathit{DLS\_04} + 4.51\times \mathit{ID3} + 0.28\times A - 0.41\times \mathit{LgD5} \tag{S1}$$

$N = 257$ $\lambda = 0.60$ $D^2 = 2.74$ $F = 25.61$ $p < 0.0001$

$$\mathit{CLASS}_{Do}\,(+/-) = -0.36 - 0.90\times \mathit{Me} - 1.40\times \mathit{nCt} - 0.79\times \mathit{NssNH} + 1.22\times \mathit{BLTD48} + 0.87\times \mathit{DLS\_04} - 0.82\times \mathit{CMC\text{-}50} - 1.86\times \mathit{nArC{=}N}\times \mathit{N\text{-}067} + 0.41\times \mathit{N\text{-}067}\times \mathit{NssNH} - 0.73\times \mathit{Me}\times \mathit{CMC\text{-}50} + 0.51\times \mathit{nR10}^2 \tag{S2}$$

$N = 257$ $\lambda = 0.59$ $D^2 = 2.88$ $p < 0.0001$

$$\mathit{Ln}\,(P+/P-) = 2.63 - 0.59\times \mathit{nCp} + 4.44\times \mathit{nArC{=}N} + 0.20\times \mathit{H\text{-}052} + 1.82\times \mathit{N\text{-}067} - 1.32\times \mathit{NssNH} + 1.09\times \mathit{BLTD48} + 4.58\times \mathit{LDS\_04} - 1.38\times \mathit{CMC\text{-}50} - 0.38\times \mathit{nO} \tag{S3}$$

[a]Measured performances of training/test set; [b]Area under the ROC curve determined on test set by non-parametric assumptions in 95% asymptotic confidence interval.

Table 2. Performances of the three best models for PBC permeability classification

Technique	Descriptor family	MCC	Accuracy	Specificity	Sensitivity	Precision	AUC (Ts)[b]
		% (Tr/Ts)[a]					
LDA (P1)	0-2D Dragon *plus* Volsurf+	0.63/0.69	81.6/84.9	81.9/85.7	81.4/84.2	82.0/88.9	0.93±0.03
QDA (P2)	0-2D Dragon	0.65/0.76	82.4/87.9	81.1/89.3	83.7/86.8	81.8/91.7	0.94±0.03
BLR (P3)	0-2D Dragon *plus* Volsurf+	0.64/0.73	82.0/86.4	79.5/89.3	84.5/84.2	80.7/91.4	0.92±0.03
VoteP	All	**0.70/0.77**	**85.2/87.9**	**85.0/96.4**	**85.3/81.6**	**85.3/96.9**	–

Mathematical equations

$$\mathit{CLASS}_{Papp}(+/-) = -5.91 + 0.01\times \mathit{P\_VSA\_s\_6} - 1.62\times \mathit{nRNR2} - 0.74\times \mathit{C\text{-}016} + 2.64\times \mathit{CATS2D\_08\_AP} + 4.23\times \mathit{LLS\_01} + 0.01\times \mathit{WN2} + 3.79\times \mathit{CACO2} \tag{P1}$$

$N = 256$ $\lambda = 0.57$ $D^2 = 2.81$ $F = 22.24$ $p < 0.0001$

$$\mathit{CLASS}_{Papp}\,(+/-) = 0.32 - 1.02\times \mathit{GATS2m} + 0.95\times \mathit{GATS2s} - 0.55\times \mathit{nRNR2} - 0.52\times \mathit{B03[O\text{-}O]} \tag{P2}$$

$-1.95\times SAdon + 0.82\times LLS\text{-}01 + 3.46\times nC\text{=}N\text{-}N\text{<}\times B04[O\text{-}Cl]$ $+ 0.37\times nRNR2\times SAdon + 0.32\times CATS2D\_03\_DD\times SAdon - 0.46\times B08[C\text{-}O]^2$ $N = 256 \quad \lambda = 0.55 \quad D^2 = 3.18 \quad p < 0.0001$	
$Ln\ (P+/P-) = 5.49 - 2.05\times nRNR2 + 3.74\times CATS2D\_07\_DP + 1.88\times CACO2 - 5.04\times GATS2m$ $- 22.48\times nFuranes - 0.02\times SAdon - 1.05\times nRCOOH$	**(P3)**

[a]Measured performances of training/test set; [b]Area under the ROC curve determined on test set by non-parametric assumptions in 95% asymptotic confidence interval.

It is important to note that there are some MDs directly related to polarizability and dispersion forces within molecules (*nCp*, *nCt*), molecular size (*nR10*, P*_VSA_v_3*), lipophilicity and hydrophobicity (*BLTD48*, *CATS2D_04_AL*, *CMC-50*), and especially, the polar, chargeable and hydrogen bond forming capacity (*A*, *Me*, *nO*, *nArC=N*, *C-005*, *N-067*, *NssNH*, *LgD5*). Beside, rule based MDs, which represent common physicochemical combination trends of known drug-like and lead-like dataset,[21,22] are selected. Generally, current finding structure-property (*Do*) relationship (S*Do*R) are rather similar with Khandelwal *et al.*'s analysis.[23]

On the other hand, the ionization state (*GATS2s*, *P_VSA_s_6*, *nRNR2*), molecular size (*GATS2m, nFuranes*, *C-016*) and hydrogen bond donor and acceptor regions (*nRCOOH*, *nRNR2*, *nC=N-N<*, *CATS2D_03_DD*, *CATS2D_07_D*, *CATS2D_08_AP*, *SAdon*, *WN2* etc.) are well correlated with Caco-2 permeability. The ADME descriptor *CACO2* was selected two times in permeability models. Please note that numeric values of this variable are result of partial least square (PLS) discriminant analysis developed by Zamora *et al.*[24] Unfortunately, the use of this descriptor does not provide precise knowledge of descriptor impacts on PBC permeability class.

Regulatory validation and applications of in silico PBC models. A robust forecast of PBC class is very useful in early drug discovery. Especially, for many NMEs whose therapeutic dose-ranges are not available in preclinical stages. This is also important for estimating possible BCS memberships, since there is a great correspondence between proposed PBC and BCS cited in regulatory guidelines [5].

Table 3. Confusion matrix of consensus system for the prediction of PBC classes

	Predicted PBC Class I	Predicted PBC Class II	Predicted PBC Class III	Predicted PBC Class IV	Total	Accuracy (%)	MCC
PBC Class I	**61**	11	18	1	91	67.0	0.62
PBC Class II	10	**59**	2	5	76	77.6	0.67
PBC Class III	7	4	**74**	12	97	76.3	0.63
PBC Class IV	1	8	8	**41**	58	70.7	0.63
Total	79	82	102	59	322		
Precision (%)	77.2	72.0	72.5	69.4			

Biopharmaceutical Screening Simulations

Finally, a large database of drugs, clinical and non-clinical trial compounds was subjected to computational prediction using *in silico* PBC consensus model. A total number of 37,202 compounds were analyzed (Figure 2). Recently, this database was classified by *in silico* BDDCS consensus models to estimate the distribution of BDDCS class.[10] In contrast to that study, obtained models here are employed for comparing the predictions and then making a round estimation of the distribution of BCS class. It is important to note that some compounds obtained non-conclusive-classification due to their condition of outliers of Ads. 1699 compounds (4.6% of prediction data) are classified as I/II, I/III, II/IV, and III/IV. Most of them (1512 compounds) are low-activity (W6) and high-activity (W9) compounds.[10] Especially, 29 compounds could not be classified by *in silico* models. Among those conclusively predicted as PBC class I, II, III and IV, there exists similar proportion between launched and clinical phase 3 drugs, between clinical phases 1 or 2 drugs and W6 or W9 compounds.

As can be appreciated from Figure 3, more than 40% of drugs and phase 3 are similar to PBC class I. The phase 3 compounds similar to PBC class II significantly overcome the PBC class III but for drugs, their percentage become similar. Compounds classified as PBC class IV take the minimal proportion in the two drug sets (7-8%). In contrast, about 50% of phase-1 and phase-2 drugs are predicted as PBC class II. This percentage is even greater (62-63%) in W6 and W9 datasets. Compounds predicted as PBC class I maintain the same proportion with respect to phase 1, 2, W6 and W9 whole dataset. There is a noticeable change of the predicted PBC class III for phase 1 and 2 drugs (15-18%) compared to W6 and W9 (7%) compounds. Particularly, compounds of W6 data set, predicted PBC class IV compounds outnumber those of predicted as PBC class III. These trends of PBC class predictions reflect the drug development process and agree, in turn, upon some points with previous findings.[10,25]

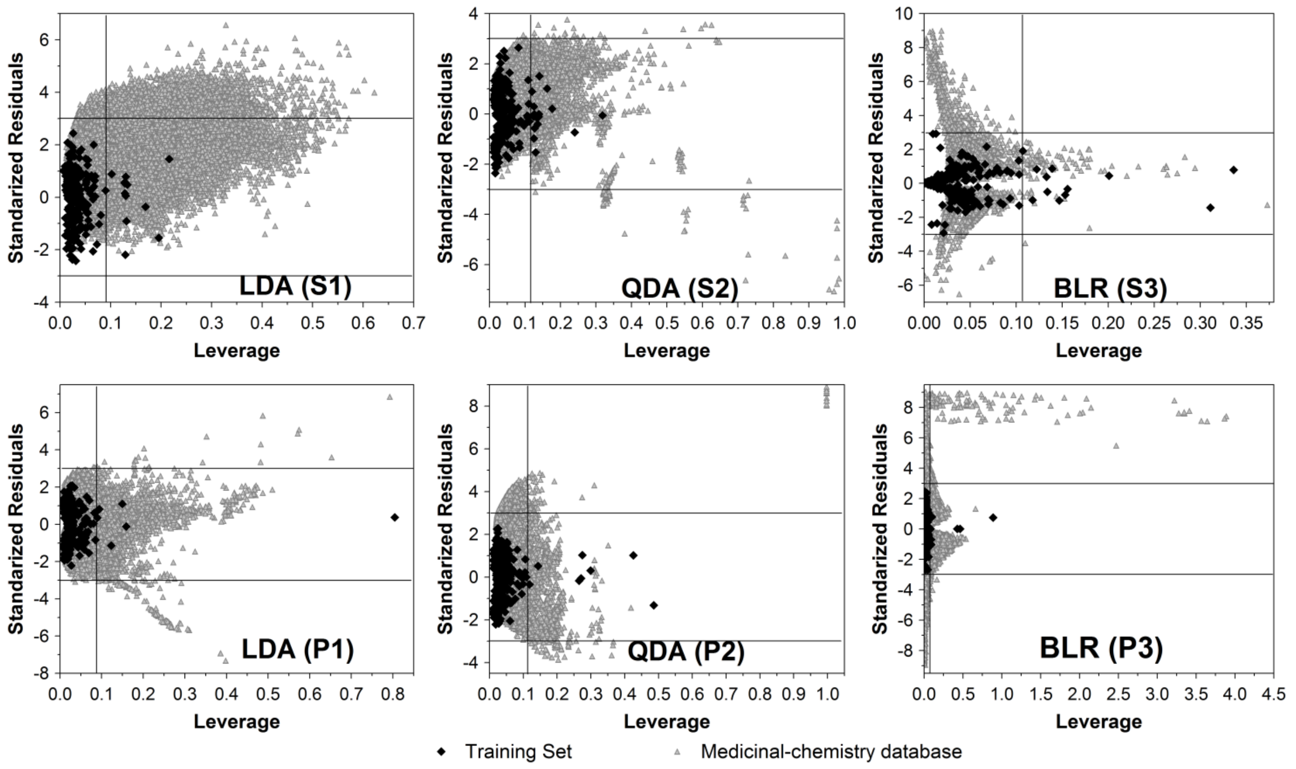


Figure 2. William's plots based on solubility and permeability models for training and screening large medicinal-chemistry database

3. Materials and Methods

Data set. BCS based-provisional classification requires both solubility and permeability measurements. In this work, a set of 322 drugs was obtained from published works. A provisional classification was executed by means of an extensive literature revision of experimental values and assigned classes, as follows.

Solubility data. The drug solubility data (in mg/mL) can be obtained from standard references,[9] such as the Pharmacopeia [26] or the Merck Index.[27] Due to the extensive survey, herein we only report the lowest solubility under the conditions listed above. In addition, scale-up guidelines were taken from Kasim *et al*. whenever solubility data was not available or was undefined.[9]

Maximum Dose Strength. Two reference sources were mainly used for searching values of maximum dose strength (mg): (i) the WHO Model List of Essential Medicines,[28] and (ii) Orange Book.[29] For drugs that are not included in these documents or exist in different market presentations, the first introduced strengths were revised and used as highest dosages. Doses in mg/kg were transformed into mg assuming 70Kg as body weight.

*Dose Number Calculations***.** The dose number (D_0) was calculated using the following equation

$$D_0 = \frac{(M0/V0)}{S} \quad (1)$$

where, M_0 is the highest dose strength (mg), S is the aqueous solubility (mg/mL) under conditions mentioned above and water volume V_0 is assumed to be 250 mL.[1,9] Drugs with $D_0 \leq 1$ were classified as high-solubility drugs. Conversely, drugs with $D_0 > 1$ were assigned as low solubility drugs.[9]

Permeability Estimations. In this work, *in vitro* Caco-2 cell permeability is used to classify drug according to BCS. For this purpose, we take advantage of our previous research where an extensive literature survey of this kind of data was processed.[30] Besides, we have adopted the same method proposed by Kim *et al*.,[31] taking the average permeability value of Metoprolol (average apparent permeability P_{app} = 20×10^{-6} cm/s) for benchmarking the high permeability class boundary. Due to the large revised literature, the mean values were listed, excluding those laid outside of the mean±2SD (standard deviation) ranges. Additionally, available data obtained on both directions apical to basolateral ($P_{app,\ A\text{-}B}$) and viceversa ($P_{app,\ B\text{-}A}$) were taken into account.

Computational methods. Taking all above together, in this work efforts have been made to establish really useful statistical predictors for BCS classes of NMEs based on two separate model series of dose number and Caco-2 cell permeability. To attain this purpose, the following computational procedures should be considered: (i) suitably computing physicochemical and molecular descriptors, (ii) rational selection of training and test sets, (iii) establishment of modeling strategy and appropriated variable selection, and (iv) ascertainment of BCS predictions for NMEs in the context of regulatory statements.

Molecular descriptor calculations. 803 simple (0-2D) descriptors belonging to 29 families implemented in Dragon software *version* 6.0,[19] and 126 molecular descriptors in VolSurf+ *version* 1.0.4 [18] were calculated.

Model building and feature selection. Three statistical classification algorithms were applied in order to detect all possible (linear or non-linear) relationships between solubility/permeability and computed parameters: LDA (Linear Discriminant Analysis), QDA (Quadratic Discriminant Analysis) and BLR (Binary Logistic Regression).

Performances of models were evaluated using false positive rate (FPr), true negative rate (TN, for specificity), true positive rate (TP, for sensitivity), Matthews Correlation Coefficient (MCC) and predictive accuracy, as defined below:

$$\text{Specificity} = TN/(TN+FP) \quad (2)$$

$$\text{Sensitivity} = TP/(TP+FN) \quad (3)$$

$$\text{Precision} = TP/(TP+FP) \qquad (4)$$

$$\text{MCC} = [(TP\times TN) \times (FP\times FN)]/[(TP+FP)(TP+FN)(TN+FP)-(TN+FN)]^{1/2} \qquad (5)$$

$$\text{Accuracy} = (TN+TP)/(TN+TP+FN+FP) \qquad (6)$$

For reliable predictions of these three *external* datasets, it is important to consider all applicability domains (ADs) defined by the chemical spaces of the training set. There are many approaches for AD estimation.[32] Here, the leverage approach, a geometric method commonly used for QSAR problems, was employed. The leverage of a compound in the original variable space is defined as $h_i = [X(X'X)^{-1}X']$, where X is the descriptor matrix derived from the training set descriptor values. The warning leverage (h*) is defined as $h^*=3(p+1)/n$, where n is the number of training compounds, and p is the number of predictor variables [32]. Compounds with $h_i > h^*$ were observed to reveal their influence on classification performance. It is not necessary to exclude them from predictions although they appear to be outside AD. However, compounds are considered to be outliers if they lay outside the ±3 standardized residual (δ) range [32].

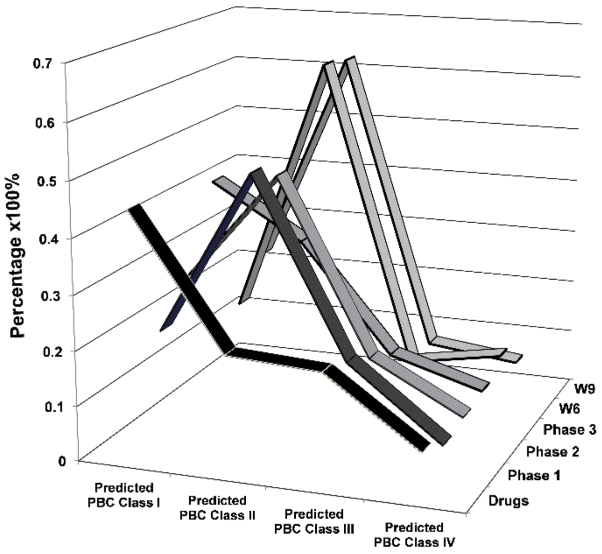


Figure 3. . Distribution comparison of computational PBC assignments of launched drugs, compounds in different drug development stages (phase 1, 2, 3), and bioactive micromolar (W6) and nanomolar (W9) compounds.[10]

4. Conclusions

In this report, a systematic study was carried out in order to standardize a BCS-based provisional classification of 322 drugs and develop computational predictions of BCS class for NMEs. It is of great interest to assign as soon as possible the probable BCS class of a drug candidate. By using extensively revised references of solubility and *in vitro* Caco-2 permeability, a very commonly used preclinical assay in pharmaceutical industry, a better *in vivo* BCS classification of drugs is anticipated. Consequently, the classification results in this study display a high concordance with BCS classification of common regulatory authorities (WHO, FDA). Other classification schemes were compared with PBC. Large additional information concerning the BCS classification of current data was analyzed in order to identify advantages as well as limitations when using PBC. As an attempt to develop QSPR models able to predict the PBC class, it was demonstrated the possibility of screening NMEs in the early phase of drug development.A combination of *in silico* and *in vitro* approaches provides a basis for robust estimation of the BCS class of NMEs without clinical information and contribute to early selection of biopharmaceutical promissory drug candidates. As a relevant limitation, this data set consists of a small number of drugs. Besides, the uncertainty of the relationship between absorption extent and proposed provisional classification (especially for low absorbed drugs) remains. A modification of BCS classification scheme (particularly for class II and III) is needed. A further compilation of *in vitro* permeability data and aqueous solubility may enhance the applicability domain of *in silico* classifications.Main text paragraph. .

Acknowledgments

H.L-T-T is supported by Vietnam National University. H.P-T, M.B, I.G-A, T.G and M.A.C-P acknowledge financial support of AECID (Grant No. 1- D/031152/10 and DCI-ALA/19.09.01/10/21526/245-297/ALFA 111(2010)29.

Author Contributions

All the authors contributed equally.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Amidon, G.L.; Lennernas, H.; Shah, V.P.; Crison, J.R. A theoretical basis for a biopharmaceutic drug classification: The correlation of in vitro drug product dissolution and in vivo bioavailability. *Pharm. Res.* **1995**, *12*, 413-420.
2. Chen, M.L.; Amidon, G.L.; Benet, L.Z.; Lennernas, H.; Yu, L.X. The bcs, bddcs, and regulatory guidances. *Pharm. Res.* **2011**, *28*, 1774-1778.
3. CDER/FDA. *Fda guidance for industry: Waiver of in vivo bioavailability and bioequivalence studies for immediate-release solid oral dosage forms based on a biopharmaceutics classification system*; Federal Drug and Food Administration: Rockville, MD, USA: Center for Drug Evaluation and Research, 2000.
4. CPMP/EWP/QWP/1401/98. *Note for guidance on the investigation of bioavailability and bioequivalence*; The European Agency for the Evaluation of Medicinal Products (EMEA): London, December 14, 2000.
5. *Annex 8: Proposal to waive in vivo bioequivalence requirements for who model list of essential medicines immediate-release, solid oral dosage forms*; Technical Report Series No. 937; WHO Expert Committee on Specification for Pharmaceutical Preparations: 2006; pp 391-461.
6. Cook, J.A.; Davit, B.M.; Polli, J.E. Impact of biopharmaceutics classification system-based biowaivers. *Mol. Pharmaceutics* **2010**, *7*, 1539-1544.
7. Pham-The, H.; Garrigues, T.; Bermejo, M.; González-Álvarez, I.; Monteagudo, M.C.; Cabrera-Pérez, M.Á. Provisional classification and in silico study of biopharmaceutical system based on caco-2 cell permeability and dose number. *Mol. Pharmaceutics* **2013**, *10*, 2445-2461.
8. Dahan, A.; Lennernäs, H.; Amidon, G.L. The fraction dose absorbed, in humans, and high jejunal human permeability relationship. *Mol. Pharmaceutics* **2012**, *9*, 1847−1851.
9. Kasim, N.A.; Whitehouse, M.; Ramachandran, C.; Bermejo Sanz, M.; Lennernas, H.; Hussain, A.S.; Junginger, H.E.; Stavchansky, S.A.; Midha, K.K.; Shah, V.P.*, et al.* Molecular properties of who essential drugs and provisional biopharmaceutical classification. *Mol. Pharmaceutics* **2004**, *1*, 85-96.
10. Broccatelli, F.; Cruciani, G.; Benet, L.Z.; Oprea, T.I. Bddcs class prediction for new molecular entities. *Mol. Pharmaceutics* **2012**, *9*, 570-580.

11. Lindenberg, M.; Kopp, S.; Dressman, J.B. Classification of orally administered drugs on the world health organization model list of essential medicines according to the biopharmaceutics classification system. *Eur. J. Pharm. Biopharm.* **2004**, *58*, 265-278.
12. Takagi, T.; Ramachandran, S.; Bermejo, M.; Yamashita, S.; Yu, L.X.; Amidon, G.L. A provisional biopharmaceutical classification of the top 200 oral drug products in the united states, great britain, spain, and japan. *Mol. Pharmaceutics* **2006**, *3*, 631-643.
13. Wu, C.Y.; Benet, L.Z. Predicting drug disposition via application of bcs: Transport/absorption/ elimination interplay and development of a biopharmaceutics drug disposition classification system. *Pharm. Res.* **2005**, *22*, 11-23.
14. Shawahna, R.; Rahman, N.U. Evaluation of the use of partition coefficients and molecular surface properties as predictors of drug absorption: A provisional biopharmaceutical classification of the list of national essential medicines of pakistan. *DARU* **2011**, *19*, 83-99.
15. Varma, M.V.; Gardner, I.; Steyn, S.J.; Nkansah, P.; Rotter, C.J.; Whitney-Pickett, C.; Zhang, H.; Di, L.; Cram, M.; Fenner, K.S.*, et al.* Ph-dependent solubility and permeability criteria for provisional biopharmaceutics classification (bcs and bddcs) in early drug discovery. *Mol. Pharmaceutics* **2012**, *9*, 1199-1212.
16. Custodio, J.M.; Wu, C.Y.; Benet, L.Z. Predicting drug disposition, absorption/elimination/transporter interplay and the role of food on drug absorption. *Adv. Drug Deliv. Rev.* **2008**, *60*, 717-733.
17. Nair, A.K.; Anand, O.; Chun, N.; Conner, D.P.; Mehta, M.U.; Nhu, D.T.; Polli, J.E.; Yu, L.X.; Davit, B.M. Statistics on bcs classification of generic drug products approved between 2000 and 2011 in the USA. *AAPS J.* **2012**, *14*, 664-666.
18. *Volsurf+*, version 1.0.4; available from Molecular Discovery Ltd., London, U.K. (http://www.moldiscovery.com).
19. *Dragon for windows (software for molecular descriptor calculator).* 6.0; Talete srl, Milano Chemometrics and QSAR Research Group: http://www.talete.mi.it/products/dragon_description.htm.
20. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in the drug discovery and development settings. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3-25.
21. Chen, G.; Zheng, S.; Luo, X.; Shen, J.; Zhu, W.; Liu, H.; Gui, C.; Zhang, J.; Zheng, M.; Puah, C.M.*, et al.* Focused combinatorial library design based on structural diversity, druglikeness and binding affinity score. *J. Comb. Chem.* **2005**, *7*, 398-406.
22. Ghose, A.K.; Viswanadhan, V.N.; Wendoloski, J.J. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* **1999**, *1*, 55-68.
23. Khandelwal, A.; Bahadduri, P.M.; Chang, C.; Polli, J.E.; Swaan, P.W.; Ekins, S. Computational models to assign biopharmaceutics drug disposition classification from molecular structure. *Pharm. Res.* **2007**, *24*, 2249-2262.
24. Zamora, I.; Oprea, T.I.; Ungell, A.L. Prediction of oral drug permeability. In *Rational approaches to drug design*, Holtje, H.D.; Sippl, W., Eds. Prous Science Press: Barcelona, Spain, 2001; pp 271-280.

25. Benet, L.Z.; Broccatelli, F.; Oprea, T.I. Bddcs applied to over 900 drugs. *AAPS J.* **2011**, *13*, 519-547.
26. *The international pharmacopoeia*. 4th ed.; World Health Organization: WHO Press, 20 Avenue Appia, 1211 Geneva 27, Switterland, 2006; Vol. 1 & 2, p 1520.
27. *The merck index*. 14th ed.; Merck Research Laboratories: Whitehouse Station, N.J., USA, 2006.
28. *Who model list of essential medicines* March 2011.
29. Electronic Orange Book. *Approved drug products with therapeutic equivalence evaluations*. 32nd ed.; Office of Generic Drugs Center for Drug Evaluation and Research, Food and Drug Administration: http://www.fda.gov/downloads/Drugs/InformationOnDrugs/UCM086233.pdf, updated August 2012.
30. Pham The, H.; Gonzalez Diaz, I.; Bermejo Sanz, M.; Mangas Sanjuan, V.; Centelles, I.; Garriges, T.M.; Cabrera Perez, M.A. In silico prediction of caco-2 permeability by a classification qsar approach. *Mol. Inf.* **2011**, *30*, 376-385.
31. Kim, J.S.; Mitchell, S.; Kijek, P.; Tsume, Y.; Hilfinger, J.; Amidon, G.L. The suitability of an in situ perfusion model for permeability determinations: Utility for bcs class i biowaiver requests. *Mol. Pharmaceutics* **2006**, *3*, 686-694.
32. Netzeva, T.I.; Worth, A.P.; Aldenberg, T.; Benigni, R.; Cronin, M.; Gramatica, P.; Jaworska, J.S.; Kahn, S.; Klopman, G.; Marchant, C.A.*, et al.* Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ecvam workshop 52. *ATLA* **2005**, *33*, 1-19.

Information Signatures of Viral Proteins: A Study of Influenza A Hemagglutinin and Neuraminidase By

Samuel Barlow, Diego F. Cucalón, Daniel J. Graham * and Jordan C. Hauck

Department of Chemistry, Loyola University Chicago, 6525 North Sheridan Road, Chicago, IL 60626, USA

* To whom correspondence should be addressed: Phone: 1-773-508-3169; FAX: 1-773-508-3086; E-Mail: dgraha1@luc.edu.

Received: 30 October 2015 / Accepted: 30 October 2015 / Published: 2 December 2015

Abstract: Hemagglutinin (HA) and neuraminidase (NA) are glycoproteins encoded by several types of viral particles. Most notably, they exercise complementary chemical functions during infection and propagation of influenza A. This research focuses on the primary structure information of the proteins, applying a computational model from previous research. Data for multiple influenza A subtypes are illustrated via information signatures and phase plots. These illuminate new ways of evaluating molecules for their virulence potential. The results further point to mutation strategies for attenuating the functions.

Introduction

Hemagglutinin (HA) and neuraminidase (NA) are glycoproteins in the surface membrane of influenza particles [1]. Infection of a host is initiated by HA while NA catalyzes the release of newly-made viral particles [2]. The antibodies of the molecules form the means of classifying the influenza A subtypes: H1N1, H2N2, H3N2, etc. [3]. At present, there are at least 16 and 9 known subtypes for HA and NA, respectively. Given the risks of viral exposure to global populations, intense effort is directed toward understanding the molecular mechanisms. Further, the design and formulation of drugs which subvert the mechanisms are on-going challenges [4].

Influenza HA and NA have presented thousands of variants. For example, two HA sequences are:

```
MKARLLILLCALSATDADTICIGYHANNST
DTVDTVLEKNVTVTHSVNLLEDSHNGKLC
RLKGIAPLQLGKCNIAGWILGNPECESLLSN
RSWSYIAETPNSENGTCYPGDFADYEELRE
QLSSVSSFERFEIFPKERSWPKHNITRGVTA
ACSHAKKSSFYKNLLWLTEANGSYPNLSKS
YVNNKEKEVLVLWGVHHPSNIEDQRTLYR
KENAYVSVVSSNYNRRFTPEIAERPKVRGQ
AGRMNYYWTLLEPGDKIIFEANGNLIAPW
YAFALSRGLGSGIITSNASMDECDTKCQTP
QGAINSSLPFQNIHPVTIGECPKYVRSTKLR
MVTGLRNIPSIQSRGLFGAIAGFIEGGWTG
MVDGWYGYHHQNEQGSGYAADQKSTQN
AINGITNKVNSVIEKMNTQFTAVGKEFNKL
EKRMENLNKKVDDGFLDIWTYNAELLVLL
ENERTLDFHDSNVKNLYEKVKNQLRNNAK
EIGNGCFEFYHKCDNECMESVKNGTYDYP
```

KYSEESKLNREKIDGVKLESMGVYQILAIY
STVASSLVLLVSLGAISFWMCSNGSLQCRIC
I

Seq. (1)

MEARLLVLLCAFAATNADTICIGYHANNST
DTVDTVLEKNVTVTHSVNLLEDSHNGKLC
KLKGIAPLQLGKCNIAGWLLGNPECDLLLT
ASSWSYIVETSNSENGTCYPGDFIDYEELRE
QLSSVSSFEKFEIFPKTSSWPNHETTKGVTA
ACSYAGASSFYRNLLWLTKKGSSYPKLSKS
YVNNKGKEVLVLWGVHHPPTGTDQQSLY
QNADAYVSVGSSKYNRRFTPEIAARPKVRD
QAGRMNYYWTLLEPGDTITFEATGNLIAP
WYAFALNRGSGSGIITSDAPVHDCNTKCQT
PHGAINSSLPFQNIHPVTIGECPKYVRSTKL
RMATGLRNIPSIQSRGLFGAIAGFIEGGWTG
MIDGWYGYHHQNEQGSGYAADQKSTQNA
IDGITNKVNSVIEKMNTQFTAVGKEFNNLE
RRIENLNKKVDDGFLDIWTYNAELLVLLEN
ERTLDFHDSNVRNLYEKVKSQLKNNAKEI
GNGCFEFYHKCDDACMESVRNGTYDYPK
YSEESKLNREEIDGVKLESMGVYQILAIYST
VASSLVLLVSLGAISFWMCSNGSLQCRICI

Seq. (2)

Two NA sequences are:

MNPNQKIITIGSICMAIGTISLILQIGNIISIWV
SHSIQTGSQNHTGICNQRIITYENNTWVNQT
YVNISNTNVVAGKDTTSMILAGNSSLCPIR
GWAIYSKDNSIRIGSKGDVFVIREPFISCSHL
ECRTFFLTQGALLNDKHSNGTVKDRSPYRA
LMSCPIGEAPSPYNSRFESVAWSASACHDG
MGWLTIGISGPDDGAVAVLKYNGIITEIIKS
WRKQILRTQESECVCVNGSCFTIMTDGPSD
GPASYRIFKIEKGKITKSIELDAPNSHYEECS
CYPDTGKVMCVCRDNWHGSNRPWVSFNQ
NLDYQIGYICSGVFGDNPRPKDGKGSCDPV
NVDGADGVKGFSYRYGNGVWIGRTKSNSS
RKGFEMIWDPNGWTDTDGNFLVKQDVVA
MTDWSGYSGSFVQHPELTGLDCMRPCFWV
ELIRGRPREKTTIWTSGSSISFCGVNSDTVN
WSWPDGAELPFTIDK

Seq.
(3)

MNPNQKIITIGSICMVVGIISLILQIGNIISIWV
SHSIQTGNQNHPETCNQSIITYENNTWVNQ
TYVNISNTNVVAGQDATSVILTGNSSLCPIS
GWAIYSKDNGIRIGSKGDVFVIREPFISCSHL
ECRTFFLTQGALLNDKHSNGTVKDRSPYRT
LMSCPVGEAPSPYNSRFESVAWSASACHD
GMGWLTIGISGPDNGAVAVLKYNGIITDTI
KSWRNNILRTQESECACVNGSCFTIMTDGP
SNGQASYKILKIEKGKVTKSIELNAPNYHY
EECSCYPDTGKVMCVCRDNWHGSNRPWV
SFDQNLDYQIGYICSGVFGDNPRPNDGTGS
CGPVSSNGANGIKGFSFRYDNGVWIGRTKS
TSSRSGFEMIWDPNGWTETDSSFSVRQDIV
AITDWSGYSGSFVQHPELTGLDCMRPCFW
VELIRGQPKENTIWTSGSSISFCGVNSDTVG
WSWPDGAELPFSIDK

Seq. (4)

The sequences offer detailed information. Yet a computer-unassisted reading of them is bewildering. This is apparent because, among other things, one cannot distinguish the extraordinary from ordinary. The above include formulae allied with the “Spanish flu” pandemic of 1918 [5]. But which ones are these? The correct answers are Seqs. (2) and (4). The reader's uncertainty is understandable given the lengths and complexities of the sequences.

Our approach to proteins has looked for guidance from information theory [6 - 10]. Here we focus on the HA and NA primary structure information. The results draw contrasts between seasonal molecules and ones with high virulence potential. The data further point to mutation strategies for re-directing and possibly attenuating the functions.

Proteins and Sequence Information

The approach builds on research from the mid-2000s. Work in this lab quantified the correlated information *CI* expressed by the naturally occurring amino acids based on their atom and covalent bond structure [6, 8]. An average < *CI* > and standard deviation σ_{CI} were established and a dimensionless quantity $Z_{CI}^{(i)}$ was based on each amino acid's *CI* contribution relative to the average *CI*, e.g.

$$Z_{CI}^{(W)} = +2.63 \qquad Z_{CI}^{(F)} = +0.691$$

$$Z_{CI}^{(M)} = -0.128 \qquad Z_{CI}^{(A)} = -0.476$$

There are twenty amino acids and thus sixteen more $Z_{CI}^{(i)}$ to note as in reference [6]. The superscript symbols refer to the amino acid while the numerical value represents the *CI* distance from the average in standard deviation (σ_{CI}) units. The sign reflects whether the amino acid contributes information above or below the natural average. The *Z*-terms largely follow chemical intuition. Tryptophan (W) features a network of aromatic bonds and functional groups; it exerts nearly $+3\sigma_{CI}$ impact in a protein. Alanine (A) is a simple aliphatic and contributes *CI* below average at ca. $-0.5\sigma_{CI}$. The methodology originated in an information study of ribonuclease A and lysozyme [8, 9].

A dimensionless function *G*(*k*) is constructed for a sequence by adding $Z_{CI}^{(i)}$ in the N- to C-terminal order; *k* is a counting index less than or equal to *N* number of residues in the protein. *G*(*k*) tracks the accumulation and fluctuations of information:

$$G(k) = Z_{CI,1}^{(i)} + Z_{CI,2}^{(i)} + Z_{CI,3}^{(i)} + \ldots + Z_{CI,k}^{(i)} = \sum_{j=1}^{k \le N} Z_{CI,j}^{(i)} \tag{1}$$

Proteins generally host a majority of low information residues. As a consequence, *G*(*k*) scales linearly with negative slope and is well accommodating of least squares analysis. The analysis establishes an ensemble of linear regression functions $L_j(k)$ with typical correlation coefficient $R^2 > 0.95$.

The ensemble leads to information signatures { $H_j(k)$ }:

$$\{ H_j(k) \} = \{ G(k) - L_j(k) \} \tag{2}$$

As examples, *G*(*k*), $\{H_j(k)\}$for Seqs. (1) and (4) appear in **Figure 1**. The proteins originate from a human host H1N1 isolate harvested in Albany, New York in 1951. The accession details are: gb:CY021821|gi:145279077|UniProtKB:A4U7A6|.

Graphs such as in **Figure 1** serve as signatures of the primary structure information. They reflect more than a molecule's local composition. If a substitution is made at site *j*, the collection in

Eq. (2) is altered. The amplitude is impacted at all sites k = 1, 2, …, N. The information signatures can be strikingly different, depending on the subtype. There are as many signatures as there are HA and NA variants.

In thermodynamics, the variance of an extensive property such as enthalpy and entropy scales with a capacity [11]. In the same way, the variance in $H_j(k)$ can be viewed in terms of a protein's functional capacity. Molecules composed of only one type of amino acid, e.g. AAAAAAAAA…., offer zero capacity. They are of no biochemical utility because they lack diversity of information. This is borne out in the signatures: their $G(k)$ trace perfect lines (R^2 = 1.000); corresponding $H(k)$ express zero amplitude.

The information signature variance is calculated as follows:

$$\sigma_H^2 = \langle H^2 \rangle - \langle H \rangle^2 \qquad (3)$$

The square root σ_H is the standard deviation, so indicated in **Figure 1** by the vertical black arrows. Linear regression computes an ensemble of $H(k)$ functions from $G(k)$, and accordingly, a distribution of σ_H. Capacities form robust descriptors of thermodynamic systems; σ_H play equally vital roles regarding protein information.

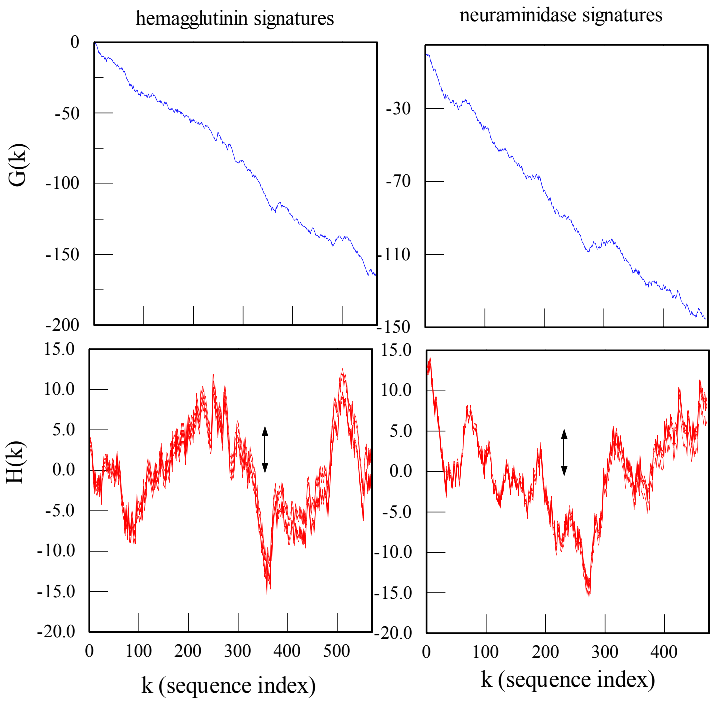


Figure 1. Information Functions for HA and NA. Plots of $G(k)$ and $H(k)$ derive from Seqs. (1) and (3) of the **Introduction**. The vertical arrows in the lower panels indicate the standard deviation in the fluctuations of $H(k)$.

Results

Nearly sixty thousand primary structures were analyzed; the HA and NA sequences were obtained as FASTA downloads from the Influenza Research Database (IRD). The data were catalogued according to viral subtype: H1N1, H2N2, H3N2, and so forth. For every molecule, $G(k)$, { $H_j(k)$ }were established along with distributions of the variance and standard deviation. A viral isolate locates a coordinate (plus error bars) on a σ_{HA}, σ_{NA} phase plot. Multiple variants establish a neighborhood of points for a subtype. The distance from one neighborhood to another is dictated by the information effects of antigenic drift and genome re-assortments.

The phase plot for influenza A is illustrated in **Figure 2**. Each filled circle marks the average σ_{HA}, σ_{NA} for the labeled subtype while the bars mark the standard deviations about the averages. Several bars are of widths less than the symbols. These reflect that a sparse number of isolates was available for analysis. That being said, every attempt was made to be exhaustive. **Figure 2** derives from HA and NA across a spectrum of hosts: human, avian, equine, bat, etc.. There will be more neighborhoods to map as new subtypes and hosts are discovered. **Figure 2** teaches two things, the first concerning boundaries. There are astronomical possible variants of HA and NA. The ones selected for viral infection and propagation are concentrated in the following ranges:

$$2.5 \le \sigma_{NA} \le 6.0$$

$$3.8 \le \sigma_{HA} \le 8.0$$

Note the ranges to be substantive despite the intense selection pressure to preserve the protein functions.

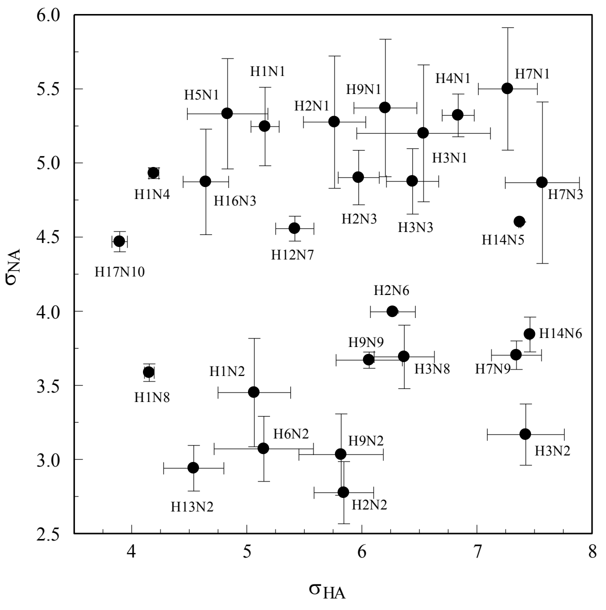


Figure 2. Phase Plot for Influenza HA and NA. Each filled circle marks the average σ_{HA}, σ_{NA} for the labeled subtype. The error bars mark the standard deviations about the averages.

The second lesson is that the neighborhood distribution is markedly uneven. A significant fraction of influenza subtypes clusters in the upper third of **Figure 2** while fewer ones occupy the lower third. Further, there are several low-density regions: these correspond to HA, NA variants which have yet to manifest, or are outright avoided by natural selection.

Information signatures discriminate the subtypes. What do things look like for proteins specific to human populations? For humans, the major circulating strains of influenza A have been H1N1, H2N2, and H3N2; the global pandemic of 1918 was attributed to the first of these [5]. The avian strain H5N1 has rarely infected humans, although it poses high virulence potential.

Figure 3 shows a phase plot based on human host isolates. Different color symbols distinguish the subtypes while the isolate years are included. Not all years are represented as the analysis was directed to complete genomes. The point locus for the 1918 pandemic year sample (Brevig Mission, >gb:AF250356|gi:8572169|UniProtKB:Q9IGQ6|) is marked in red. It is considerably removed from the H1N1 neighborhood. Its nearest neighbors derive from H5N1 isolates.

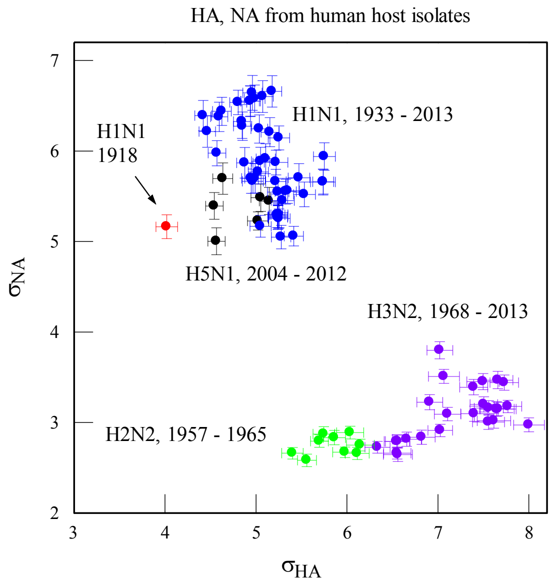


Figure 3. Phase Plot for Human Host HA and NA. The blue symbols are placed by H1N1 samples collected between 1933 and 2013. The red symbol is placed by HA and NA from pandemic year 1918. The black, green, and violet symbols are placed, respectively, by H5N1, H2N2, and H3N2 samples.

Discussion

HA and NA exercise complementary functions: the former enables attachment of influenza particles to a cell surface while the latter catalyzes the release [1, 2]. Given the essentialness of the functions, there is significant pressure for variants to manifest over time. Variants enable the virus to sidestep host immune responses and to thwart drug therapy. There are >10^5 HA and NA sequences on record, yet this is a paltry number compared to the possibilities.

Thermodynamic analysis of a system commences with variables and functions of state. This has been the approach to HA and NA in constructing information *G* and *H*. The functions track the information accumulation and fluctuation in a manner dependent on *all* the amino acids. No one site or region is viewed as more important than others.

One learns several things, the first being an information method of evaluating HA, NA pairs. All sequences are confounding by their complexity. However, using a spreadsheet and $Z_{CI}^{(i)}$ look-up table, it is straightforward to compute *G*-functions plus regression *L* and signature *H*. The signatures lead immediately to σ_{HA}, σ_{NA} and the neighborhood of the phase diagram. To be sure, the evaluation is not without tentativeness given the overlap of neighborhoods. However, the analysis points to the information and virulence similarities of the viral subtypes.

The second insight is the contrast between seasonal- and pandemic-year proteins. HA and NA from the 1918 pandemic places an outlier point on the phase plot. This placement stems from a lower fluctuation amplitude, compared with that of seasonal proteins. This suggests that lower chemical noise in the primary structures underpins a more invasive chemical function.

The third insight is a strategy for re-directing—and possibly attenuating—the functions. Natural selection favors molecules which promote viral infection and suppresses variants that serve otherwise. We conjecture that the latter type place state points in the low density regions of **Figure 2**.

Figure 4 revisits **Figure 2** and includes four pathways. Each tracks a succession of substitutions on Seqs. (1) and (4). With each substitution, there is a displacement of the σ_{HA}, σ_{NA} coordinate. As the primary structures belong to the H1N1 subtype, each pathway commences near the center of the H1N1 neighborhood and terminates in a zero-to-low density region. The paths are annotated by the following ordered-pair sequences:

Pathway 1	**Pathway 2**	**Pathway 3**	**Pathway 4**
HA NA	**HA NA**	**HA NA**	**HA NA**
E199G, D199V	P504A, H144K	Y209N, C279A	E51R, T362I
F432V, D329C	Y246K, V291N	L335T, D79W	R238K, T381G
R514I, C238F	P135Y, G342R	S179K, A271K	G411V, R52W
S275Y, S166P	A302S, K369R	K328N, L127H	K511Q, I20S
E260D, P326D	W553M, N235T	A379H, M188C	L547C, K150F
E449G, P337T	W553K, G333H	V428I, W458E	D456W, V149W
L250F, H126A	I530E, Y208L	R344Q, S168H	G76W, L22Y
N177T, H185M	L512A, P328K	N448D, V75K	Q386I, G109T
L417M, N141P	S160N, N171Q	L37M, P169R	A156D, N325I

The pathways (and countless more) are readily charted using a forced random walk algorithm. One selects a target locale on the phase plot. The sequences are then subject to trial substitutions. With each trial, *G*, *H*, and σ_{HA}, σ_{NA} are computed. The substitutions are accepted if the state point is inched closer to the target and declined otherwise. Each pathway in **Figure 4** is traversed via nine pair-substitutions. This demonstrates that the proteins do not have to be radically altered for the information signatures to move out of the virulent neighborhood of origin. In **Figure 4**, pathways 1 and 2 direct HA and NA away from *all* the subtype neighborhoods. In contrast, pathways 3 and 4 cross territory allied with highly virulent subtypes. In re-directing HA and NA functions, the upward-going pathways 1 and 2 would seem preferable. Molecules with information removed from the active neighborhoods would likely offer diminished potency, yet stimulate some production of host antibodies. This would enhance the overall immunity of host populations.

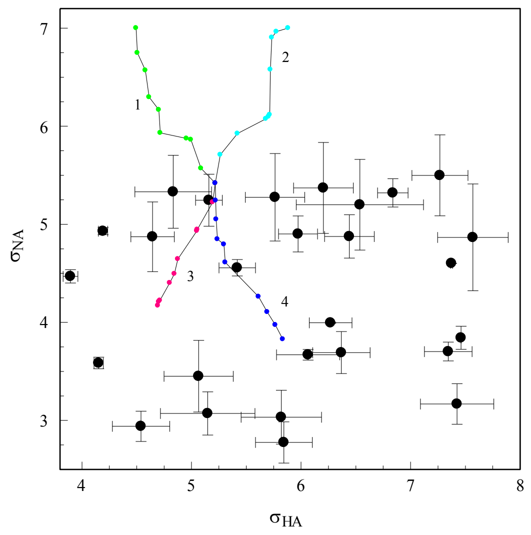


Figure 4. Pathways for Re-locating Phase Points. Each pathway tracks a succession of amino acid substitutions on the H1N1 Seqs. (1) and (3) of the **Introduction**. The pathways are annotated above.

Summary and Closing

The primary structure information expressed in influenza HA and NA was investigated using a model established in previous research. The model enabled computation of signatures based on the accumulation and fluctuation of information. The signatures were encapsulated in *G*, *H*-functions and phase plots. These illuminated information methods for discriminating variants and attenuating the molecular functions.

Acknowledgements

The authors appreciate the input of Donald May about information theory in reference to biological systems. The authors equally appreciate the insights of Miguel Ballicora about natural selection and viral systems. The authors thanks Loyola University Chicago for support of this research.

References

1. Levine AJ (1992) Viruses, Scientific American Library, Chapter 8.
2. Voyles BA (1993) The Biology of Viruses, Mosby-Year Book, St. Louis.
3. Kawaoka Y, Neumann G in Influenza Virus: Methods and Protocols (2012), Kawaoka Y, Neumann G eds, Humana Press, New York, Chapter 1.
4. Roberts NA (2001) Prog. Drug Res. 56:195-237.
5. Kolata GB (1999) Flu: the Story of the Great Influenza Pandemic of 1918 and the Search for the Virus That Caused It, Farrar, Straus, and Giroux, New York.
6. Graham DJ, Malarkey C, Schulmerich MV (2004) J Chem Info Comp Sci, 1601.
7. Graham DJ (2013) Prot J 32:275-287. 10.1007/s10930-013-9485-2.
8. Graham DJ, Greminger JL (2009) Mol Divers 10.1007/s11030-009-9211-3.
9. Graham DJ, Greminger JL (2011) Mol Divers 10.1007/s11030-011-9307-4.
10. Graham DJ, May D, Grzetic S, Zumpf J (2012) Prot J 31:550-563. 10.1007/s10930-012- 9432-7.
11. Goodstein DL (1985) States of Matter, Dover, New York, Chapter 1.

SciForum
Mol2Net

Genome-Wide Discriminatory Information Patterns of Cytosine DNA Methylation

Robersy Sanchez * and Sally A. Mackenzie *

N300 Beadle Center, University of Nebraska, Lincoln, NE 68588

* Author to whom correspondence should be addressed; E-Mails: robersy@unl.edu (R.S.); sally.mackenzie@unl.edu (S.A.M.)

Received: 5 October 2015 / Accepted: 5 October 2015 / Published: 2 December 2015

Abstract: Cytosine DNA methylation (CDM) is a highly abundant epigenetic heritable but reversible chemical modification to the genome. Herein, a machine learning approach, was applied to analyze the accumulation of epigenetic marks in 150 methylomes from *Arabidopsis thaliana* ecotypes. We hypothesize that these marks are chromosomal footprints that account for different ontogenetic and phylogenetic and histories of individual members of the sampling population. Our results support this hypothesis and suggest a statistical-physical relationship between CDM changes and single nucleotide polymorphism (SNPs). Furthermore, the genome-wide redistribution of CDM changes ensures the thermal stability of the DNA molecule preserving the integrity of the genetic message continuously stressed by thermal fluctuations in the cell environment.

Keywords: Epigenetics, Epigenomics, Information thermodynamics, Linear discriminant analysis

1. Introduction

Cytosine DNA methylation (CDM) is one of the molecular processes that result in epigenetic modifications to the genome. In particular, cytosine methylation is a widespread regulatory factor in living organisms. Changes introduced by DNA methylation can be inherited from one generation to the next. Some methylation changes can regulate gene expression and cause genomic imprinting [1,2]. Cytosine methylation arises from the addition of a methyl group to a cytosine's C5 carbon residue. Distinct pathways regulate methylation status by the action of methyltransferases [3]. The addition or removal of a methyl group to a cytosine C5 residue produces a change of information that is recognized by the molecular transcription machinery and can be verified by current sequencing technologies [2]. However, it is still undefined whether or not the observed methylation changes could be linked to genome-wide information patterns.

The development of DNA bisulfite conversion methodology coupled with next-generation sequencing approaches (Bis-seq) allows determination of the methylation status of nearly every cytosine in a genome. In this way, the methylation status of particular cytosine sites is often expressed in terms of methylation level $p_i = \#C_i / (\#C_i + \#nonC_i)$, where $\#C_i$ and $\#nonC_i$ represent the numbers of methylated and non-methylated read counts observed at the genomic coordinate i, respectively. At tissue level, the methylation status (methylated or non-methylated) of cytosine C_i at the genomic coordinate i can be analyzed as a random variable that takes value "methylated" with probability p_i and "non-methylated" with probability $1 - p_i$. Then, the formula $H(p(x_i)) = -\sum_i p(x_i) log\ _2 p(x_i)$ (1) of Shannon's entropy of a random event with probability distribution $p(x_i)$ can be applied to estimate the uncertainty of the methylation events at given cytosine site i as: $H(C_i) = -p(C_i) log_2 p(C_i) - (1 - p(C_i)) log_2 (1 - p(C_i))$ (2).

The entropy defined by Eq. 2 is therefore the expected value of the logarithm base 2 of the methylation level [4]. Assuming that, as a result of variations in environmental conditions, a change of methylation status in a genomic region *R* takes place, the uncertainty decrease in the genomic region *R* leads to a gain of information given by:

$$I_R = -\left(\sum_{i \in R} H\left(C_i^{after}\right) - \sum_{i \in R} H\left(C_i^{before}\right)\right) \qquad (3)$$

Where C_i^{before} and C_i^{after} stand for the methylation status before and after the variations of environmental conditions, respectively [5]. Eq.3 expresses an information theoretical derived concept with a thermodynamic and biophysical meaning [5,6].

Herein, our study is focussed in the analysis of the genome-wide CDM information patterns induced by the changes in the enviromental conditions. In particular, we analyzed whether or not these information patterns carry discriminatory information in the form of chromosomal footprints.

2. Results and Discussion

The genome-wide evaluation of Eq.3 indicates the existence of methylation hotspots along the chromosomes (Fig.1). Genomic regions (GRs) can be classified according to the value of the I_R as: 1) highly variable methylation regions, 2) variable regions, and 3) low variable or constant regions. The regions with information gain (orange to black lines in the heatmaps color bar) or loss (light yellow to sky-blue) (Fig. 1) are observed at specific positions with a high line density in the pericentromeric region. The lines in yellow correspond to regions where the difference between the entropies $H_R^{ecotype}$ and H_R^{Col-0} is close to zero. According to Eq. 3, methylation hotspots are the ecotype chromosomal regions with a remarkable decrease in uncertainty with respect to Col-0.

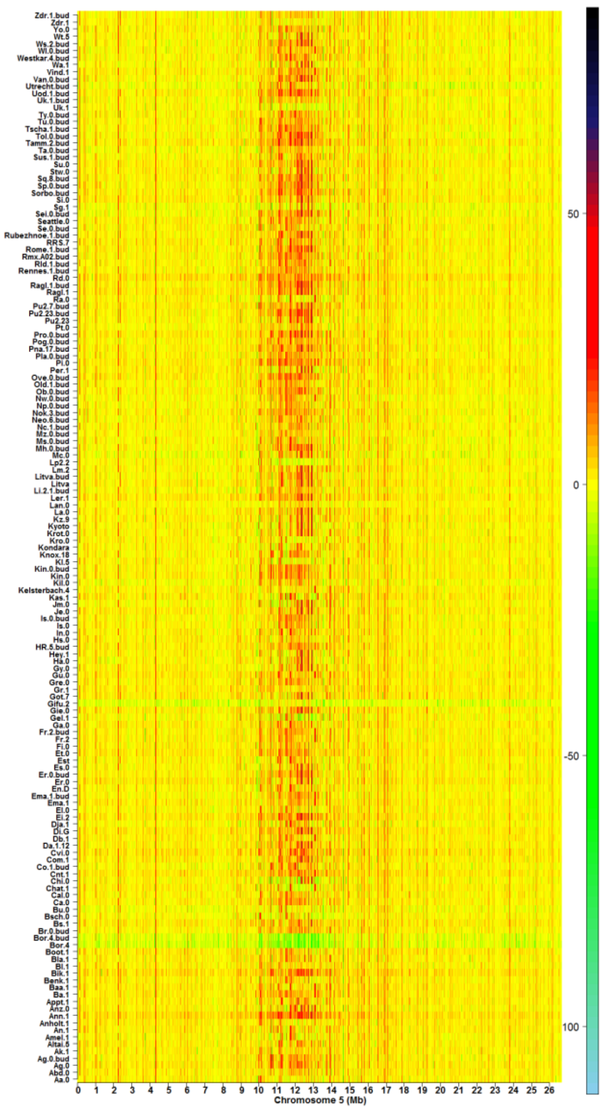
Methylation hotspots shared by a set of individuals at fixed chromosomal positions suggest the existence of specific informative landmarks (Fig. 1 and 2). That is, most of the CDM changes observed in natural variation and silencing mutants occur at specific methylation GRs, which are delineated in the heatmaps as chromosomal landmarks. These landmarks frequently cover transposable elements (TEs) and protein-coding regions (Fig. 2).

Discriminatory Informative Patterns in natural *Arabidopsis* ecotypes

The heatmaps suggest the existence of specific landmark informative patterns in all chromosomes across the ecotype samples that may or may not be shared by several individuals. These patterns comprise chromosomal regions carrying discriminatory information. That is, it is possible to distinguish between the individuals and among subsets of individuals by considering their discriminatory information patterns.

Figure 1. Methylation hotspots along chromosome 5 from 150 *Arabidopsis thaliana* ecotypes [7]. The color bar indicates the

magnitude of I_R values.



Applying hierarchical clustering based on the levels of C-DMRs, Schmitz *et al.* [7] found that the 150 *Arabidopsis thaliana* ecotypes from North America and Asia reflect their geographical distributions. Herein, the consecutive application of principal component analysis (PCA) and linear discriminant analysis (LDA) to the same ecotype set supports the hypothesis that the landmark patterns constitute chromosomal footprints that may account for ontogenetic and phylogenetic differences among individuals (Fig.3 A and C). This analysis supports not only the ecotype classification according to their geographical location for North America and Asia [7]), but also for all geographical regions.

These footprints are not only connected to the environment, but also to the single nucleotide polymorphisms (SNPs) detected throughout the DNA sequences (Fig.3 B and D). The classification of the *Arabidopsis thaliana* ecotypes according to their geographical distribution was retrieved not only from their landmark patterns, but also from their SNPs patterns (Fig.3). A summary of the classification results is presented in Table 1.

The similarity between the hierarchical clusters suggests that some statistical-physical relationship must exist between the SNPs and methylation changes. The two (2D) and three-dimensional (3D) kernel density plots presented in Fig.4 support the last hypothesis. The 2D kernel density plots indicate that the frequency of normalized read-counts supporting SNPs decrease with the increment of methylation changes, expressed here in terms of gain or loss of information I_R (Fig. 4A). This statistical trend is emphasized in the empirical 3D kernel density plots (Fig. 4B) and in the modelled Farlie-Gumbel-Morgenstern copula distribution built from the non-linear fit of the marginal distributions (Fig. 4C).

Figure 4 suggests that most of the observed CDM changes tend to preserve the integrity of the message carried by the DNA molecule, which is challenged by thermal fluctuations in the cell environment. This is consistent with the report that CDM changes alter the mechanical properties of the DNA molecule [8]. Thus, a statistical-physical relationship between CDM changes and SNPs is expected. Indeed, depending on the DNA sequence context, the addition or removal of a methyl group to a cytosine residue could increase or decrease the local thermodynamic stability of the DNA molecule and the nucleosomes [8–12]. The density plots of the experimental data indicate that the greatest frequency of SNPs is found in those GRs where the methylation status remains unchangeable with respect to the control (Col-0, Fig. 4).

Figure 2. Annotation of several hotspots on chromosome 2 from eigth *Arabidopsis* gene silencing mutants involving methylation.

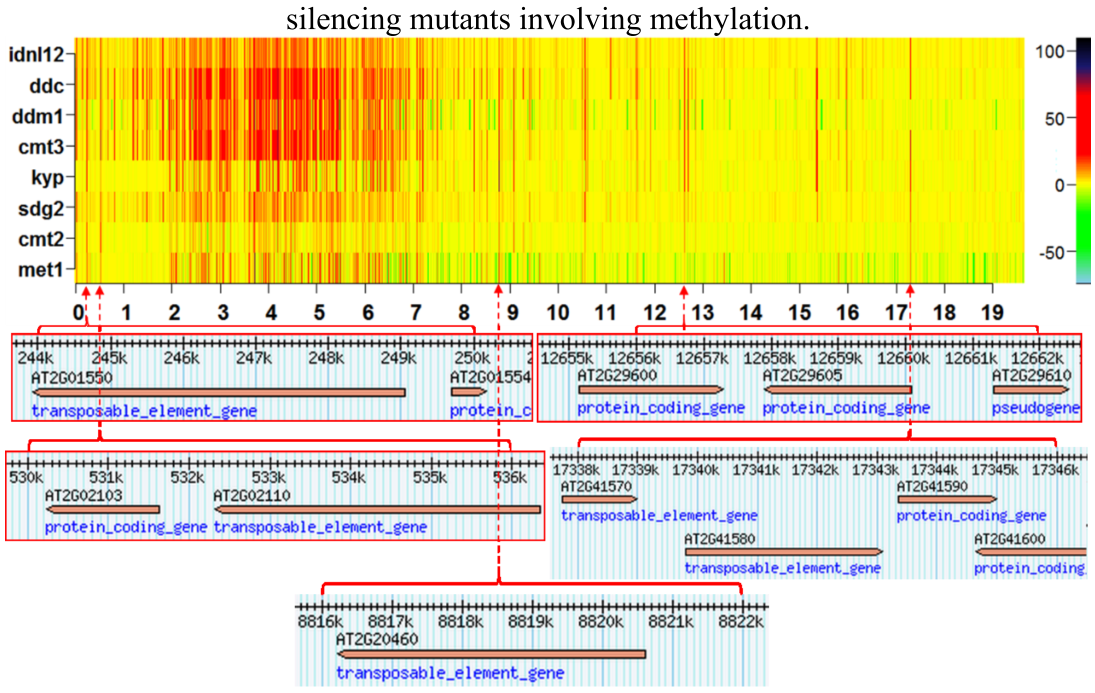


Figure 3. Classification of the Arabidopsis thaliana ecotypes according to their geographical distribution. **A** and **B**, LDAs based on I_R and SNPs, respectively. **C** and **D**, fan dendrograms based on the individual coordinates estimated from the LD functions. The dendrograms were built by applying hierarchical clustering with Euclidean distance and UPGMA as agglomeration method.

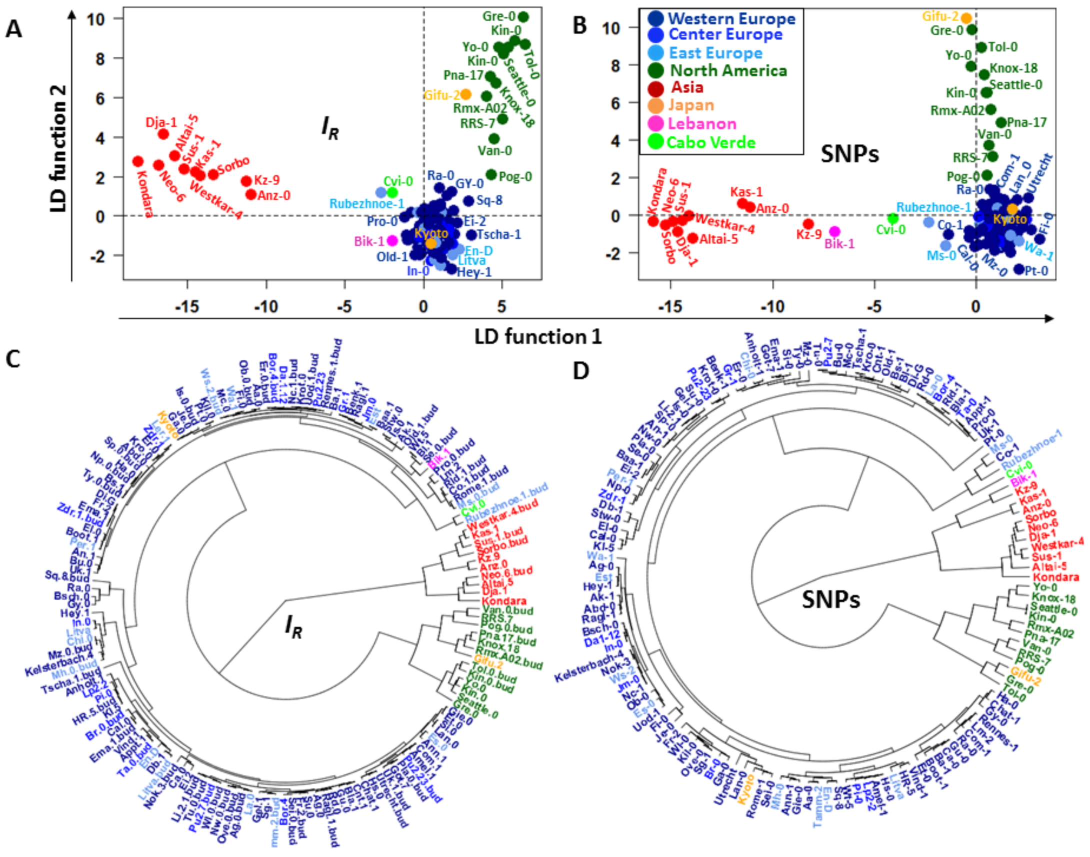
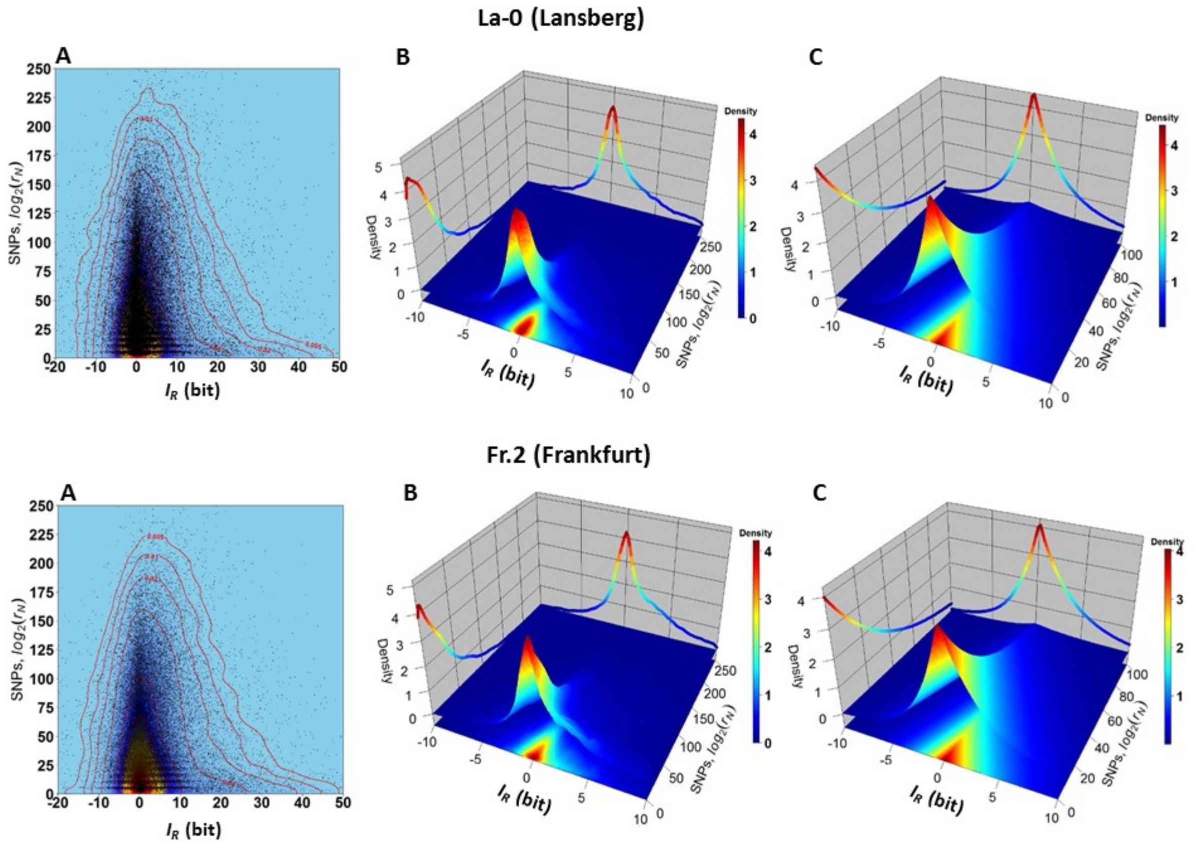


Table 1. Performance of the classifications presented in Fig. 3.

Sample [a]	Classifier	Accuracy Mean	2.5% quantile	97.5% quantile
CG ecotypes (2482 DIRs)	AUC+PCA+LDA	93.08352	88.05678	97.4359
	AUC+PCA+SVM	93.52517	91.83673	95.2381
	AUC+SVM	96.42381	95.91837	96.59864
SNPs ecotypes (2590 DIRs)	AUC+LDA	90.85758	85.42125	95.89744
	AUC+PCA+SVM	95.01642	94.02985	96.26866
	AUC+SVM	95.77007	95.23810	95.91837

[a] 1000 ten-fold cross-validations were performed for each classifier.

Figure 4. A: 2D kernel density plot. B: 3D kernel density plot. C: 3D plot of the density probability distribution of the Farlie-Gumbel-Morgenstern copula built from the non-linear fit of the marginal distributions estimated for LC_R (a Weibull PDF) and I_R (a Skew-Laplace PDF). These estimations were performed for several *Arabidopsis* ecotypes. The results for the ecotypes La-0 and Fr.2 are shown.

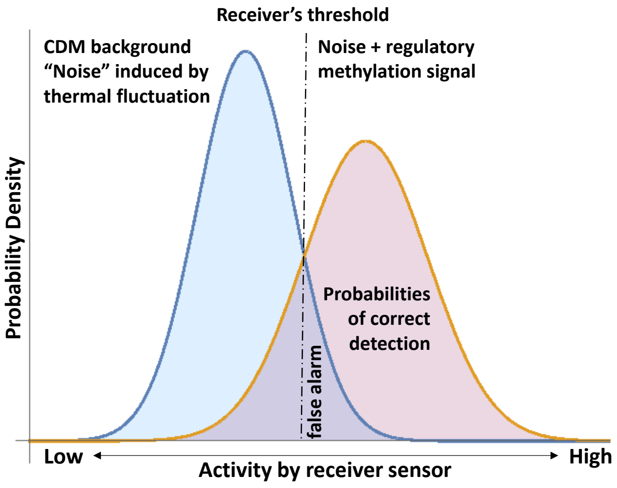


Hence, for an *Arabidopsis* plant, the adaptation to a new environment implies a genome-wide redistribution of CDM changes that will ensure the thermal stability of DNA. These are frequent methylation changes, which dynamically can vary from cell to cell in the same tissue. CDM changes induced by thermal fluctuations are the simplest natural explanation to the "spontaneously occurring variations" of DNA methylation in *Arabidopsis thaliana* plants propagated by single-seed descent throughout generations [13,14].

An important subset of CDM changes regulates the process of gene expression and functional adaption to the environment [12]. These are specific molecular signals from the regulatory methylation machinery. At this point, the challenge is whether or not we would be able to sort out the regulatory methylation signals from the CDM background ("noise") induced by thermal fluctuations. This challenge has been already confronted (although in a different field, see references [15,16]) and a concrete application in the context of CDM is illustrated in Fig. 5. It is not possible to separate the regulatory methylation signal from the CDM background induced by thermal fluctuation. Even a simple regulatory methylation change could alter the mechanical properties of the DNA molecule [2,8,10] and, consequently, it could require an additional local readjustment. Therefore, the receiver (a device used by the experimenter to detect the signal) must set up a criterion for response, in this case, a threshold level of activity in its sensor (i.e., a function of the methylation levels). This threshold in combination with the PDFs for noise and signal plus noise determine the probabilities of correct detection [17] (Fig. 5).

Hence, any statistical analysis of the regulatory signals of CDM changes must consider the statistical thermodynamics subjacent to the methylation process. This concept conveys a suitable approach to discriminate the regulatory signals from the "noise" induced by the thermal fluctuations.

Figure 5. Signal detection in noise according to reference [15,16] and, here, applied to the detection of regulatory CDM signals.


Receiver's threshold
CDM background "Noise" induced by thermal fluctuation
Noise + regulatory methylation signal
Probability Density
Probabilities of correct detection
false alarm
Low
Activity by receiver sensor
High

3. Materials and Methods

Equation 3 was used to compute the I_R for several samples with methylation data available in online databases (see below).

Arabidopsis thaliana methylation data

According to Eq. 3, I_R is computed for a subject sample with respect to a given reference sample. The I_R values were computed for 150 Arabidopsis ecotypes [7]. The TSV files taken from NCBI GEO under accession GSE43857 [7] were read and transferred to R software version 3.2.1 [18] by using the Bioconductor (version 2.14) R-package *GenomicFeatures* [19]. Ecotype Col-0 was used as reference (152 ecotypes including Col-0). The mutant data used in Fig. 2 were reported in reference [20] (GEO accession numbers GSE39901).

Machine learning approach

To test the hypothesis that different environmental conditions must leave different landmark patterns on chromosomes, a machine learning approach was followed.

The estimation of the area under the ROC curve (AUC) for the current multiple-class classification problem was performed according to reference [21] and applied to reduce the space dimension and to detected potential discriminant informative regions. This method was applied by using the R-package *HandTill2001*. Principal component analysis (PCA) was also used to reduce space dimensions.

AUC and PCA outputs were used with two classifiers: linear discriminant analysis (LDA) and support vector machine (SVM). These computations were performed by using the R-packages *adegenet* and *e1071*, respectively.

Logarithm of the normalized reads counts

The lists of SNPs and 1-bp deletions with a quality score of 25 and above of each ecotype samples were taken from 1001 Genomes Data Center (http://1001genomes.org/datacenter/; http://1001genomes.org/data/Salk/releases/). For a given number of non-repetitive reads supporting the base substitution (r), the normalized reads counts (r_N) were estimated as $r_N = r\ Concordance$, where $Concordance$ stand for the read ratios supporting a predicted feature to the total coverage. Next, the sum of logarithm base 2 of DNA-base substitution counts at a given region R was computed as:

$$LC_R = \sum_{i \in R} log_2\left(r_{N_i}\right) \qquad 4.$$

4. Conclusions

The CDM changes observable at the heatmaps do not take place at random genomic regions, but at specific locations in chromosomes, hotspots of methylation changes, which are noticed as chromosomal landmarks in the heatmaps. The phylogenetic and ontogenetic history of each individual is reflected in the variations of landmark patterns, which like footprints carries discriminatory information about the individual.

Results indicate that, as a statistical tendency, most of the CDM changes preserve the thermodynamic stability of the DNA molecules. In addition, our study also leads to a new open practical problem: the discrimination between the regulatory methylation signals from the CDM background ("noise") induced by thermal fluctuations.

Acknowledgments

This work was supported by a grant from the Bill and Melinda Gates Foundation (OPP1088661) to S.M.

Conflicts of Interest

The authors declare no conflict of interest

References

1. Belanger AS, Tojcic J, Harvey M, Guillemette C (2010) Regulation of UGT1A1 and HNF1 transcription factor gene expression by DNA methylation in colon cancer cells. BMC Mol Biol 11: 9. doi:10.1186/1471-2199-11-9.

2. Dantas Machado AC, Zhou T, Rao S, Goel P, Rastogi C, et al. (2015) Evolving insights on how cytosine methylation affects protein-DNA binding. Brief Funct Genomics 14: 61–73. doi:10.1093/bfgp/elu040.
3. Law J a, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11: 204–220. doi:10.1038/nrg2719.
4. Shannon C. E (1948) A Mathematical Theory of Communication. Bell Syst Tech J 27: 379–423.
5. Schneider TD (1991) Theory of molecular machines. II. Energy dissipation from molecular machines. J Theor Biol 148: 125–137.
6. Tribus M, McIrvine EC (1971) Energy and Information. Sci Am 225: 179–188. doi:doi:10.1038/scientificamerican0971-179.
7. Schmitz RJ, Schultz MD, Urich M a, Nery JR, Pelizzola M, et al. (2013) Patterns of population epigenomic diversity. Nature 495: 193–198. doi:10.1038/nature11968.
8. Severin PMD, Zou X, Gaub HE, Schulten K (2011) Cytosine methylation alters DNA mechanical properties. Nucleic Acids Res 39: 8740–8751. doi:10.1093/nar/gkr578.
9. Římal V, Socha O, Štěpánek J, Štěpánková H (2015) Spectroscopic Study of Cytosine Methylation Effect on Thermodynamics of DNA Duplex Containing CpG Motif. J Spectrosc 2015: 1–8. doi:10.1155/2015/842810.
10. Yusufaly TI, Li Y, Olson WK (2013) 5-Methylation of cytosine in CG:CG base-pair steps: a physicochemical mechanism for the epigenetic control of DNA nanomechanics. J Phys Chem B 117: 16436–16442. doi:10.1021/jp409887t.
11. Portella G, Battistini F, Orozco M (2013) Understanding the connection between epigenetic DNA methylation and nucleosome positioning from computer simulations. PLoS Comput Biol 9: e1003354. doi:10.1371/journal.pcbi.1003354.
12. Flores KB, Wolschin F, Amdam G V (2013) The role of methylation of DNA in environmental adaptation. Integr Comp Biol 53: 359–372. doi:10.1093/icb/ict019.
13. Schmitz R, Schultz M, Lewsey M (2011) Transgenerational epigenetic instability is a source of novel methylation variants. Science (80-) 334: 369–373. doi:10.1126/science.1212959.
14. Becker C, Hagmann J, Müller J, Koenig D, Stegle O, et al. (2011) Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature 480: 245–249. doi:10.1038/nature10555.
15. Wiley RH (2006) Signal Detection and Animal Communication. Adv Study Behav 36: 217–247. doi:10.1016/S0065-3454(06)36005-6.
16. Wiley RH (2013) Signal Detection, Noise, and the Evolution of Communication. In: Brumm H, editor. Animal Communication and Noise. Berlin Heidelberg: Springer-Verlag, Vol. 2. pp. 7–31. doi:10.1007/978-3-642-41494-7.
17. Wiley RH (2013) A receiver–signaler equilibrium in the evolution of communication in noise. Behaviour 150: 1–37. doi:10.1163/1568539X-00003063.
18. R Core Team (2014) A language and environment for statistical computing.
19. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, et al. (2013) Software for computing and annotating genomic ranges. PLoS Comput Biol 9: e1003118. doi:10.1371/journal.pcbi.1003118.

20. Stroud H, Greenberg MVC, Feng S, Bernatavichute Y V, Jacobsen SE (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell 152: 352–364. doi:10.1016/j.cell.2012.10.054.
21. Hand DJ, Till RJ (2001) A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. Mach Learn 45: 171–186.

SciForum
Mol2Net

^{14}N NMR Spectroscopy Study of Binding Interaction between Sodium Azide and Hydrated Fullerene

Tamar Chachibaia[1,2,]* **and Manuel Martin Pastor**[3]

1. Department of Analytical Chemistry, Food Science and Nutrition, Faculty of Pharmacy, University of Santiago de Compostela, Spain
2. Department of Public Health and Epidemiology, Faculty of Medicine, Tbilisi State University, Georgia
3. Magnetic Resonance Unit, CACTUS, University of Santiago de Compostela, Spain

* Author to whom correspondence should be addressed; E-Mail: nanogeorgia@gmail.com.

Received: 1 November 2015 / Accepted: 4 November 2015 / Published: 4 December 2015

Abstract: The presence of human pharmaceutical compounds in surface waters is an emerging issue in environmental science. Low levels of many active pharmaceutical ingredients are detected in the aquatic environment as a result of pharmaco-chemical industrial waste spill-offs in draining water. In the manufacturing of pharmaceutical drug substances azides are used as reagents or when they are generated somehow in the synthesis, it may be necessary to demonstrate that these impurities are sufficiently removed to levels below an appropriate safety threshold. Sodium azide is an example of an azide for which the environmental exposure limits have been reasonably well characterized. The treatment of waste and industrial water can be conducted by removing dissolved materials and ions in water using membrane separation technology with ultra- and nanofiltration (NF) and reverse osmosis (RO) membranes. To achieve better effluent water quality, tertiary treatment with activated carbon adsorption is used. To analyze the risk of pharmaceuticals in the environment, a proposed validated methodology by NMR spectroscopy will support the evaluation of the eco-toxicological hazards during the early development process of pharmaceuticals.

Keywords: sodium azide; fullerene; 14N NMR spectroscopy; nanofiltration.

1. Introduction

The presence of pharmaceutical active compounds (PhACs) in the surface, drinking, and wastewaters is an emerging issue in environmental science. [1,2,3,4,5,6,7,8,9]. Low levels of many pharmaceutical active compounds are detected in the aquatic environment as a result of pharmaco-chemical industrial waste spill-off in draining water. It may be necessary to demonstrate that these impurities are sufficiently removed to levels below an appropriate safety threshold.

Sodium azide is an example of an azide for which the environmental exposure limits have been reasonably well characterized.

In the manufacturing of pharmaceutical drug substances azides are used in the synthesis, or they are generated as intermediate substance. Sodium azide (NaN_3) is widely used as starting molecule in the synthesis of Sartans, for the treatment of hypertension since the 1990s [10]. Some of these products have reached a market volume of several 100 t/a with an upward trend and are therefore considered as blockbusters.

Besides Sartant it is used in the synthesis other pharmaceuticals, such as Alfentanil (analgesic), Azosemid (diuretic), (anti-inflammatory) Broperamol and others.

There are required regulations for the analysis of pharmaceutical wastewater treatment and removal using membrane bioreactors (MBR). The waters treatment can be conducted by removing dissolved materials and ions in water using membrane separation technology with ultra- and nanofiltration (NF) and reverse osmosis (RO) membranes. [11]

To achieve better effluent water quality, tertiary treatment with activated carbon adsorption is used [12]. Activated carbon filters, which may contain fullerene retains extremely effective by mechanical filtration effect [13,14,15,16]. In 2009 by Chae and coworkers was developed technology of membrane coating by hydrated fullerene for improvement of filtration properties of membranes.

To analyze the risk of pharmaceuticals in the environment, proposed validated methodology by NMR spectroscopy will support the evaluation of the eco-toxicological hazards of PhACs during the early development process - the questions brought forward by many academic and regulatory scientists.

Aim:

Aim was to study deviation of signals of sodium azide obtained by ^{14}N NMR spectroscopy under influence of hydrated fullerene to examine binding properties of sodium azide with hydrated fullerene. In current study we propose innovative method for detection of sodium azide by ^{14}N NMR spectroscopy.

Background:

Detection and inactivation of sodium azide in the environment is global issue, due to its widespread use in many spheres of human activity, in pharmaco-chemical industry in the synthesis of pharmaceuticals, as well pesticides, direct use in agricultural sphere, as herbicide, pesticide and insecticide, wine fermentation inhibitor, in automotive industry in the content of detonators of airbags and bactericidal agent for inhibition of germ growth.

The Organization for Economic Co-operation and Development (OCDE) [17] has included sodium azide in the list of 5,235 High Production Volume Chemicals (HPV) with a production or import greater than 1,000 tons per year (McKeen, 2010).

Sodium azide (NaN_3) has inhibitory effect on heme-containing mitochondrial respiratory chain enzyme Cytochrome C Oxidase, which is cause of CNS anoxia and hypoxia in case of acute intoxication, while in case of chronic exposure to its lower doses long-term outcome is dementia.

Against NaN_3 not exists any antidote, and thus the only method of treatment remains hemodialysis affected patients. In case of chronic intoxication prevention is possible by administration of antioxidants to workers.

For prevention of sodium azide impact is important to decrease the risk of exposure, as from occupation workplace atmosphere and also from the environment.

The environmentalist and atmospheric scientists are concerned about the safety of the use of sodium azide. Despite the widespread opinion of proponents of sodium azide use in water and soil, arguing that this chemical undergoes rapid hydrolysis and degradation (Rodríguez-Kábana & Robertson, 2000, 2001). [18], their opponents (Betterton, 1999, 2003, 2010).[19] claimed that this is not exactly what it can be anticipated, since they discovered water and soil samples containing residual amounts of sodium azide.

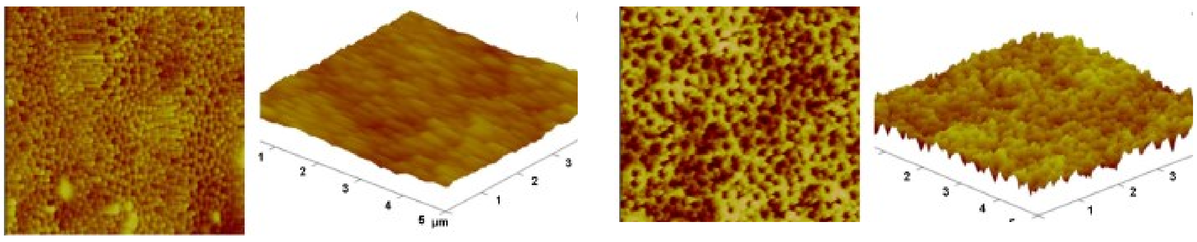
One of the most important issues is control of runoff waters and adequate membrane filtration barrier setup.

For industrial wastewater, as well municipal and hospital dialysis centers water treatment systems are using modern membrane filtration technologies.

Physicochemical modifications of membrane materials have been tried to improve performance of membrane processes for a long time. Numerous studies dealing with the surface modification of membranes have achieved by coating or grafting a functional group on the prepared membrane surface. [20].

In addition to the conventional ways of surface modification, researchers are focusing on the application of nanomaterials to modify the membrane properties thanks to recent developments of nanotechnologies. Among others, fullerene (C60) is the potential candidate expected to show an improved performance when used to modify the membrane properties. [21,22,23] Chae et al. (2009) examined the modification of ceramic microfiltration membranes coated by fullerene solution dip-coat-evaporation procedure. [24]. The dip coating procedure consisted of an initial immersion of membrane into solution for 2 seconds, drip-draining of excess solvent, followed by solvent evaporation under vacuum for several days. The surface concentration of C60 on the membranes was varied by repeating the dip-coating procedure anywhere from one to nine times. The final concentration of C60 on the membrane was determined by measuring the change in membrane weight. The surface morphology of the membrane coated with various amounts of C60 was investigated using a scanning probe microscope (SPM). The C60 nanoparticles used in this study were not chemically bound to the membrane surface.

Pic. 1. Surface morphology of the ceramic membranes: (left) anodisc 200 nm with C60 0.030mg cm^{-2} and (right) anodisc 200 nm with C60 0.058mg cm^{-2} (Source: doi:10.1016/j.memsci.2008.12.023. Courtesy Chae at el. 2009).



2. Materials and methods:

We studied binding properties between sodium azide and hydrated fullerene, without adding any catalyst, heating or microwave irradiation, under conventional conditions, to see if any interaction may occur to recommend in water filtration and pharmaceutical waste-water treatment applications, e.g. in different phases of filtration, as in membranes and carbon filters enriched by fullerenes.

The experimental part of this project is performed in the Magnetic Resonance Unit at the Center of Technology Innovation and Transfer (CACTUS) of the University of Santiago de Compostela. Experiments were conducted during 2012-2014 and obtained results analyzed.

University of Santiago de Compostela (USC) is equipped with NMR spectroscopy and propriety technology of MESTRE Labs, which is the software used worldwide.

The Magnetic Resonance Unit at the University of Santiago de Compostela provides the optimum research instrumentation required for this part of the project. The NMR facility provides three state-of-the-art high magnetic field NMR spectrometers of 500 MHz and 750 MHz.

Experimental

C60hyfn production, characterization and preparation of C60FWS

For C60FWS preparation (C60HyFn water solution), C60 fullerene samples with purity of more than 99.5% (MER Corporation, Tuscon, AZ, USA) have been used. C60FWS was produced without using of any solubilizers or chemical modification [25].

C60HyFn concentration of 8.88×10–4 M was used as stock solution for preparing C60FWS prior the experiment. This method is based on transferring of fullerene from organic solution into the aqueous phase with the help of ultrasonic treatment. To obtain C60FWS is possible with C60 concentration up to 5.5×10–3 M (~4 mg/ml).

Titration:

We performed ^{14}N NMR spectroscopy of pure sodium azide water solution and obtained satisfactory results with visualization of two peaks corresponding to three atoms of nitrogen with chemical shifts corresponding to 204.78 ppm and 56.06 ppm.[26]

We added hydrated fullerene 50 mM solution (144 mg/l) (IPAC, Ukraine, Kharkov) to sodium azide water solution with titration.

In our study we performed two series of experiments, with different concentrations of sodium azide. First, with molar concentration of 1M, and second, with 10M solution.

We performed titration by fullerene water solution with decreasing concentrations. Thus, ratios of NaN_3:C60 were <10:1 in the first set of experiment. In the first series of experiment we used standard addition method of titration. Standard is hydrated fullerene C60 and with decreasing the ratios of NaN_3:C60 which were subsequently 10:1, 1,36:1, 0,88:1 and 0,38:1.

In the second set of experiments we used higher concentration of sodium azide 10M with ratio of NaN_3:C60 which was equal to 100:1.

3. Results

The sample prepared at molar ratio NaN_3:C60 100:1 show a small change in the peak position and so does a change in the linewidth respect to the other samples explored in the titration study. Those changes could be indicative of a weak binding interaction between NaN_3 and C60. At high molar ratio NaN_3:C60 100:1. The ^{14}N peaks of sodium azide have observable CSPs and changes in Linewidth. The two effects are stronger for the two external nitrogens of sodium azide (signal B) than for the central nitrogen (signal A). The mentioned effects could indicate a weak binding interaction between NaN_3 and C60.

Fig. 1 ^{14}N NMR titration study. C60 fullerene added to NaN_3 water solution. Superimposition of two spectra of sodium azide: black line is corresponding to pure sodium azide water solution and green to sodium azide titrated with fullerene water solution at lowering concentrations. At low molar ratio NaN_3:C60 (<= 10 :1) no appreciable change of ^{14}N chemical shift or linewidth occurs for two peaks of NaN_3 (Signals A and B).

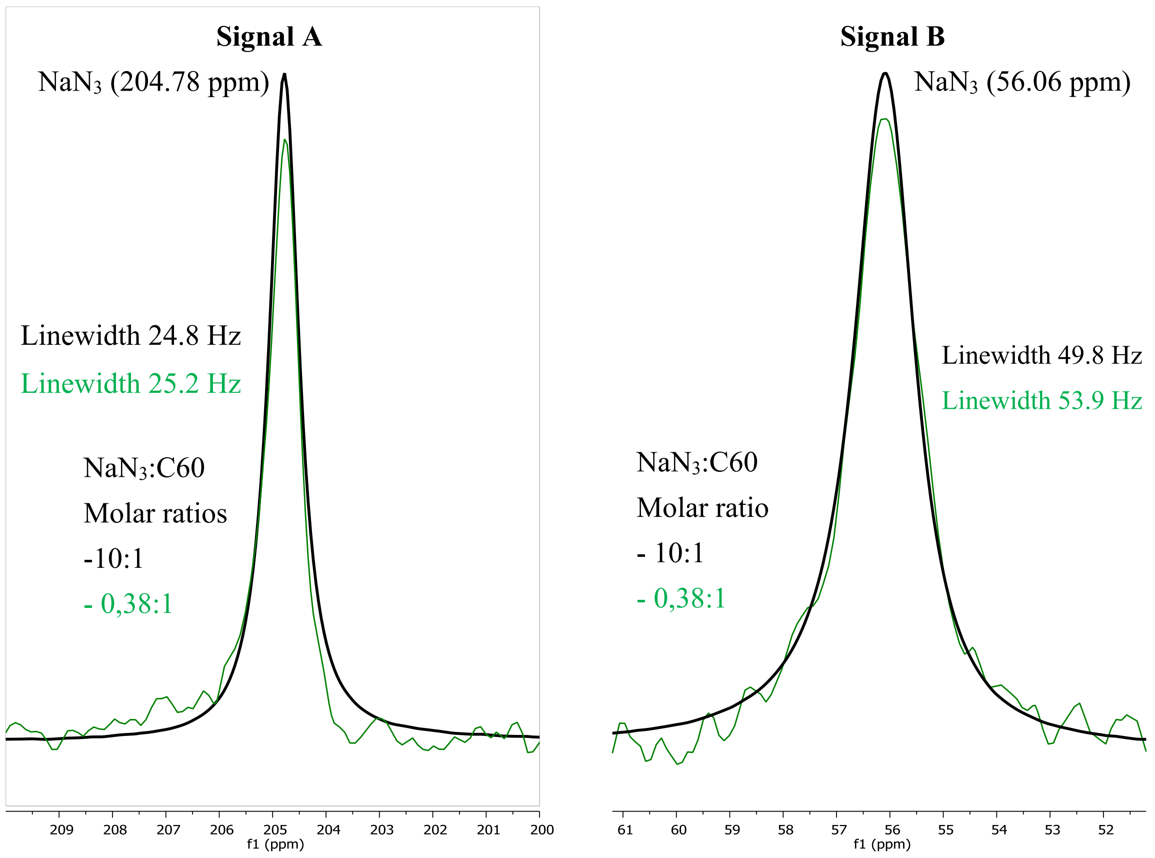
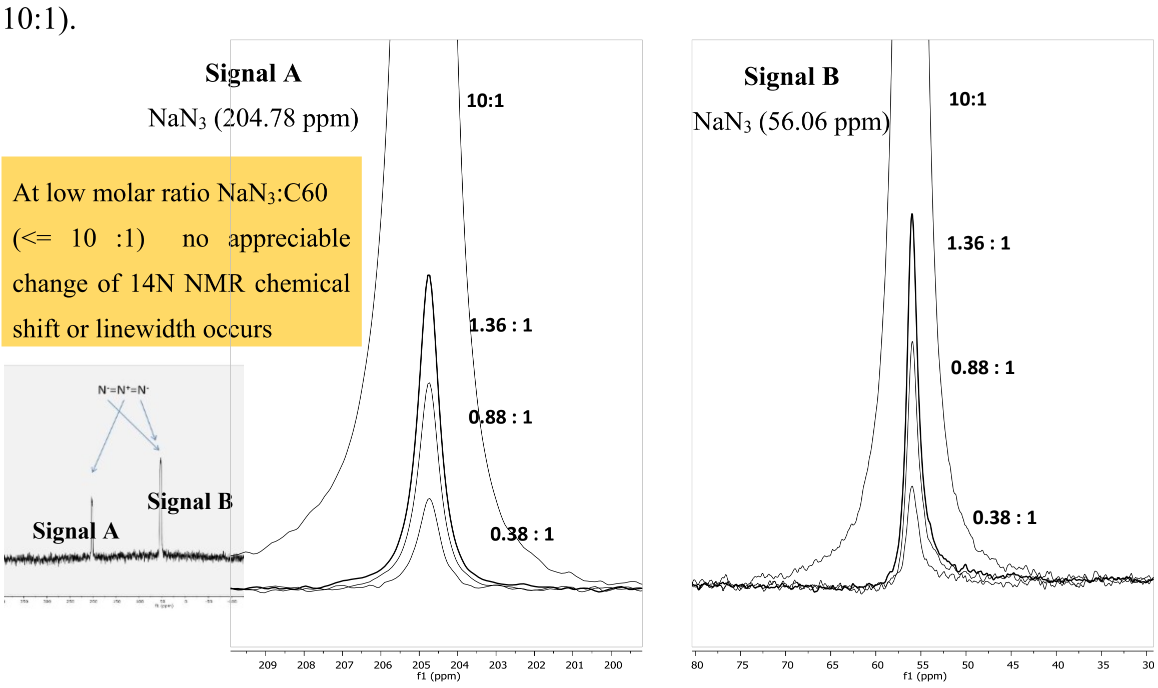


Figure 2. ^{14}N NMR superimposition of four different spectra at low molar ratios of NaN_3:C60 (<= 10:1).



At low molar ratio NaN_3:C60 (<= 10 :1) no appreciable change of 14N NMR chemical shift or linewidth occurs

Figure 3. Superimposition of two spectra at high ratio of NaN_3:C60 (>10 :1), particularly 100:1 There are some subtle changes of ^{14}N chemical shift and linewidth of both NaN_3 peaks.

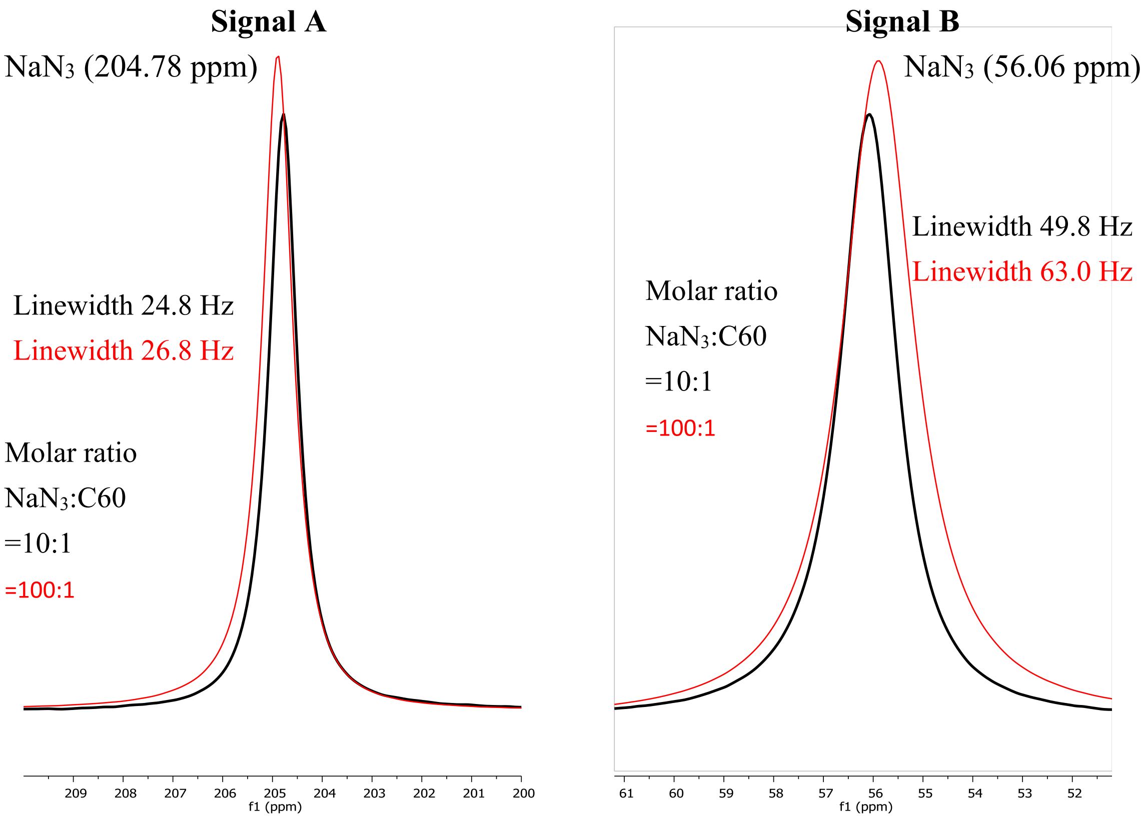
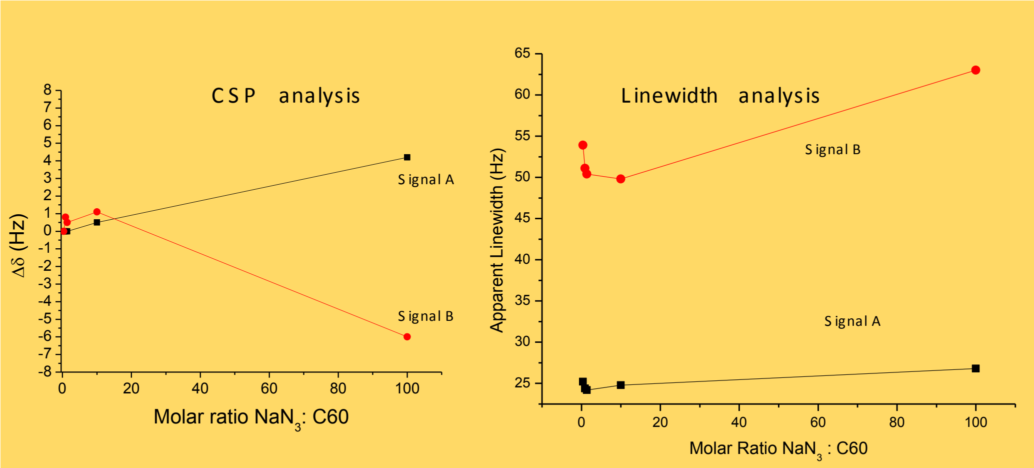


Figure 4. Chemical Shift Perturbations (CSP) and Linewidth study of ^{14}N peaks of NaN_3 at several molar ratios.



To confirm obtained results it is required extra experiment measuring the ^{14}N spectrum of a sample containing only sodium azide in water at the same concentration of 10 M without C60 to discard a false positive. There were anticipated two possible outcomes of this experiment: if the changes mentioned above does not occur in this sample, then they must be due to C60, and therefore the weak-binding is confirmed, but if the changes mentioned above do occur also in this sample, then there is likely an effect of the high concentration of NaN_3 in the sample, but we cannot say that we have detected a weak-binding between C60 and NaN_3.

The results obtained with the control sample of pure NaN_3 at 10 M demonstrated that there are changes in the chemical shift position and line-broadening related to the molar ratio 100:1 of NaN_3:C60 in the sample. These results can be interpreted as binding interaction occurring between NaN_3 and C60 molecules, from the two ^{14}N peaks of NaN3, the one that is more affected is the one that resonates at approximately 56 ppm.

Discussion:

Some solutes, such as C60 can have a dynamic interaction of binding with other solutes, such as Na N_3. They are exchanging during time between a bound state when the two molecules are packed together and unbound state in which the two molecules are independent. Chemically this is represented by equilibrium:

N_3 + C60 <====> N_3:C60
Free state Bound state

The double arrow indicates that the equilibrium moves in both direction during time, in our case the timescale of this reaction is likely below <100 ms for moving in both directions.

In the instant that the two molecules are bound their atoms are packed together and as consequence their electron clouds introduce a slight shielded/de-shielded of the external magnetic field that is felt by the nuclei of any of the two molecules. Thus, the ^{14}N NMR peaks of N_3 molecule feels on average a slightly different magnetic field when C60 is present and therefore resonate at a slightly different chemical shift than when C60 is not present. We made a titration with C60 and followed the changes of chemical shifts of ^{14}N. The fact that changes are observed by addition of C60 demonstrates that the two molecules interact and bind together dynamically.

The influence of C60 in the chemical shifts of NaN_3 is not linear with the concentration, i.e. the two molecules interact one to one proportion as indicated above, it has a quadratic dependence with the concentration.

4. Conclusion:

The results demonstrate that there are changes in the chemical shift position and line-broadening related to the molar ratio NaN_3:C60 in the sample (100:1). One of two peaks of ^{14}N, which resonates at 56 ppm and corresponds to two external nitrogen atoms of NaN_3 is more affected by influence of hydrated fullerene C60. These results can be interpreted as binding interaction occurring between NaN_3 and C60 molecules.

Acknowledgments

The authors express gratefulness for financial support company Chemanteck (Galicia) and its founders Raquel Sendón García, Ana Rodríguez Bernaldo de Quirós.

Special acknowledgments to Prof. Dr.: Perfecto Paseiro Losada for scientific advising and Miguel Angel Lago Crespo for technical assistance, University of Santiago de Compostela.

Conflicts of Interest

The authors declare no conflict of interest.

References

1 Petrovic, M., Gonzalez, S., Barcelo, D. Analysis and removal of emerging contaminants in wastewater and drinking water. *Trends Anal Chem* 2003, 22, 685–696.

2 Focazio, M.J., Kolpin, D.W., Barnes, K.K., Furlong, E.T., Meyer, M.T., Zaugg, S.D., Barber, L.B., Thurman, E.M. A national reconnaissance for pharmaceuticals and other organic wastewater contaminants in the United States: II) Untreated drinking water sources. *Science of the Total Environment*, 2008, 402, 201-216.

3 Buser, H.R., Poiger, T., Muller, M.D. Occurrence and environmental behavior of the chiral pharmaceutical drug ibuprofen in surface and in wastewater. *Environ Sci Technol* 1999, 33, 2529–2535.

4 Heberer, T. Occurrence, fate, and removal of pharmaceutical residues in the aquatic environment: A review of recent research data. *Toxicol Lett* 2002, 131, 5–17.

5 Ternes, T.A. Occurrence of drugs in German sewage treatment plants and rivers. Water Res 1998, 32, 3245–3260.

6 Metcalfe, C.D., Koenig, B.G., Bennie, D.T., Servos, M., Ternes, T.A., Hirsch, R. acidic drugs in the Occurrence of neutral and acidic drugs in effluents of Canadian sewage treatment plants. *Environ Toxicol Chem* 2003, 22, 2872–2880

7 Castiglioni, S., Bagnati, R., Fanelli, R., Pomati, F., Calamari, D., Zuccato, E. Removal of pharmaceuticals in sewage treatment plants in Italy. *Environ Sci Technol* 2006, 40, 357–363.

8 Joss, A., Keller, E., Alder, A.C., Göbel, A., McArdell, C.S., Ternes, T., Siegrist, H. Removal of pharmaceuticals and fragrances in biological wastewater. *Water Res* 2005, 39:3139–3152

9 Giger, W., Alder, A.C., Golet, E.M., Kohler, H.E, McArdell, C.S., Molnar, E., Siegrist, H., Suter, M.F. Occurrence and fate of antibiotics as trace contaminants in wastewaters, sewage sludges, and surface waters. *Chimia* 2003, 57, 485–491

10 Haase, J., Large - scale Preparation and Usage of Azides in book "Organic Azides: Syntheses and Applications", Ed. S.Brase and K. Banert. John Wiley & Sons, USA, 2010. ISBN: 978-0-470-51998-1, Ch 2, 30-51

11 Roh, J., Bartels, C., Wilf, M. Use of Dendrimers to Enhance Selective Separation of Nanofiltration and Reverse Osmosis Membranes. Desalination and Water Purification Research and Development Report # 140. 2009.
www.usbr.gov/pmts/water/publications/reports.html

12 Jones, O., Voulvoulis, N., Lester, J. Human Pharmaceuticals in Wastewater Treatment Processes. Critical Reviews in Environmental Science and Technology, 2005, 35, 401–427.

13 Buseck, P., S. Tsipursky, R. Hettich. Fullerenes from the Geological Environment. *Science* 1992, 257(5067), 215-7. (DOI:10.1126/science.257.5067.215).

14 Krasnovyd, S.V., Konchits, A.A., Shanina, B.D., Valakh, M.Y., Yanchuk, I.B., Yukhymchuk, V.O., Skoryk, M.A. Local structure and paramagnetic properties of the nanostructured carbonaceous material shungite. *Nanoscale Research Letters* 2015, 10, 78. doi.org/10.1186/s11671-015-0767-9

15 Buseck, P.R. Geological fullerenes: review and analysis. *Earth and Planetary Science Letters.* 2002, 203, 781–792. doi: 10.1016/S0012-821X(02)00819-1.

16 Augustyniak-Jabłokow, M.A., Yablokov, Y.V., Andrzejewski, B., Kempinrski, W., Łos, S., Tadyszak, K. et al. EPR and magnetism of the nanostructured natural carbonaceous material shungite. *Phys Chem Minerals*. 2010, 37, 237–47. doi: 10.1007/s00269-009-0328-9.

17 McKeen, S. (Ed). High Production Volume (HPV) Chemicals. The Organization for Economic Co-operation and Development. Environment Directorate. Paris. 2010.

18 a) Rodríguez-Kábana, R. and Robertson, D.G. Nematicidal and herbicidal properties of potassium azide. *Nematropica*, 2000, 30,146.

b) Rodríguez-Kábana, R. Pre-plant applications of sodium azide for control of nematodes and weeds in eggplant production. Proceedings Annual International Research Conference on Methyl Bromide Alternatives and Emissions Reductions. Nov. 5-9, 2001, San Diego, CA. 6-1.

c) Rodríguez-Kábana, R. Efficacy of aqueous formulations of sodium azide with amine-protein stabilizers for control of nematodes and weeds in tomato production. Proceedings Annual International Research Conference on Methyl Bromide Alternatives and Emissions Reductions. Nov. 5-9, 2001, San Diego, CA. 7-1.

e) Rodríguez-Kábana, R. and Abdelhaq, H. Sodium azide for control of root-knot nematodes and weeds in green pepper and tomato production in the Souss valley. Proceedings Annual International Research Conference on Methyl Bromide Alternatives and Emissions Reductions, Nov. 5-9, 2001, San Diego, CA. p 8 -1.

19 a) Betterton, E. A. Environmental Fate of Sodium Azide Derived from Automobile Airbags. *Critical Reviews in Environmental Science and Technology* 2003, 33, 423-458.

b) Betterton, E. A., Lowry, J., Ingamells, R., Venner, B. Kinetics and mechanism of the reaction of sodium azide with hypochlorite in aqueous solution. *J. Hazardous Materials* 2010, 182, 716–722.

c) Betterton, E.A. & Craig, D. Kinetics and Mechanism of the Reaction of Azide with Ozone in Aqueous Solution. *J. Air & Waste Management Association* 1999, 49(11), 1347-1354.

20 Park, H., Chang, I., Lee K. (Ed.). Membranes and Module Modification. Principles of Membrane Bioreactors for Wastewater Treatment. CRC press Taylor & Francis Group, New York. 2015, 281-282.

21 Jassby, D., Chae, S.R., Hendren, Z., Wiesner, M. R. Membrane filtration of fullerene nanoparticle suspensions: Effects of derivatization, pressure and electrolyte concentration. *J. of Colloid and Interface Science,* 2010, 346(2), 296-302.

22 Chae, S. R., Therezien, M., Budarz, J. F., Wessel, L., Lin, S., Xiao, Y. and Wiesner M. R. Comparison of the photosensitivity and bacterial toxicity of spherical and tubular fullerenes of variables aggregate size. *J. of Nanoparticle Research,* 2011, 13, 5121-5127.

23 Chae, S.R., Hotze, E.M., Wiesner, M. R. Possible applications of fullerene nanomaterials in water treatment and reuse, Nanotechnology Applications for Clean Water. William Andrew Publishing, New York, USA 2009 (ISBN: 9780815515784).

24 Chae, S. R., Wang, S., Hendren, Z., Wiesner, M. R., Watanabe, Y., Gunsch, C. K. Effects of fullerene nanoparticles on Escherichia coli K12 respiratory activity in aqueous suspension and potential use for membrane biofouling control. J. of Membrane Science, 2009, 329, 68-74. doi:10.1016/j.memsci.2008.12.023

25 a) Andrievsky, G.V., Kosevich, M.V., Vovk, O.M., Shelkovsky, V.S., Vashchenko, L.A. On the production of an aqueous colloidal solution of fullerenes. J Chem Soc Chem Commun 1995,1281-1282.

b) Andrievsky, G.V., Klochkov, V.K., Karyakina, E.L., Mchedlov-Petrossyan, N.O. Studies of aqueous colloidal solutions of fullerene C60 by electron microscopy. Chem Phys Lett 1999, 300: 392-396.

c) Andrievsky, G.V., Klochkov, V.K., Bordyuh, A.B., Dovbeshko, G.I. Comparative analysis of two aqueous-colloidal solutions of C60 fullerene with help of FTIR reflectance and UV-VIS spectroscopy. *Chem Phys Lett* 2002, 364, 8-17.

d) Avdeev, M.V., Khokhryakov, A.A., Tropin, T.V., Andrievsky, G.V., Klochkov. V.K. et al. Structural features of molecular-colloidal solutions of C60 fullerenes in water by small-angle neutron scattering. *Langmuir* 2004, 20, 4363-4368.

26 Chachibaia, T. & Pastor, M. The State-Of-The-Art chemical analytical method for detection of sodium azide by 14N NMR spectroscopy. *J. Nano Studies* 2015, 11: 8-15.

SciForum
Mol2Net

The Symmetry-Adapted Configurational Ensemble Approach to the Computer Simulation of Site-Disordered Solids

Ricardo Grau-Crespo [1],* and Said Hamad [2]

[1] Department of Chemistry, University of Reading, Whiteknights, Reading RG6 6AD, UK

[2] Departamento de Sistemas Físicos, Químicos y Naturales, Universidad Pablo de Olavide, Carretera de Utrera km. 1, 41013 Seville, Spain

* Author to whom correspondence should be addressed; E-Mail: r.grau-crespo@reading.ac.uk; Tel.: +44 118-378-7180.

Received: 8 October 2015 / Accepted: 10 November 2015 / Published: 4 December 2015

Abstract: Site-occupancy disorder, defined as the non-periodic occupation of lattice sites in a crystal structure, is a ubiquitous phenomenon in solid-state physics and chemistry. Examples are mineral solid solutions, synthetic non-stoichiometric compounds and metal alloys. The experimental investigation of these materials using diffraction techniques only provides averaged information of their structure. However, many properties of interest in these solids are determined by the local geometry and degree of disorder, which escape an "average crystal" description, either from experiments or from theory. In this paper, I describe a methodology for the computer simulation of site-disordered solids, based on the consideration of configurational ensembles and statistical mechanics, where the number of occupancy configurations is reduced by taking advantage of the crystal symmetry of the lattice. Thermodynamics and non-thermodynamic properties are then defined from the statistics in the symmetry-adapted configurational ensemble. I will briefly summarize and discuss some recent applications of this methodology to problems in materials science.

Keywords: site disorder; density functional theory; solid state chemistry; statistical mechanics

Mol2Net YouTube channel: *http://bit.do/mol2net-tube*

1. Introduction

There are many problems in materials science involving the solution of two or more crystalline materials with some level of randomness in the occupancy of lattice sites. We define site disorder as the kind of disorder that results from *non-periodic* occupation of lattice sites in a crystal structure. Amorphous disorder differs from lattice site disorder in that the latter does not destroy the long-range periodicity of the lattice sites, except possibly with small local atomic displacements with respect to lattice sites. Examples of site-disordered materials include metallic alloys, mineral solid solutions, and synthetic non-stoichiometric compounds. The computational modeling of these solids is challenging because periodic boundary conditions cannot be applied in the same straightforward way as in the simulation of ordered crystals.

2. Classification of site-disorder models

The models employed in the literature to investigate site-disordered solids can be classified in three broad groups (**figure 1**).

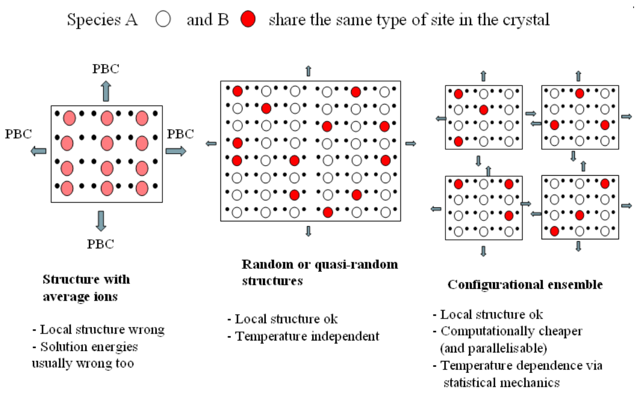

Species A ○ and B ● share the same type of site in the crystal
PBC
PBC
PBC
PBC
Structure with average ions
- Local structure wrong
- Solution energies usually wrong too
Random or quasi-random structures
- Local structure ok
- Temperature independent
Configurational ensemble
- Local structure ok
- Computationally cheaper (and parallelisable)
- Temperature dependence via statistical mechanics

Figure 1. Classification of methods to represent site-disorder in solids

The first group comprises all methods in which a sort of average atom is defined, thus allowing recovering the perfect periodicity of the crystal. In the context of classical calculations, based on analytical interatomic potentials, this is done by making each site experience a potential which is the mean or weighted average of all possible configurations corresponding to disordered atomic positions. This approach is implemented, for example, in the GULP code [1,2], and can sometimes be useful for preliminary simulations of very disordered (random) systems. An equivalent method in the world of quantum-mechanical simulations is the virtual crystal approximation (VCA), where the potential felt by electrons is the one generated by average atoms, *i.e.* average of potentials of atoms that can occupy a given site and its periodic images. The main drawback of this kind of methods is that the local structure around each particular ion in a real material is very poorly represented by the geometry around these average ions.

In the second group of methods for treating site-disordered solids, a large periodic supercell is employed with a more or less random distribution of ions at the sites. This type of representation is computationally more expensive than the "average-ion" models, but it provides a better description of the local geometries found in the real system. It is assumed that (a) distribution of atoms on lattice sites is random, and (b) the supercell is large enough to include a large number of possible local arrangements of ions, amounting to spatial averaging over configurations. A useful variation of this model is the special quasi-random structure, where the ion positions in the supercell are chosen to mimic as closely as possible the most relevant near-neighbor pair and multisite correlations of a random substitutional alloy [3]. Special quasi-random structures are particularly useful in evaluation of the electronic structure and related properties such as magnetic moments of site-disordered solids. Their main limitation comes from the inflexibility of the ion distribution in the structure, which is fixed to mimic a disordered (a truly random) solid. However, we often desire to investigate varying degrees of disorder, for example, short-range ordering can be present depending on the temperature used in the synthesis, for which we need a more flexible representation.

The third type of methods is the multi-configurational supercell approach, which is the focus of the present paper. Within this approach, an infinite site-disordered solid is modeled with a set of configurations with of various site occupancies in a supercell representing a piece of the solid. Each configuration corresponds to a particular arrangement of the atoms within the supercell, and has associated a probability of occurrence. The idea behind the method is that listing all possible configurations and their probabilities can provide a reasonable description of the distribution of the ions and their level of disorder, at least within the range marked by the supercell size.

In order to attribute experimental meaning to this kind of representation, we can image the case of a site-disordered surface that is being studied with an electron microscope, capable of taking atomic resolution photographs of the surface. If we take a very large number of photographs at randomly selected sections of the surface, we have all the information required to describe any distribution pattern with ordering range shorter than the size of the photographs. The probability of a given configuration can then be defined as its frequency of appearance in the limit of a very large number of images. In the limit of infinite size of the system, this scheme becomes exact.

From a theoretical point of view, the central challenge in the multi-configurational supercell approach is the calculation of the probabilities of occurrence of the configurations, under the assumption of configurational equilibrium. A typical approximation consists of assigning an energy to each configuration, and then applying a formalism based on Boltzmann-Gibbs statistical mechanics, in such a way that those configurations with lower energies have higher probabilities, and the dispersion of the distribution is controlled by the temperature: at low temperatures only the most stable configurations occur, whereas at high temperatures more configurations are accessible to the ions, leading to higher degree of disorder.

We should note that the existence of a well-defined energy characterizing the stability of each configuration is not trivial, as strictly speaking the energy contribution from a given ionic configuration in a cell depends on the configuration of the ions in the neighboring cells. Only for large supercells, where *intracell* interactions are much more important than *intercell* interactions, the energy of a given configuration can be considered a function of the arrangement of the ions within the supercell. In what follows we will consider that this is the case, i.e., the energy of a given configuration is independent of the distribution of cations next to the cell boundary. Then, we could as well assume that the distribution of ions in the region adjacent to supercell is identical to that in the supercell used in simulations. Thus, we can simply calculate the configuration energy using periodic boundary conditions, which is straightforward using modern computer programs for solid state simulations.

Orthogonally to the classification of the methods in terms of the representation of disorder, we can also classify the methods in terms of the types of methods used for evaluating the energy of the periodic configurations in the crystal. This is actually a typical task in computational physics and chemistry, and the different methods and approximations involved have been widely discussed elsewhere (e.g. [4]). We only present here a simple classification in increasing order of sophistication and computational cost:

Type I methods: in this case the energy is a function of only the site occupancies. Electronic and structural (geometric) degrees of freedom are not included explicitly, but are implicit within a model that yields the energy from nearest-neighbor (NN), next-nearest-neighbor (NNN) configurations or longer-distance effective pair interactions, or including terms for clusters of more than two ions. These types of methods include Ising-type models of alloys and cluster expansion methods, and have been used extensively in determination of phase diagrams of alloys.

Type II methods: including geometric relaxations explicitly, but electronic effects only implicitly. The energy is evaluated in this case via a classical interatomic potential function or force field, which is a function of the ionic coordinates. These are quite efficient computationally and can be used in studies of symmetry-breaking structural phase transitions such as those in a ferroelectric or a shape memory alloy. In such a case, the structural degrees of freedom are directly relevant to the problem and cannot be integrated out.

Type III methods: they include explicit geometric and electronic relaxation for each configuration. This category comprises all quantum-mechanical methods, including those based on the density functional theory (DFT) and its extensions, Hartree-Fock (HF) and post-HF approaches, hybrid DFT/HF, semi-empirical methods like tight-binding, etc. Computationally, these methods are quite expensive and can be used only with relatively small supercells. They have to be used in problems that involve electronic phase transitions (such as magnetic or metal-insulator transition), and those where site-specific chemistry is relevant, for example in catalysts.

In the evaluation of energies of a large number of configurations to investigate the thermodynamics of disorder, type I methods are still the most commonly employed. Not only they are computationally cheaper, but they are also easier to integrate with sampling algorithms (e.g. Metropolis – Monte Carlo) within a single computer program. In contrast, energy evaluations using type II and type III methods are more expensive and typically require a specialized program that deals with only one geometric configuration at a time. Configurational sampling in these cases requires multiple calls to the quantum-mechanical or interatomic potential code from an external program, and has very long running times. However, there are good reasons to move towards more sophisticated (type II and III) methods to evaluate energies in a multi-configurational simulation. First, such calculations not only provide energies, but many other properties too (e.g. local geometries and cell parameters, and in the case of type III methods, electronic structure information). Any property that can be obtained for each configuration can be averaged over the ensemble to obtain effective values for the disordered solid. Second, methods of type II and III also provide access to vibrational properties of the solid and its response to external pressure, therefore allowing an integration of configurational and vibrational degrees of freedom in the construction of complex phase diagrams (as a function of composition, temperature and pressure). Finally, if interactions in the system are long-range, energy evaluations in terms of simple type I methods might not provide good precision, or would require a large number of terms and parameters [5]. The main disadvantage of methods of type II and III, while evaluating a large number of configurations, is their computational cost, but with the developments in computer hardware and efficient algorithms, they are becoming much more affordable.

3. Statistical formulation of the method

3.1 The Boltzmann equations

We now present the statistical formulation of the configurational equilibrium in a site-disordered binary system. For the sake of simplicity, in this initial formulation all configurations are constrained to have the same compositions, and we will ignore vibrational and pressure effects in the thermodynamics; the corresponding generalizations are introduced later in this chapter.

The extent of occurrence of the each configuration (labeled with an index *k*) is described in this approximation by a Boltzmann-like probability which is calculated from the energy E_k of the configuration and the temperature *T*:

$$P_k = \frac{1}{Z}\exp(-E_k / k_B T) \tag{1}$$

where k_B is Boltzmann's constant (it is formally equivalent to use the gas constant *R* instead, and expressing the molar energies of supercells, but we follow here the usual notation in statistical mechanics in terms of k_B),

$$Z = \sum_{k=1}^{K}\exp(-E_k / k_B T) \tag{2}$$

is the partition function, and *K* is the total number of configurations with the given composition in the supercell. For a binary system, *K* can be calculated as a number of combinations:

$$K = \frac{N!}{(N-n)!n!} \tag{3}$$

where *N* is the number of exchangeable sites in the supercell, and *n* is number of ions of one of the species (it can also be a vacancy) that can occupy these sites. The molar concentrations are then $x=n/N$ for the first species, and 1-*x* for the second species.

The definition of the configurational ensemble and the associated probabilities allows us to obtain the effective value in the disordered solid of any quantity that can be theoretically obtained for each ordered configuration. If A_k is the value of the given magnitude for configuration *k*, then the effective value in the disordered solid is:

$$A = \sum_{k=1}^{K} P_k A_k \tag{4}$$

In this way, it is possible to obtain effective values even for quantities like the cell parameters, which are not strictly defined but still have experimental meaning in a solid with non-periodic distribution of ions on lattice sites. Equation (4) also allows us to obtain the effective energy of the solid from the configurational energies E_k, as:

$$E = \sum_{k=1}^{K} P_k E_k \qquad (5)$$

In evaluating the thermodynamic stability of a disordered solid at a given temperature, not only the energy but also the configurational multiplicity of the system should be taken into account, which is done by introducing the (configurational) free energy:

$$F = -k_{\mathrm{B}} T \ln Z = -k_{\mathrm{B}} T \sum_{k=1}^{K} \exp(-E_k / k_{\mathrm{B}} T) \qquad (6)$$

The difference per temperature unit between the average energy and the free energy defines the configurational entropy, which can also be expressed in terms of the probabilities P_k:

$$S = \frac{E - F}{T} = -k_{\mathrm{B}} \sum_{k=1}^{K} P_k \ln P_k \qquad (7)$$

We can distinguish two important limiting cases here:

Perfect order: this occurs when one configuration (say k=1) is much more stable than the rest, *i.e.*, it is separated from the other configurations by an energy difference much larger than $k_{\mathrm{B}}T$. In this case P_1=1, while P_k=0 for k≠1, and the system will have zero configurational entropy. The configurational free energy simply corresponds to the energy of the most stable configuration.

Perfect disorder: this occurs when the energies of all the configurations are very similar (again in comparison with $k_{\mathrm{B}}T$) or formally in the limit $T \to \infty$. In this case, all configurations have the same probability $P_{\mathrm{k}}=1/K$ and the configurational entropy reaches its maximum possible value:

$$S_{\max} = k_B \ln K = k_B \ln \frac{N!}{[N(1-x)]![Nx]!} \qquad (8)$$

which, in the limit of an infinitely large supercell ($N \to \infty$ at constant x), using Stirling's formula, converts to the well-known expression:

$$S_{\mathrm{ideal}} = -k_B N (x \ln x + (1-x) \ln(1-x)) \qquad (9)$$

Intermediate to these two limiting cases, there is a continuum of situations with varying degrees of ordering, which can be described within the same formalism, leading to temperature-dependent entropy values given by Eq. (7). It is clear from the equations above that any finite supercell is unable to describe exactly the perfect disorder limit. In order to correct for this, it is convenient to re-write the free energy (Eq. 6) as:

$$F = -k_{\mathrm{B}} T \ln K - k_{\mathrm{B}} T \ln\left(\frac{1}{K} \sum_{k=1}^{K} \exp\left(-E_k / k_{\mathrm{B}} T\right) \right) \qquad (10)$$

as suggested by Becker et al.[6]. The first term represents the entropy contribution in the limit of perfect disorder, while the second term contains the energy contribution plus the correction to the entropy contribution due to partial ordering. The first term can then be adjusted to its correct value, *i.e.*:

$$\begin{aligned} F &= N k_{\mathrm{B}} T \left(x \ln x + (1-x) \ln(1-x) \right) \\ &- k_{\mathrm{B}} T \ln\left(\frac{1}{K} \sum_{k=1}^{K} \exp\left(-E_k / k_{\mathrm{B}} T\right) \right) \end{aligned} \qquad (11)$$

which is equivalent to amending the temperature-dependent entropy by a term

$\Delta S_{\text{corr}}=S_{\text{ideal}}-S_{\max}$, thus guaranteeing the correct behavior in the limit of perfect disorder. However, this adjustment also breaks down in the description of the perfect order limit, by introducing a spurious configurational entropy contribution to the free energy (which should be zero in this limit). Therefore this correction should be applied only to simulations of systems with nearly-perfect disorder.

3.2 Including vibrational contributions

This basic methodology can be made more sophisticated to include other effects, depending on the problem in hand. For example, in order to consider vibrational contributions to the thermodynamics of a disordered solid within the multi-configurational formalism, we can write the total partition function of the system as:

$$Z=\sum_{k=1}^{K}\sum_{v}\exp(-E_{k,v}/k_{\text{B}}T) \qquad (12)$$

where v is an index (or strictly speaking, a collection of indices) that characterizes each vibrational state (with energy $E_{k,v}$) of configuration k. In terms of the vibrational partition function $Z_k^{(vib)}$ and the vibrational free energy $F_k^{(vib)}$ of each configuration, this becomes:

$$Z=\sum_{k=1}^{K}Z_k^{(vib)}=\sum_{k=1}^{K}\exp(-F_k^{(vib)}/k_{\text{B}}T) \qquad (13)$$

which, by comparison with Eq. (6), indicates that the formalism including vibrational contributions is equivalent to the one introduced in the previous section, except that now the vibrational free energy $F_k^{(vib)}$ of each configuration should be used to define the probabilities (1) instead of energy E_k of the configuration. Using methods of type II and III, where the energy depends explicitly on ionic coordinates, vibrational free energies can be evaluated from the vibrational frequencies of each configuration, by invoking the harmonic approximation[7]. This, of course, adds considerably to the cost of the simulations, especially if the equilibrium geometry of each configuration is obtained by minimizing the free energy and not just the energy, but can become affordable if methods of Type II are being used (e.g. ref. [8]).

Analogously, if we want to introduce the effect of a finite external pressure p, we should use the Gibbs free energy:

$$G_k^{(vib)}=F_k^{(vib)}+pV_k=H_k-TS_k^{(vib)} \qquad (14)$$

to calculate the configurational probabilities, where V_k, H_k and $S_k^{(vib)}$ are the supercell volume, the enthalpy and the vibrational entropy, respectively, for the particular configuration. In this case, the effective thermodynamic potentials for the disordered solid are:

$$H=\sum_{k=1}^{K}P_kH_k \qquad (15)$$

$$G=-k_{\text{B}}T\sum_{k=1}^{K}\exp(-G_k^{(vib)}/k_{\text{B}}T) \qquad (16)$$

and

$$S=\frac{H-G}{T}=\sum_{k=1}^{K}P_kS_k^{(vib)}-k_{\text{B}}\sum_{k=1}^{K}P_k\ln P_k \qquad (17)$$

where in the last expression for the entropy, the first term is the vibrational contribution and the second term is the configurational contribution. The correction defined by expression (11) can be analogously applied here to treat highly disordered systems.

3.3 Accessing the configurational space

Computing energies and other properties of all possible configurations of ions in a mixed solid can be a rather demanding task, even for relatively small cells. Let us consider the case of a body-centered cubic (bcc) binary alloy, with a 2×2×2 supercell, which has only 16 exchangeable sites. The number of configurations as a function of the substitution fraction *x* increases very quickly and reaches a maximum at *x*=0.5, when there is a total of 12870 configurations (**figure 2**). This number is tractable with methods of type I, or even type II, but already becomes too expensive for type III methods. In a 3×3×3 supercell, with 54 exchange sites, the maximum is ~2×10^{15} configurations, which becomes very expensive even for type I methods.

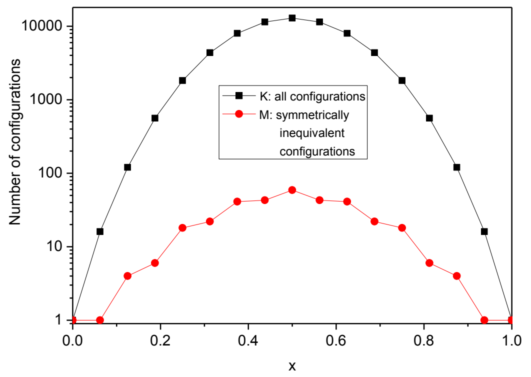


Figure 2. Number of configurations (*K*) in a model of a binary alloy with bcc structure using a 2x2x2 supercell, in comparison with the number of symmetrically inequivalent configurations.

It is therefore clearly necessary to find strategies to reduce the number of configurations to evaluate. We will discuss here three possible routes: *i)* taking advantage of the crystal symmetry, *ii)* random sampling, and *iii)* importance sampling using Metropolis – Monte Carlo algorithms.

3.4 Taking advantage of the crystal symmetry

If we are dealing with relatively small supercells, for example, when doing quantum-mechanical calculations for each configuration, it is possible to reduce the number of configurations by taking advantage of the crystal symmetry of the lattice [9]. Within this approach, two configurations are considered equivalent when they are related by a symmetry (an isometric) operation, for example, a reflection. A list of all possible isometric transformations is provided by the group of symmetry operators in the parent structure (the original structure without any substitutions). They include the symmetry operators in the space group of the crystal unit cell (scaled in an appropriate way to account for the cell multiplicity of the supercell), the supercell internal translational operators, and the combinations between them. It is then possible (at least for small systems) to start with all possible configurations through explicit enumeration and reduce to those which are symmetrically inequivalent for energy/properties evaluation. We also need to keep track of the degeneracy of each independent configuration, that is, how many times it repeats in the whole configurational space (this is similar to a number of members in the star of k-points in the Brillouin zone in the representation and group theory of crystals). This algorithm is implemented in the SOD (Site Occupancy Disorder) program [9,10].

It is necessary to slightly adapt the equations for configurational statistics to operate in the reduced space of inequivalent configurations. If E_m is the energy and Ω_m is the degeneracy of the independent configuration *m* (*m*=1,…, *M*), its contribution to energy or other properties needs to be weighted by a probability:

$$P_m = \frac{\Omega_m}{Z}\exp(-E_m / k_\mathrm{B}T) \qquad (18)$$

which means that if we want to compare the stability of two independent configurations in energetic terms, we should not use their energies E_m but instead the value $E_m - k_B T \ln \Omega_m$. Average values can be obtained using the equation analogous to (4):

$$A = \sum_{m=1}^{M} P_m A_m \qquad (19)$$

For scalar properties, *i.e.* if A_m is the same for all the Ω_m equivalent configurations that the inequivalent configuration *m* represents. For example, if we are modeling a cubic system, we cannot obtain the average cell parameter *a* from the cell parameters a_m of the inequivalent configurations, as this result could be different from the direct average of the b_m or c_m values, breaking the cubic symmetry. We therefore need to find first a related magnitude that is invariant in the subspace of equivalent configurations, *e.g.* the volume V_m in the given example. We can then define the average cell parameter of the cubic system as:

$$a = \left(\sum_{m=1}^{M} P_m V_m\right)^{1/3} \qquad (20)$$

Otherwise, one needs to explicitly symmetrize the property obtained using Eqn. (19) with all the symmetry operations used in obtaining the reduced set of configuration (for example, electric dipole or polarization in a ferroelectric).

Symmetry reduction is only practical when working with relatively small supercells. This is typically the case when properties other than the energy are being evaluated, using type II and type III methods. When evaluating the thermodynamic functions of the solution, the enthalpy tends to converge very quickly with supercell size, but the convergence of the entropy, which depends on configuration counting, is much slower. Therefore, for a complete thermodynamic characterization of solid solutions it is generally necessary to consider supercells much larger than those tractable by symmetry-reduction methods.

4. Examples of applications

4.1 Using symmetry-adapted ensembles to identify favourable ion distributions

We start with the simplest possible use of the multi-configurational representation of the ionic distribution in a mixed solid: finding the most stable configurations. We use the iron oxide γ-Fe_2O_3 (maghemite) as a first example.

Maghemite is the second most stable polymorph of iron (III) oxide. Its magnetism, chemical stability and low cost led to its wide application as magnetic pigment in electronic recording media since the late 1940's [11]. Maghemite nanoparticles are also widely used in biomedicine, because their high magnetic moment allows manipulation with external fields, while they are biocompatible and potentially non-toxic to humans [12,13]. Despite the compositional simplicity, its precise structure has been the subject of debate for decades. Like magnetite (Fe_3O_4), maghemite exhibits a spinel crystal structure, but while the former contains both Fe^{2+} and Fe^{3+} cations, in maghemite all the iron cations are in trivalent state, and the charge neutrality of the cell is guaranteed by the presence of cation vacancies. The debate about the maghemite structure has focused on the degree of ordering of these vacancies in the solid.

The unit cell of magnetite can be represented as $(Fe^{3+})_8[Fe^{2.5+}]_{16}O_{32}$, where the brackets () and [] designate tetrahedral and octahedral sites, respectively. The maghemite structure can be obtained by creating 8/3 vacancies out of the 24 Fe sites in the cubic unit cell of magnetite. These vacancies are known to be located in the octahedral sites [14] and therefore the structure of maghemite can be approximated as a cubic unit cell with composition $(Fe^{3+})_8[Fe^{3+}_{5/6}\diamond_{1/6}]_{16}O_{32}$. If the cation vacancies were randomly distributed over the octahedral sites, as it was initially assumed, the space group would be Fd3m like in magnetite. However, there is a evidence about a higher degree of ordering. Braun [15], for example, noticed that maghemite exhibits the same superstructure as lithium ferrite ($LiFe_5O_8$), which is also a spinel with unit cell composition $(Fe^{3+})_8[Fe^{3+}_{3/4}Li^{1+}_{1/4}]_{16}O_{32}$, and suggested this was due to similar ordering in both compounds. In the space group $P4_332$ of lithium ferrite, there are two types of octahedral sites, one with multiplicity 12 in the unit cell, and one with multiplicity 4, which is the one occupied by Li. In maghemite, the same symmetry exists if the Fe vacancies are constrained to these Wyckoff 4b sites, instead of being distributed over *all* the 16 octahedral sites. It should be noted, however, that some level of disorder persists in this structure, as the 4b sites have fractional (1/3) iron occupancies. Finally, there is also evidence of a fully ordered structure, exhibiting a tetragonal cell with space group $P4_12_12$ and $c/a\approx3$ (spinel cubic cell tripled along the *c* axis) [16,17].

A computational investigation of the energetics of vacancy ordering in maghemite was presented in [18]. A 1x1x3 supercell of the cubic structure was used to obtain the spectrum of energies of all the ordered configurations which contribute to the partially disordered $P4_332$ cubic structure. The energies were evaluated using long-range Coulomb contributions and classical interatomic potentials to describe short-range interactions (parameters derived by Lewis and Catlow[19]). The core-shell model of Dick and Overhauser[20] was employed to account for the polarizability of the anions. The calculations were performed with the GULP code [1,2,21]. Although not as sophisticated as quantum mechanical calculations, this methodology allows for accurate calculations of ion relaxations and configuration energies.

The total number of combinations of the 4 Fe ions on the so-called L sites of the supercell (**figure 3a**) is 12!/(4! ×8!)=495, but only 29 of these are inequivalent, as determined using the SOD program [9]. The calculated energies for these 29 configurations is shown in Fig. 4b. Only one of these configurations has the space group $P4_12_12$, found by Shmakov*et al.* [16] for fully ordered maghemite. This configuration is indeed the most stable one, with a significant energetic separation from the second most stable configuration (32 kJ/mol). The energy range covered by the configurational spectrum is quite wide (~850 kJ/mol), indicating that full disorder is very unlikely. The distinctive feature of the most stable configuration ($P4_12_12$) is the maximum possible homogeneity of iron cations and vacancies over the L sites. This configuration is the only one in which vacancies never occupy three consecutive layers; there are always two layers containing vacancies separated by a layer without vacancies, which instead contains Fe^{3+} cations in the L sites (*e.g.* positions L1 - L4 - L7 - L10) and the $P4_12_12$ configuration is therefore the one that minimizes the electrostatic repulsion between these cations.

In order to interpret the energy differences in the configurational spectrum in terms of the degree of vacancy ordering in the solid, we can calculate the probability of occurrence of each independent configuration. **Figure 4** shows the probabilities of the most stable configuration ($P4_12_12$) and of the second most stable configuration (with space group $C222_1$) as a function of temperature. At 500 K, a typical synthesis temperature for maghemite [16], the cumulative probabilities of all the configurations excluding the most stable $P4_12_12$ is less than 0.1%. This contribution increases slowly with temperature, but at 800 K this cumulative probability, which measures the expected level of vacancy disorder, is still less than 2%. At temperatures above 700-800 K maghemite transforms irreversibly to hematite (α-Fe_2O_3), and considering higher temperatures is therefore irrelevant. It thus seems clear that perfect crystals of maghemite in configurational equilibrium should have a fully ordered distribution of cation vacancies. Further analysis of the cation distribution in this oxide can be found in Ref. [18].

4.2 Configurational averages in the bulk and the surfaces: $Ce_{1-x}ZrO_2$ solid solutions

We discuss now some applications of the concept of configurational average, using the $Ce_{1-x}ZrO_2$ solid solution as a case study. This material is used as a support for the noble metals in the catalyst employed for the reduction of harmful emissions from car exhausts. A computational study of this solid solution was presented in ref. [22].

A supercell with 36 atoms was used there to model the bulk system, in particular the Ce-rich part of the solid solution ($0<x<0.5$ in $Ce_{1-x}Zr_xO_2$), which exhibits cubic symmetry ([23,24]). In this case, all calculations were performed using quantum-mechanical calculations, based on the density functional theory (DFT), as implemented in the VASP code [25,26]. From the calculations, it was immediately clear that the lowest-energy configurations were those where all the Zrions are grouped together, indicating a tendency to ex-solution. The tendency to ex-solution within bulk phases can be quantified by calculating the enthalpy of mixing:

$$\Delta H_{\mathrm{mix}} = H[\mathrm{Ce}_{1\text{-}x}\mathrm{Zr}_x\mathrm{O}_2] - (1-x)H[\mathrm{CeO}_2] - xH[\text{c-ZrO}_2] \qquad (21)$$

where $H[\mathrm{CeO}_2]$ and $H[\text{c-ZrO}_2]$ are the DFT energies per formula unit of ceria and cubic zirconia, respectively, and $H[\mathrm{Ce}_{1\text{-}x}\mathrm{Zr}_x\mathrm{O}_2]$ is the effective energy of the solid solution, calculated as a configurational average. The resulting enthalpy of mixing is strongly positive, in agreement with recent calorimetric measurements[24] (**Figure 5**).

Assuming a regular solid solution model (*e.g.*[27], [28]), the enthalpy of mixing at low Zr content was fitted with a polynomial of the form:

$$\Delta H_{\mathrm{mix}} = Wx(1-x) \qquad (22)$$

as in previous experimental work ([24,29]), which gives W=38 kJ/mol. This result is intermediate between the value of 28 kJ/mol obtained by [29]) from fitting a regular solution model to experimental solubility data, and the value of 51 kJ/mol obtained by [24]) by fitting directly to calorimetric measurements. The positive values of the enthalpy of mixing suggest that cation ordering is not a stabilizing factor in ceria-zirconia solid solutions, at least for the compositions examined here, and confirm that the Zrions have an energetic preference to segregate or form a separate Zr-rich phase. The origin of this tendency to is the difference between the ionic radii of the cations (r[Ce^{4+}]=0.97 Å and r[Zr^{4+}]=0.84 Å, for 8-fold coordination, according to [30]). It should be noted that real samples, where homogeneity at the atomic level can be achieved using special synthesis methods (*e.g.*[23,31]), might not experience this trend unless subjected to temperatures high enough to overcome the cation diffusion barriers.

In order to describe the thermodynamic stability of the solid solution at any finite temperature, entropies and free energies of mixing should be also calculated. It was found that, even assuming ideal configurational entropy, the resulting free energy of mixing is positive except for very small values of x. Furthermore, since Zr-rich phases are known to be monoclinic ([32]) at the temperatures of interest here, the mixing free energy should be calculated with respect to the more stable monoclinic zirconia phase (m-ZrO_2), which makes the mixed phase even less stable with respect to phase separation. In order to estimate the solubility limit of Zr in CeO_2, the mixing free energy function:

$$\Delta G_{\text{mix}}(x,T) = Wx(1-x) \\ +\Delta H_t x + RT[x\ln x + (1-x)\ln(1-x)] \qquad (23)$$

was considered, where the enthalpy of the monoclinic-cubic zirconia phase transformation ΔH_t =8.8 kJ/mol[33] was introduced. The use of the ideal entropy is justified because at very low Zr content the disorder should be nearly perfect. This analytical function allows the interpolation to x values smaller than those directly obtainable with the simulation supercell, and its minimum with respect to x at a given temperature provides an estimation of the solubility limit. **Figure 6** shows that the maximum equilibrium solubility of Zr from monoclinic zirconia into the ceria structure is ~0.4 mol% at 973 K, and increases to 2 mol% at 1373 K. Thus, although ceria-zirconia solid solutions in the whole range of compositions can be synthesized under adequate conditions (*e.g.*[31]), these results taken together with previous experimental evidence clearly show that these solid solutions are metastable with respect to phase separation into Ce-rich and Zr-rich phases. This phase separation can actually occur in a close-coupled catalytic converter, where temperatures of up to 1373 K could lead to rearrangement of the cations in the solid solution.

Simulations of the distribution of cations near the (111) surface of the solid were performed in the same study, by using the periodic slab model shown in Fig. 8. The number of configurations in the slab was reduced by only including those keeping the inversion symmetry of the cell and then selecting the symmetrically inequivalent ones. The equilibrium zirconium content of a particular cation layer parallel to the (111) surface, depends both on the overall zirconium content of the slab and on the temperature, and can be calculated by taking the configurational average:

$$c_l = \frac{\sum_m f_{ml}\Omega_m \exp(-E_m / k_\mathrm{B}T)}{\sum_m \Omega_m \exp(-E_m / k_\mathrm{B}T)} \qquad (24)$$

where f_{ml} is the fraction of sites occupied by Zr in the layer *l* for configuration *m*. The results are shown in Fig. 8 for temperatures between 800 and 1600 K. The most obvious feature of the cation distribution is the low concentration of Zr at the top (111) layer. Even for the 50:50 solid solution, at the highest temperature considered (1600 K), the equilibrium Zr content of the surface is only ~10%. The dependence of the calculated concentrations on temperature is relatively weak, especially at the top layer, but it is clear that increasing temperatures lead to more homogeneity in the composition of the interior of the slab, by equalizing the Zr content in the second and third layers. Thus, according to these results, the redistribution of cations at high temperatures should occur with significant Ce-enrichment of the (111) surface of ceria-zirconia, regardless of the overall composition of the solid solution. These conclusions are discussed in detail, in comparison with the experimental evidence, in ref. [22]

4.3 Stability of titanium oxynitrides

TiN and TiO2 (the most stable nitride and oxide phases of titanium) have an impressive number of interesting properties and potential applications in key technological fields. However, the properties are very different from one to other, and a complete change in the electronic and geometric structures takes place when TiN is oxidized to TiO_2 and when TiO_2 is nitrided to TiN. A number of intermediate phases of general composition TiO_xN_y, called "oxynitrides", appear in these complex processes. Obviously, the properties of the oxynitrides will be similar to those of the respective pure nitride and oxide when their compositions are close to those of the pure systems, and they will change progressively from those of the nitride to those of the oxide and vice versa when the compositions move to intermediate values. In principle, one could control and modulate the properties of the system by controlling the composition of the oxynitrides. In that way, potentially interesting combined properties could be obtained. But there are a number of questions to solve: are those oxynitrides stable phases which we are able to synthesize? What are their structures? Are their properties a result of the combination of those of the pure solids? Are we really able to control these properties as a function of the composition TiO_xN_y?

In order to answer these questions we made use of the SOD code. We performed[34] DFT calculations using a Generalized Gradient Approximation (GGA) implemented in the VASP code.[25] A plane-wave cutoff energy of 500 eV was used. We chose a supercell model of (1x1x3) for TiN (24 atoms) and one of (1x1x1) for α-TiO (20 atoms). These models allow us to change the composition progressively while still having a computationally affordable size, since we optimize the geometries with high accuracy (cutoff of 500 eV, saturation of k-points and demanding convergence criterions) for all the possible different configurations for each composition (240 configurations). The calculations were carried out using a (7x7x3) mesh for TiN and oxynitrides with NaCl-type structure, and a (6x4x6) mesh for TiO and oxynitrides with α-TiO-type structure.

We performed an exhaustive study of all the possible configurations of the systems, i.e. we studied all the different arrangements of the N and O atoms in the unit cells. For example, if we want to model a $TiN_{1-x}O_x$ system with NaCl-type structure, in which the N/O ratio is 0.5, we use the supercell (1x1x3) for TiN (which has 24 atoms), and we substitute 6 out of the 12 N atoms by O atoms. The number of different possibilities in which we can carry out the 6 substitutions is 940. In principle, if we wanted to make sure that we have found the most stable configuration of the system $TiN_{0.5}O_{0.5}$ we should perform the 940 geometry optimisations, which would be an impossible task, given the current computer time limitations. In order to perform the exhaustive study of all the configurations, while still using an acceptable amount of computer resources, we employed the SOD code, which makes use of the symmetry of the system to reduce drastically the number of configurations.

In the previous example, most of the 940 configurations are found to be equivalent. Two configurations are equivalent when they are related by an isometric operation (such as translations, rotations or reflections within the supercell, which are consistent with the symmetry operations of the crystal). Using the SOD code to remove the equivalent configurations we find that the number of non-equivalent configurations of the cited example is only 34, which is tractable with our computer resources. Employing the SOD code we also performed a statistical analysis of all the possible configurations of the $TiN_{0.5}O_{0.5}$ system, with which we obtained the energy of the system as a weighted average of the 940 configurations.

In order to study the $TiN_{1-x}O_x$ systems with NaCl-type structures, for x=0, 0.16, 0.33, 0.5, 0.66, 0.83 and 1, we substituted respectively 0, 2, 4, 6, 8, 10 and 12 of the 12 N atoms in the NaCl-type TiN supercell by O atoms. Using the SOD code we calculated the number of non-equivalent configurations, which is 1, 5, 21, 34, 21, 5 and 1 respectively. The total number of configurations we studied with the NaCl-type structure is therefore 88.

In the case of the TiN1-xOx systems with alpha-TiO structures, the supercell had 10 Ti atoms and 10 N or O atoms. The concentrations studied are x=0, 0.2, 0.4, 0.6, 0.8 and 1, which are achieved by substituting 0, 2, 4, 6, 8 and 10 of the 10 O atoms in the alpha-TiO supercell by N atoms. The number of non-equivalent configurations in this case is 1, 15, 60, 60, 15 and 1 respectively, giving a total number of 152. Note that, even with a smaller number of atoms in the supercell (20 as opposed to 24), the number of non-equivalent configurations in the case of the α-TiO structure is almost twice as large as that in the case of the NaCl-type structure. The reason of that is the high symmetry of the latter structure, which allows a great reduction of the number of configurations. The structural evolution of $TiN_{1-x}O_x$ compounds has been studied through the evolution of the formation energy with both the composition (x) and the structure (NaCl and α). The formation energy was calculated as follows:

$$E_f(TiN_{1-x}O_x) = E(TiN_{1-x}O_x) - nE(Ti\,\text{bulk}) - \frac{n(1-x)}{2}E(N_2) - \frac{nx}{2}E(O_2) \qquad (25)$$

Where *n* is the number of Ti atoms, E(Ti bulk) is the energy of the bulk of metallic Ti per Ti atom, and E(N_2) and E(O_2) are the energies of the isolated N_2 and O_2 molecules respectively.

Obviously, when *x* is close to 0 the system will have tendency to arrange itself as NaCl-type structure, since that is the most stable structure for TiN. Analogously, when x is close to 1 the system will try to arrange itself as alpha structure since this is the most stable structure for TiO. What we have to calculate is the limit composition at which the change of crystal structure takes place, and whether this limit composition depends on the temperature. **Figure 8** shows the evolution of the formation energy with the composition for both structures at 10 K. As it was predicted to happen, the NaCl-type structure is the preferred one for compositions close to x=0 (TiN), while the alpha structure is the most stable for compositions close to x=1 (TiO). The crossing point is found to be at the limit composition of x=0.55-0.60, approaching 0.60 as the temperature increases (curves were calculated at 10 K, 300 K and 600 K). The system will tend to acquire the NaCl-type structure for compositions x<0.6 while it will try to be ordered as alpha structure for compositions x>0.6. In the later case one should expect to find a number of vacancies in both Ti and N/O sublattices, since this is one of the main characteristics of the alpha structure. This is important from a technical point of view, since the presence of vacancies may change drastically the surface stability of the solid and generate highly reactive surfaces, which would be a serious drawback for microelectronic devices or technologies based on thin-films, but it could become interesting from a chemical point of view.

5. Conclusions

Site disorder is an important phenomenon that affects the structure and properties of materials. We have reviewed here some strategies for modeling and simulations to capture the physics and chemistry of disorder in influencing various properties of materials. These typically involve access to configurational information at different levels, like electronic properties, atomic displacements, and have varied computational cost. While cluster expansions have been used extensively in determination of phase diagrams of alloys and similar problems, we have emphasized here the methods that attempt essentially an exact statistical thermodynamic analysis using the SOD technique with a relatively smaller system, but having access to as much information and properties as possible. Such an approach is becoming quite practical in understanding and design of disordered materials, thanks to advances in computers and algorithms.

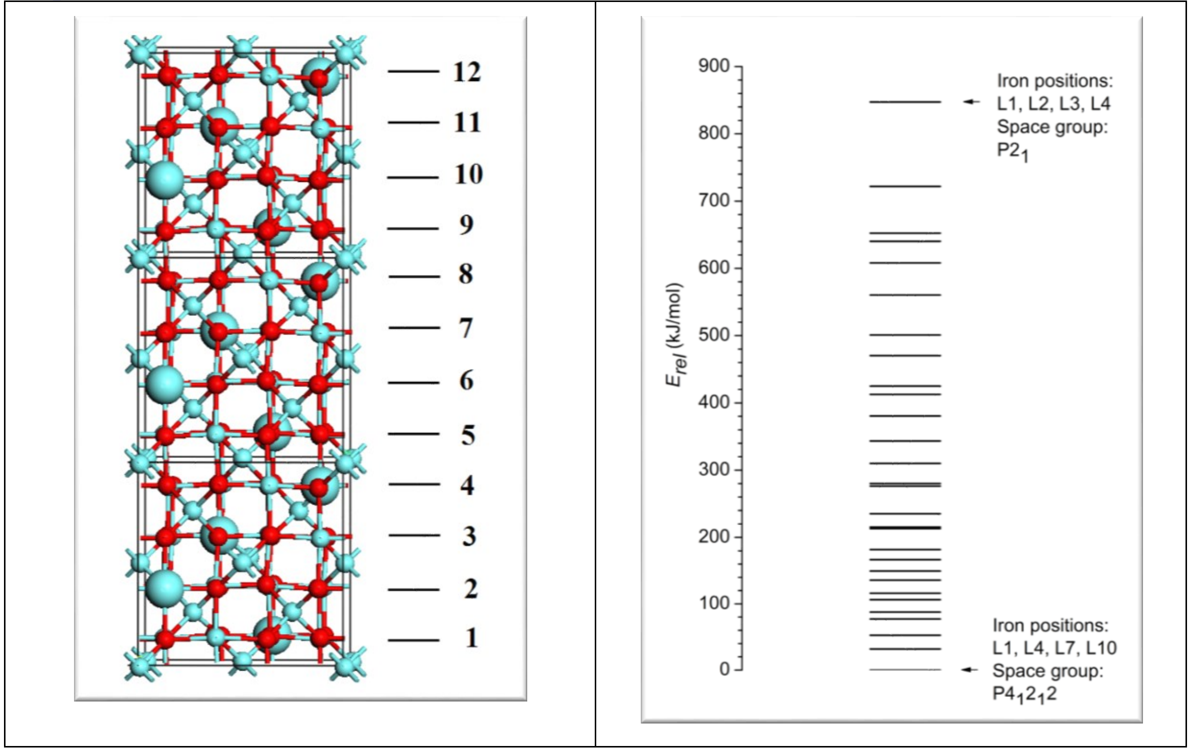


Figure 3. a) The exchangeable sites in the maghemite tetragonal cell: 4 Fe ions and 8 vacancies are distributed over these "L" sites. b) The calculated configurational spectrum.

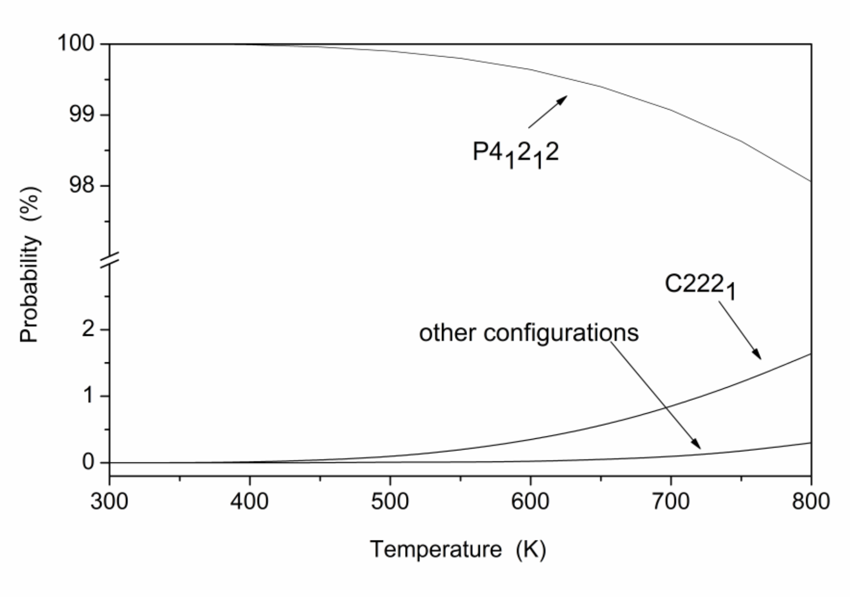


Figure 4. Probabilities of the two most stable configurations in the maghemite supercell as a function of temperature.

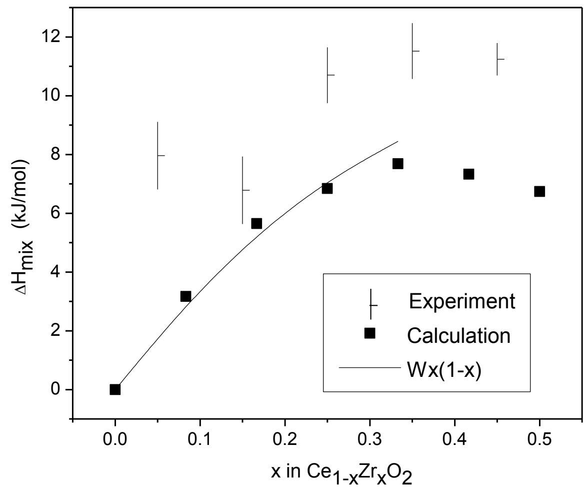


Figure 5. Calculated enthalpies of mixing for $Ce_{1-x}Zr_xO_2$ in comparison with experimental results [24]. The curved line represents the fitting of a regular-solution quadratic polynomial to the calculated values for low Zr concentrations.

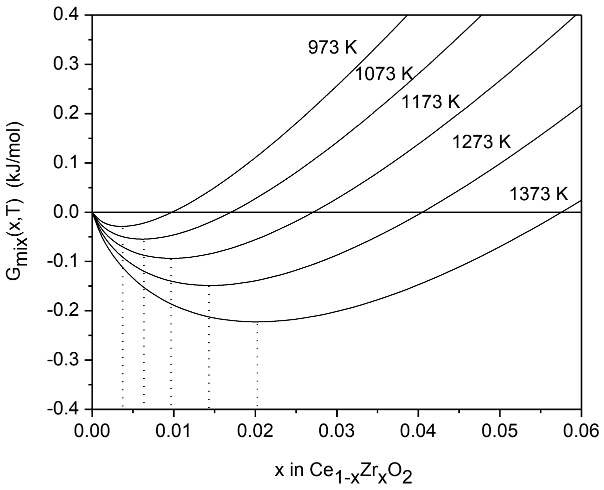


Figure 6. Free energies of mixing for low Zr concentrations. The vertical dotted lines mark the solubility limit of Zr in CeO_2 at the particular temperature.

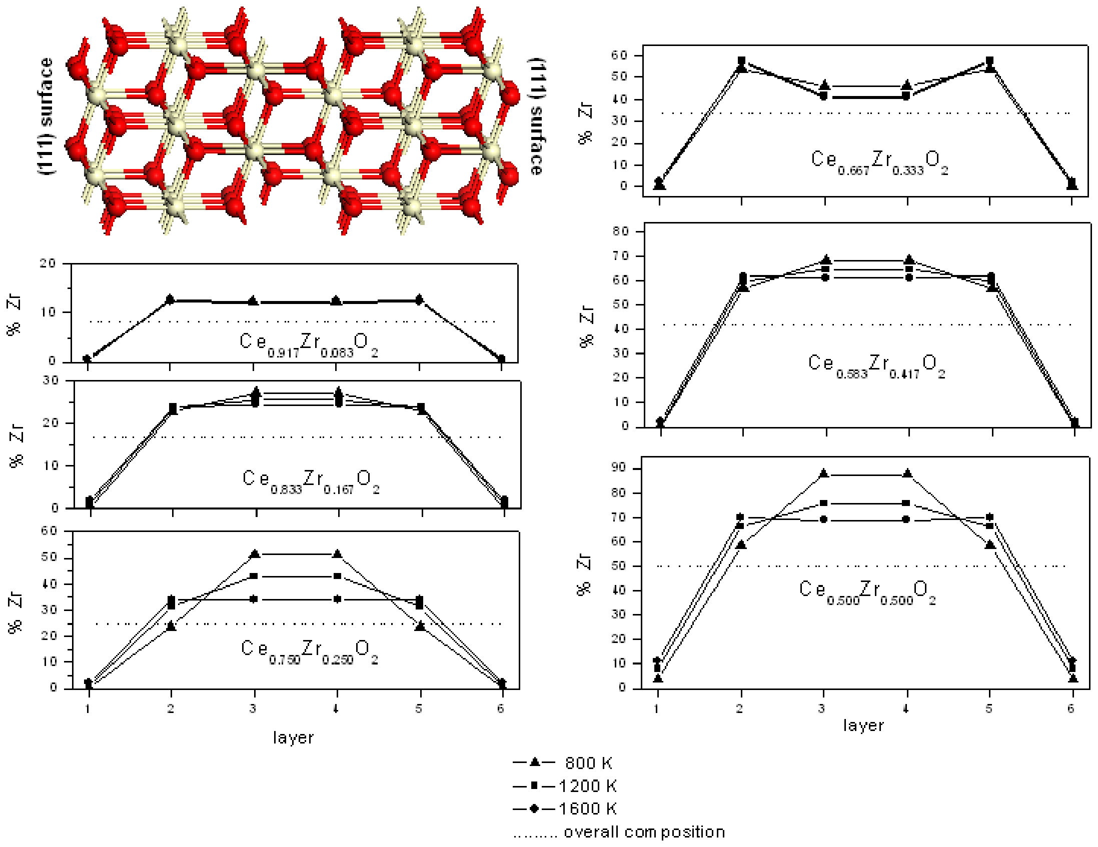


Figure 7. Calculated equilibrium concentrations of Zr as a function of the distance to the (111) surface in the $Ce_{1-x}Zr_xO_2$ solid solution. Because of the slab construction, layers 1, 2 and 3 are equivalent to layers 6, 5 and 4, respectively.

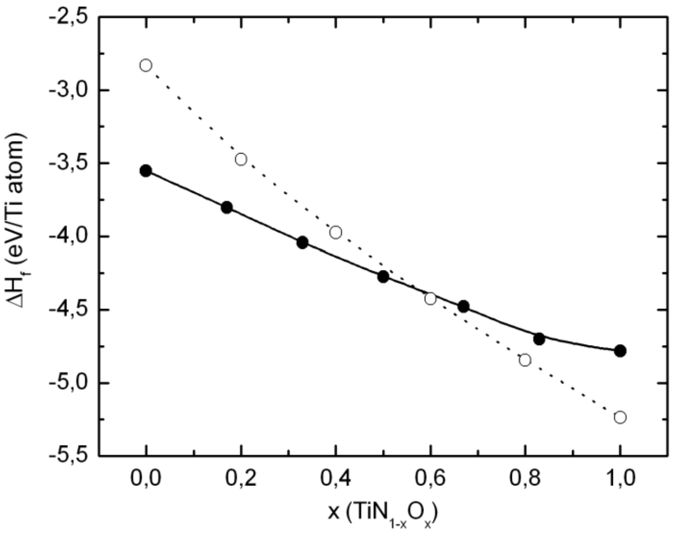


Figure 8. Evolution of the formation energy per Ti atom (eV) with the composition for the both structures NaCl-type (black circles) and α-type (white circles).

Acknowledgments

We are very grateful to all our co-authors in the case studies shown here, and in particular to Prof. Umesh Waghmare, Prof. Richard Catlow, and Prof. Nora de Leeuw for their contributions to the development of the ideas presented in this summary.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1 Gale, J. D. GULP: A computer program for the symmetry-adapted simulation of solids. *J. Chem. Soc.-Faraday Trans.* **93**, 629-637 (1997).

2 Gale, J. D. & Rohl, A. L. The General Utility Lattice Program (GULP). *Mol. Simul.* **29**, 291-341 (2003).

3 Zunger, A., Wei, S. H., Ferreira, L. G. & Bernard, J. E. Special quasirandom structures. *Physical Review Letters* **65**, 353 (1990).

4 Catlow, C. R. A. *Computer Modelling in Inorganic Crystallography*. (Academic Press Limited 1997).

5 van de Walle, A. & Ceder, G. The effect of lattice vibrations on substitutional alloy thermodynamics. *Reviews of Modern Physics* **74**, 11 (2002).

6 Becker, U., Fernandez-Gonzalez, A., Prieto, M., Harrison, R. & Putnis, A. Direct calculation of thermodynamic properties of the barite/celestite solid solution from molecular principles. *Physics and Chemistry of Minerals* **27**, 291-300 (2000).

7 Dove, M. T. *Introduction to Lattice Dynamics*. (Cambridge University Press, 1993).

8 Benny, S., Grau-Crespo, R. & De Leeuw, N. H. A theoretical investigation of α-Fe_2O_3– Cr_2O_3 solid solutions. *Physical Chemistry Chemical Physics* **11**, 808 - 815 (2009).

9 Grau-Crespo, R., Hamad, S., Catlow, C. R. A. & de Leeuw, N. H. Symmetry-adapted configurational modelling of fractional site occupancy in solids. *Journal of Physics-Condensed Matter* **19**, 256201 (2007).

10 Grau-Crespo, R. & Hamad, S. *SOD (Site-Occupancy Disorder)*, <https://sites.google.com/site/rgrauc/sod-program> (2007).

11 Dronskowski, R. The little maghemite story: A classic functional material. *Advanced Functional Materials* **11**, 27-29 (2001).

12 Pankhurst, Q. A., Connolly, J., Jones, S. K. & Dobson, J. Applications of magnetic nanoparticles in biomedicine. *J. Phys. D-Appl. Phys.* **36**, R167-R181 (2003).

13 Levy, M. *et al.* Magnetically induced hyperthermia: size-dependent heating power of gamma-Fe2O3 nanoparticles. *Journal of Physics-Condensed Matter* **20** (2008).

14 Waychunas, G. A. Crystal chemistry of oxides and oxyhydroxides. *Reviews in Mineralogy and Geochemistry* **25**, 11-68 (1991).

15 Braun, P. B. A superstructure in spinels. *Nature* **170**, 1123 (1952).

16 Shmakov, A. N., Kryukova, G. N., Tsybulya, S. V., Chuvilin, A. L. & Solovyeva, L. P. Vacancy Ordering in Gamma-Fe2o3 - Synchrotron X-Ray-Powder Diffraction and High-Resolution Electron-Microscopy Studies. *J Appl Crystallogr* **28**, 141-145 (1995).

17 Jorgensen, J. E., Mosegaard, L., Thomsen, L. E., Jensen, T. R. & Hanson, J. C. Formation of gamma-Fe2O3 nanoparticles and vacancy ordering: An in situ X-ray powder diffraction study. *Journal of Solid State Chemistry* **180**, 180-185 (2007).

18 Grau-Crespo, R., Al-Baitai, A. Y., Saadoune, I. & De Leeuw, N. H. Vacancy ordering and electronic structure of γ-Fe2O3 (maghemite): a theoretical investigation. *Journal of Physics: Condensed Matter* **22**, 255401 (2010).

19 Lewis, G. V. & Catlow, C. R. A. Potential Models for Ionic Oxides. *Journal of Physics C-Solid State Physics* **18**, 1149-1161 (1985).

20 Dick, B. G. & Overhauser, A. W. Theory of the dielectric constant of alkali halide crystals. *Physical Reviews* **112**, 90-103 (1958).

21 Gale, J. D. GULP: Capabilities and prospects. *Z. Kristallogr.* **220**, 552-554 (2005).

22 Grau-Crespo, R., De Leeuw, N. H., Hamad, S. & Waghmare, U. V. Phase separation and surface segregation in ceria-zirconia solid solutions. *Proceedings of the Royal Society A-Mathematical Physical and Engineering Sciences*, DOI:10.1098/rspa.2010.0512 doi:10.1098/rspa.2010.0512 (2011).

23 Cabanas, A., Darr, J. A., Lester, E. & Poliakoff, M. Continuous hydrothermal synthesis of inorganic materials in a near-critical water flow reactor; the one-step synthesis of nano-particulate $Ce_{1-x}Zr_xO_2$ (x=0-1) solid solutions. *J. Mater. Chem.* **11**, 561-568 (2001).

24 Lee, T. A., Stanek, C. R., McClellan, K. J., Mitchell, J. N. & Navrotsky, A. Enthalpy of formation of the cubic fluorite phase in the ceria–zirconia system. *Journal of Materials Research* **23**, 1105-1112 (2008).

25 Kresse, G. & Furthmuller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15-50 (1996).

26 Kresse, G. & Furthmuller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169-11186 (1996).

27 Prieto, M. Thermodynamics of Solid Solution-Aqueous Solution Systems. *Thermodynamics and Kinetics of Water-Rock Interaction* **70**, 47-85 (2009).

28 Ruiz-Hernandez, S. E., Grau-Crespo, R., Ruiz-Salvador, A. R. & De Leeuw, N. H. Thermochemistry of strontium incorporation in aragonite from atomistic simulations. *Geochimica Et Cosmochimica Acta* **74**, 1320-1328 (2010).

29 Du, Y., Yashima, M., Koura, T., Kakihana, M. & Yoshimura, M. Thermodynamic evaluation of the ZrO_2–CeO_2 system. *Scripta Metall. Mater.* **31**, 327 (1994).

30 Shannon, R. D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallographica* **A32**, 751-767. (1976).

31 Cabanas, A., Darr, J. A., Lester, E. & Poliakoff, M. A continuous and clean one-step synthesis of nano-particulate $Ce_{1-x}Zr_xO_2$ solid solutions in near-critical water. *Chemical Communications*, 901-902 (2000).

32 Garvie, R. C. in *High temperature oxides. Part II* (ed M.A. Alper) 117 (Academic Press, 1970).

33 Navrotsky, A., Benoist, L. & Lefebvre, H. Direct calorimetric measurement of enthalpies of phase transitions at 2000 degrees–2400 degrees C in yttria and zirconia. *Journal of the American Ceramic Society* **88**, 2942 (2005).

34 Graciani, J., Hamad, S. & Sanz, J. F. Changing the physical and chemical properties of titanium oxynitrides TiN(1-x)Ox by changing the composition. *Phys. Rev. B* **80**, 184112 (2009).

SciForum
Mol2Net

Categorization of Continuous Variables in a Logistic Regression Model Using the R Package CatPredi

Irantzu Barrio [1,*], María-Xosé Rodríguez-Álvarez [2] and Inmaculada Arostegui [1,3]

[1] Departamento de Matemática Aplicada, Estadística e Investigación Operativa, Universidad del País Vasco UPV/EHU, Leioa, Spain; E-Mail: inmaculada.arostegui@ehu.eus

[2] Departamento de Estadística e Investigación Operativa, Universidade de Vigo, Vigo, Spain; E-Mail: mxrodriguez@uvigo.es

3 BCAM—Basque Center for Applied Mathematics, Bilbao, Spain

* Author to whom correspondence should be addressed; E-Mail: irantzu.barrio@ehu.eus; Tel.: +34-94-601-2504.

Published: 4 December 2015

Abstract: Prediction models are gaining importance in many areas such as medicine, meteorology, finance, toxicology, etc. In this context, a common distribution for the response variable is the binomial distribution and hence the logistic regression model is a commonly used regression modeling approach. Although it is not recommended from a statistical points of view due to loss of information and power, the categorization of continuous variables is a common practice in the development of prediction models. However, there are no unified criteria for the selection of the cut points in the categorization process. In order to provide valid cut points whenever a categorization is going to be performed, we have developed a valid methodology to categorize continuous variables in a logistic regression model based on the maximization of the AUC. This methodology has been implemented in an `R` package called `CatPredi`. This is a package of R functions that allows the user to categorize a continuous predictor variable in a univariate or multiple logistic regression model. It provides the optimal location of cut points for a chosen number of cut points and returns the estimated and bias-corrected discriminative ability index for this model. Additionally, it allows a comparison of two categorization proposals for different number of cut points and the selection of the optimal number of cut points.

Keywords: categorization; R package; prediction model

Mol2Net YouTube channel: *http://bit.do/mol2net-tube*
YouTube link: *please, paste here the link to your personal YouTube video, if any.*

1. Introduction

Prediction models are gaining importance in many areas such as medicine, meteorology, finance, toxicology, etc. In this context, a common distribution for the response variable is the binomial distribution and hence the logistic regression model is a commonly used regression modeling approach. Although it is not recommended from a statistical points of view due to loss of information and power, the categorization of continuous variables is a common practice in the development of prediction models. However, there are no unified criteria for the selection of the cut points in the categorization process. In order to provide valid cut points whenever a categorization is going to be performed, we have developed a valid methodology to categorize continuous variables in a logistic regression model based on the maximization of the discriminative ability of the model measured by the area under the ROC curve – AUC.

2. Methods

We have developed a methodology to categorize continuous variables in a logistic regression model. The proposed methodology consists on the maximization of the AUC. Two alternative algorithms have been proposed to select the optimal cut points to categorize continuous variables named *AddFor* and *Genetic* respectively. This methodology has been presented elsewhere (Barrio et al. 2015). This methodology has been implemented in an `R` (R Core Team 2015) package which is explained below.

3. The `CatPredi` Package

`CatPredi` is a package of `R` functions that allows the user to categorize a continuous predictor variable either before or during the development of a prediction model. `The CatPredi` package can be used to categorize a predictor variable in a univariable or a multivariable setting. It provides the optimal location of cut points for a chosen number of cut points, fits the prediction model with the categorized predictor variable and returns the estimated and bias-corrected discriminative ability index for this model. Additionally, it allows a comparison of two categorization proposals for a different number of cut points and the selection of the optimal number of cut points.

The `CatPredi` package has been designed similarly to other packages in R. It has a main function called `catpredi()` which categorizes a continuous predictor variable in a logistic regression model.

Numerical and graphical summaries of the fitted objects can be obtained by using `print.catpredi, summary.catpredi` and `plot.catpredi` for `catpredi` type objects. Furthermore, one more main function has been developed `comp.cutpoints` to obtain the optimal number of cut points in a logistic regression model. Table 1 contains a description of all the functions available in the package.

Below, we give a general overview of the package and its general use.

2.1 catpredi() function

The `catpredi()` function provides the optimal cut points to categorize a continuous predictor variable in a logistic regression model.

This function creates an object of class `catpredi`. The main arguments of this function are presented in Table 2. The call to the function is as follows:

```
catpredi(formula, cat.var, cat.points, data, method = c("addfor","genetic"),range=NULL, correct.AUC=TRUE, control = controlcatpredi())
```

In the formula argument users must specify the prediction model setting in which they want to categorize the predictor variable *X* specified in the `cat.var="X"` argument. If the model is a univariate logistic regression model, then the formula would be specified as `Y~1`, with *Y* being the response variable available in the data set specified in the argument `data`. However, if the model is a multiple logistic regression model, and the aim is to categorize the predictor variable *X* together with another predictor *Z*, then the formula would be specified as `Y~Z`. Additionally, in the argument `cat.points` the user must specify the number of cut points to look for. The range argument allows for modifying the range of the predictor variable *X* in which to look for the cut points. By default it would be NULL, which represents the entire range of *X*. Finally, if `correct.AUC` is set to TRUE, the bias-corrected AUC would be estimated.

A numerical summary of the results of the categorization method can be obtained by calling the functions `print.catpredi()` or `summary.catpredi()`. When the method selected is the *AddFor*, the summary returns the estimated AUC for each of the selected cut points. For example, if `cat.points = 2` is chosen, it returns the estimated AUC for one and two cut points. Additionally, if `correct.AUC=TRUE` is chosen it returns the bias-corrected AUC for two cut points. If the method selected is the *Genetic*, estimated cut points, AUC and bias-corrected AUC will be given only for the selected number of cut points.

Table 1. Summary of the functions in the `CatPredi` package.

Function	**Description**
`catpredi()`	Returns an object with the optimal cut points to categorize a continuous predictor variable in a logistic regression model.
`controlcatpredi()`	Function used to set several parameters to control the selection of the optimal cut points in a logistic regression model
`print.catpredi()`	Print method for objects of type `catpredi`.
`summary.catpredi()`	Produces a summary of the `catpredi` object.
`plot.catpredi()`	Plots the relationship between the continuous predictor and the response variable obtained by fitting a Generalized Additive Model (GAM), together with the location of the optimal cut points.
`comp.cutpoints()`	Compares two objects of type `catpredi`.
`print.comp.cutpoints()`	Print method for objects of type `comp.cutpoints`

Table 2. Summary of the arguments in the `catpredi()` function.

Argument	Description
`formula`	A formula giving the model to be fitted.
`cat.var`	Name of the continuous variable to categorize.
`cat.points`	Number of cut points to look for.
`data`	Data frame containing all needed variables.
`method`	The algorithm selected to search for the optimal cut points- `"addfor"` if the *AddFor* algorithm is chosen; otherwise, `"genetic"`.
`range`	The range of the continuous variable in which to look for the cut points. By default `NULL`, i.e., the entire range.
`correct.AUC`	A logical value. If `TRUE` the bias-corrected AUC is estimated.
`control`	Output of the `controlcatpredi()` function.

4. Conclusions

We have developed a user-friendly R package, named `CatPredi`, to obtain optimal cut points to categorize continuous predictor variables in a logistic regression model in practice, either in a univariable or multivariable setting. The `CatPredi` package can be freely download from https://sites.google.com/site/biostit/lineas-de-investigacion/software/catpredi.

Acknowledgments

This work was supported by grants MTM2010-14913, 2012111008, IT620-13, UFI11/52, MTM2011-28285-C02-01, MTM2013-40941-P MTM2014-55966-P. MXRA thanks the Agrupamento INBIOMED from DXPCTSUG-FEDER unha maneira de facer Europa (2012/273).

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Barrio, I.; Arostegui, I; Rodríguez-Álvarez, MX; Quintana, JM. A new approach to categorising continuous variables in prediction models: Proposal and validation. *Statistical Methods in Medical Research* **2015 (in press)**.

2. R Core Team. R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. 2015.

SciForum
Mol2Net

Applying a Novel Web-Tool for Performing Virtual Screening Experiments

Vinicius R. Seus [1],*, Jorge Gomes [2] and Karina S. Machado [2]

[1] Universidade Federal do Rio Grande—FURG, Campus Carreiros: Av. Itália km 8 Bairro Carreiros

[2] Universidade Federal do Rio Grande—FURG, Campus Carreiros: Av. Itália km 8 Bairro Carreiros; E-Mails: jorge_hcg@hotmail.com (J.G.); karina.machado@furg.br (K.S.M.)

* Author to whom correspondence should be addressed; E-Mail: viniciusseus@gmail.com or E-Mail: viniciusseus@furg.br.

Received: 22 October 2015 / Accepted: 30 November 2015 / Published: 4 December 2015

Abstract: The use of *in silico* methods for identifying new drugs to a target of interest is a step of a process called Rational Drug Design. The *insilico* analysis of a set of drug candidates is performed by a computational technique named as Virtual Screening (VS). In a previous work, we have developed a novel web tool for configuring different types of VS experiments using AutoDock Vina docking software. The presented tool is a framework that generates python scripts to run VS experiments in the users' computer according to the users' configuration on the framework web interface. In this paper we propose to apply the developed framework in a specific VS experiment considering one target receptor and a set of ligands. For this VS experiment the researcher informs the location of receptor and the ligands files as well as their formats. It is also possible to set receptor and ligand flexibility. After this, the user indicates the output folder where all the results in the user's computer will be stored after the script execution. Then, the user should configure the box area that indicates where ligands will be docked in the receptor molecule. The box size and the center must be configured, the variation of box center could be configured if user wants to execute an experiment that search the binding site in the entire molecule. In this way, this paper demonstrates the usage of the proposed framework for VS where we considered as receptor the structure of the human voltage-dependent anion channel (PDB code: 2JK4) and as ligands different types of carbon nanotubes. In the experiments performed we have defined both receptor and ligands as rigid and considered only one box for representing the receptor binding site.

Keywords: Protein Data Bank; Virtual Screening; Rational Drug Design; Molecular Docking.

1. Introduction

In Bioinformatics, one of most important research area is the Rational Drug Design (RDD) [3], a process where the costs and time involved are high. One of the RDD steps is the molecular docking, a computational technique used for the approximate determination of the interaction energy between the macromolecule receptor and a ligand candidate inhibitor [5]. One of the major challenges in RDD is related to the understanding of the behavior of biological macromolecules receptors as proteins and how they interact with a set of different small molecules or ligands, a strategy called Virtual Screening (VS) [2].

Here, in order to approximate the in-silico step of RDD process to the in vitro and in vivo tests the VS should consider that under physiological conditions, biomolecules experience various types of movement and conformational changes often crucial to their functions [7]. This flexible behavior of biological macromolecules can be simulated by molecular dynamics (MD) trajectories [2]. To incorporate flexibility in the receptors in a VS process, a possible approach is the implementation of a series of molecular docking using in each experiment a different receptor conformation generated by DM [4].

For the execution of a VS process, in addition to the target receptor (rigid or flexible), it is necessary a set of compounds which may be obtained in different public databases. These ligands should be in a specific file format and can be considered as rigid or flexible.

Another problem is when the target binding site is unknown. Thus, the entire conformational space of the target molecule should be investigated by further increasing the complexity of this kind of experiment. In a previous work we proposed a framework for virtual screening where the user can easily configure a VS experiment[7]. Different types of configurations can be set by researchers, from the generation of a simple experiment containing a receptor and a binder to a more complex experiment that will contain in its implementation, several receptors with various ligands and also the possibility of variation box (receptor molecule coordinate where the drug candidates, also called ligands, will be tested).

Therefore, this paper aims at presenting the application of the previously web framework for a VS experiment. In this experiment we consider only one rigid receptor and a group of different carbon nanotubes. This framework uses the AutoDock Vina [8] to perform the molecular docking.

.

2. Results and Discussion

In this section we present the application of the proposed framework for a virtual screening experiment. The experiment performed to analyze the application of the framework consists in one target receptor, called human voltage-dependent anion channel (PDB ID: 2JK4) and a set of single walled carbon nanotubes (SWCNT-pristine amchair).

The framework's web interface consists in two parts. Part one (Figure 1) is called Input/Output Area where the configuration of the receptor(s) and ligand(s) are filled. The part two (Figure 2) is called Box Area where the box size and the center of the box (receptor molecule coordinate where the drug candidates will be tested) are filled.

Firstly, according to Figure 1, before any receptor or ligand configuration is performed the user must answer a question about the operational system (OS) where the generated

Python script is going to be performed. For this case it was selected "Linux" OS. The second question regards the amount of times that the user wants to run the same experiment. This is implemented in the framework since the AutoDock Vina starts its tests by choosing a random initial position for the ligand inside the delimited box, a parameter called seed. For this case "No" was selected.

The next area to configure is the receptor, it provides to the user the possibility of selecting a flexible receptor model or only one receptor structure. In the present case, we considered as the receptor the protein human voltage-dependent anion channel PDB ID: 2JK4 that is stored in the path "/home/script_vs/receptor/" according to figure 1. The receptor type was selected as PDBQT. The PDBQT format indicates that this receptor input file has already the charges for each atom and hydrogens were added. After this, the user needs to configure the ligand area. In this case, the ligand is one set of SWCNT (SWCNT-pristine amchair) and it is located in "/home/script_vs/ligand/". It is important to notice that the ligand type was selected as "MOL2 (file with multiple structures)" thus the the file "SWCNT.mol2" can have either one or a set of ligands. In our case study, the input ligand file is composed by many ligand files that corresponds to different types of carbon nanotubes. This functionality was implemented when user wants to perform a virtual screening considering many ligands in an experiment. In addition, the user can choose to consider the ligand(s) as rigid by selecting the checkbox *Rigid Ligand or flexible by not selecting the checkbox.*

Another parameter of a VS experiment (Figure 1) is the "Output Folder" that indicates where results should be stored on the user's computer during the execution of the Python script. If the user selects the type of receptor or ligand input file different of PDBQT format, the MGL Tools space appears in the interface called *MGL Tools path*. This path is provided by the user indicating where the MGL Tools are installed on his own computer.

The Box area (Figure 2), is the module where the user configures the simulation box that represents the binding site of receptor. During the molecular docking simulations AutoDock Vina analyses the best conformation and orientation of the ligand(s) inside the framework's interface *Box section* according to the figure 2. For the present experiment we configured the field *Box Size* with the following values: X = 20, Y = 20 and Z = 20. For the *Box Center* field we configured the following values: X = 28.097, Y = 3.078 and Z = 8. The user can also configure a variation on the box: the user fills in an initial area (X, Y and Z fields of the Box center area), a final area (X-final, Y-final and Z-final in Box center area) and a particular step (X-step, Y-step and Z-step in the Box Center area).

After the configuration of the Box Area, the user needs to answer a question in relation to the analysis module. This module allows the user to obtain a script to perform the analysis of the receptor-ligand interaction using the software LigPlot[7]. The question in the interface is "Do you want LigPlot results?". According to the figure 1, we have answered "Yes", thus a script that have to be executed after the mais script execution is generated. The user needs to inform where LigPlot is installed in his own computer in the field "LigPlot Folder" as presented in figure 2.

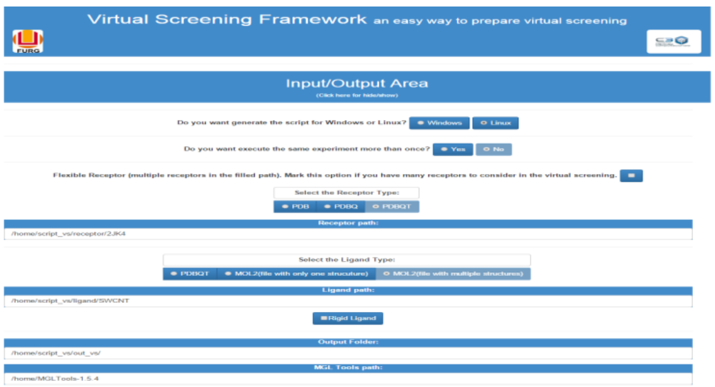
The LigPlot script reads all results generated by the main script creating spreadsheets to facilitate the analysis of contact between residues of receptor and ligand atoms. Finally, to generate the main script, the user needs to click in "Generate Script" and execute it in his own computer to perform the VS experiment.

For the performed experiment we consider SWCNT-pristine amchair. The average FEB was -14.44 and the standard deviation was 2.38. It indicates that almost all tested carbon nanotubes had a good interaction with the protein human voltage-dependent anion channel.

.

.

.


Virtual Screening Framework an easy way to prepare virtual screening

Input/Output Area
(Click here for hide/show)

Do you want generate the script for Windows or Linux? ● Windows ○ Linux

Do you want execute the same experiment more than once? ● Yes ○ No

Flexible Receptor (multiple receptors in the filled path). Mark this option if you have many receptors to consider in the virtual screening.

Select the Receptor Type: ● PDB ● PDBQ ○ PDBQT

Receptor path: /home/script_vs/receptor/2JK4

Select the Ligand Type: ● PDBQT ● MOL2(file with only one strucuture) ○ MOL2(file with multiple structures)

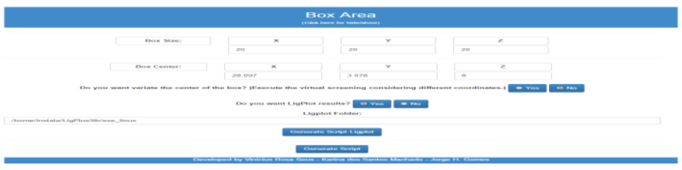
Ligand path: /home/script_vs/ligand/SWCNT

Rigid Ligand

Output Folder: /home/script_vs/out_vs/

MGL Tools path: /home/MGLTools-1.5.4

Figure 1. Input/Output section in the Framework interface.


Box Area
(Click here for hide/show)

	X	Y	Z
Box Size:	20	20	20
Box Center:	28.097	3.078	8

Do you want variate the center of the box? (Execute the virtual screening considering different coordinates.) ● Yes ○ No

Do you want LigPlot results? ○ Yes ● No

Ligplot Folder: /home/instala/LigPlus/lib/exe_linux

Generate Script Ligplot

Generate Script

Developed by Vinicius Rosa Seus - Karina dos Santos Machado - Jorge H. Gomes

Figure 2. Box section in the Framework interface.

3. Materials and Methods

This section presents the materials and methods applied in the development of the proposed experiment using the VS experiment.

This is a client-server architecture implementation that works in the web environment. For performing the VS experiment the user needs to have installed Python and AutoDock Vina in his own computer.

For the development of the protein-ligand interaction module the software LigPlot was integrated to our tool. This software also needs to be installed on user's computer.

According to [9], LigPlot is a software that evaluates the intermolecular interactions that occurs in a receptor-ligand complex formation.

. The main objective of this work was to implement a tool that generates scripts enabling users to perform different types of VS experiments.

In relation to the experiment used as example, most of the results were good according to the average FEB, showing that interactions between receptor and ligands for all experiments are present.

With this tool, researchers from different fields can perform VS experiments according to their necessities. More importantly, with the use of the presented tool the VS process is automated making it easier to run complex VS experiments..

4. Conclusions

The main objective of this work was to implement a tool that generates scripts enabling users to perform different types of VS experiments.

In relation to the experiment used as example, most of the results were good according to the average FEB, showing that interactions between receptor and ligands for all experiments are present.

With this tool, researchers from different fields can perform VS experiments according to their necessities. More importantly, with the use of the presented tool the VS process is automated making it easier to run complex VS experiments.

Acknowledgements

K. Machado acknowledges the support from CNPQ (Process number 477462/2013-8). Vinicius Rosa Seus receives a CAPES Master Degree fellowship.

References and Notes

1. Cavassoto, C. N., Orry, A. J.. Ligand docking and structured-based virtual screening in drug discovery. *Current Topics in Medicinal Chemistry.* **2007**, *7(10)*, 1006-14.
2. Durrant, J., Mccammon, J.. Molecular dynamics simulation in drug discovery. *BMC Biology* **2011**, *9(71)*, 1-9.
3. Kuntz, I. D.. Structure-Based Strategies for Drug Design and Discovery. *Science* **1992**, *257(5073)*, 1078–1082.
4. Lin, J-H., Perryman, A. L., Schames, J. R., Mccammon, J. A.. Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *Journal of the American Chemical Society* **2002**, *124,* 5632-5633.
5. Morris, G. M., Huey, R., Lidstrom, W., Sanner, M., Belew, R. K., Goodsell, D. S., Olson, A. J. AutoDock 4 and AutoDock Tools 4: automated docking with selective receptor flexibility. *J. Computational Chemistry* **2009**, 16, 2785-91.
6. Seus, V. R., Machado, K. S.. A Framework for Virtual Screening. In: BSB & Xmeeting 2013, **2013**, Recife-PE. Xmeeting & BSB Abstracts Book **2013**. v.1 p. 199-199.
7. Totrov, M.; Abagyan R.. Flexible ligand docking to multiple receptor conformations: a pratical alternative. *Current Opinion in Structural Biology.* **2008**, 178-184.
8. Trott, O., Olson, A. J.. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Computational Chemistry* **2010**, *31*, 455-461.
9. Wallace, A. C., Laskowski, R. A., Thornton, J. M.. Ligplot: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **1996**, *8*, 127–134.

SciForum
Mol2Net

Prediction of the Total Antioxidant Capacity of Food Based on Artificial Intelligence Algorithms

Estela Guardado Yordi [1,2]*, Raúl Koelig [1], Yailé Caballero Mota [1], Maria João Matos [2], Lourdes Santana [2], Eugenio Uriarte [2]and Enrique Molina [1,2]

[1] Universidad de Camagüey Ignacio Agramonte Loynaz, Circunvalación Norte Km 51/2, Camagüey, Cuba; E-Mials: raul.koelig@reduc.edu.cu (R.K.); yaile.caballero@reduc.edu.cu (Y.C.M.); enrique.molina@reduc.edu.cu (E.M.)

[2] Universidade de Santiago de Compostela, Campus Vida, Santiago de Compostela; E-Mials: mariacmatos@gmail.com (M.J.M.); lourdes.santana@usc.es (L.S.); eugenio.uriarte@usc.es (E.U.)

* Author to whom correspondence should be addressed; E-Mail: estela.guardado@reduc.edu.cu; Tel.: +53-32-261192.

Received: 12 October 2015 / Accepted: 9 November 2015 / Published: 4 December 2015

Abstract: The growing increase in the amount and type of nutrients in food created the necessity for a more efficient use applied to dietetics and nutrition. Flavonoids are exogenous dietetic antioxidants and contribute to the total antioxidant capacity of the food. This paper aims to explore the data using different algorithms of artificial intelligence to find the one that best predict the total antioxidant capacity of food by the oxygen radical absorbance capacity (ORAC) method. A record of composition data based on the Database for the Flavonoid Content of Selected Foods and the Database for the Isoflavone Content of Selected Foods, was created. The KNN (K-Nearest Neighbors) and supervised unidirectional networks MLP (MultiLayer Perceptron) technics were used. The attributes were: a) amount of flavonoid (mean), b) class of flavonoid, c) Trolox equivalent antioxidant capacity (TEAC) value of each flavonoid, d) probability of clastogenicity and clastogenicity classification by Quantitative Structure-Activity Relationship (QSAR) method and e) total polyphenol (TP) value. The variable to predict the activities was the ORAC value. For the prediction, a cross-validation method was used. For the KNN algorithm the optimal K value was 3, making clear the importance of the similarity between objects for the success of the results. It was concluded the successful use of the MLP and KNN techniques to predict the antioxidant capacity in the studied food groups.

Keywords: Flavonoid, Artificial intelligence, MultiLayer Perceptron algorithm, K-Nearest Neighbors algorithm.

Mol2Net YouTube channel: http://bit.do/mol2net-tube

1. Introduction

In the recent years, database including information about the emerging food composition database were created [1-3]. These databases are centred in the composition of bioactive substances including flavonoids. Flavonoids are present in several sources in the vegetal kingdom and display a large range of biological properties. They are already proved their benefits for health [4,5]. Therefore, their study is a topic of interest. One of the most important activities is related to their antioxidant capacity [1,6-8].

An antioxidant is a substance, even in small amounts comparing to the substrate, that is able to decrease the oxidation of that substrate [9]. Furthermore, the antioxidant activity is correlated to the prevention of chronic diseases of high prevalence in different countries [10].

The food composition database of flavonoids has huge chemical information due to the structural diversity of the compounds included on it. This database provides researchers with new values on the flavonoid content of many foods in order to better ascertain the impact of flavonoid consumption against several chronic diseases [2, 3]. Flavonoids, particularly flavan-3-ols, have been associated with the reduction in the risk of cardiovascular diseases by modulating different mechanisms of primary and secondary prevention [11].

This project was developed taking into account the possibility of generating predictive information related to the data found in the food composition database. In particular, we were looking for a tool to predict the antioxidant capacity of food containing different compounds with flavonoid scaffold (dietary exogenous antioxidants). This project was focus on the idea that a dietary antioxidant is a substance that significantly decreases the adverse effects of reactive species, such as reactive oxygen and nitrogen species, on normal physiological function in humans [12].

The data regarding the composition of food is complex and extensive [13]. Therefore, it is hard to process all the information regarding the different assays presented in the literature. However, the processing of the information is still performed by classic statistical methodologies [14,15]. When the problem is complex and mediated by non-lineal behaviours, it could be studied either by a multivariate perspective or by using artificial intelligence technics (AI) [16]. In particular, artificial neuronal networks (ANN) are able to develop a predictive model that automatically includes relationships between the analysed variables, with no necessity of included them in the model.

In the biomedical field, several unidirectional supervised networks were used, specially based on the MultiLayer Perceptron (MLP) [16]. However, as far as we know, these technics were never been used related to food composition databases. Therefore, the current work is centred in the development of an AI algorithm that allow the prediction of the total antioxidant capacity of the food, based on quantitative information, topologic-structural and the bioactivity of flavonoids.

2. Results and Discussion

The studied food was divided in 11 groups (**Figure 1**). Vegetables, vegetable spices and herbs, are the groups with more flavonoid-containing aliments: 39 % and 37 %, respectively.

The monomeric dietary flavonoids present in the studied data (class of flavonoid attribute) are from the chemical subclasses: flavonols, flavones, flavanones and flavan-3-ols (**Table 1**). Flavonoids from the anthocyanidin subclass can be found in several aliments. However, they were not included in this study due to their structure, which invalided the application of Topological Substructural Molecular Design, TOPSMODE approach [17].

In **Table 2** it is shown the chemical structure, the SMILE codes and some examples of sources of the studied flavonoids.

Oxygen radical absorbance capacity (ORAC) was the studied parameter related to the antioxidant potential of the studied compounds. The results obtained in the predictions show that the assigned weights to each attribute were correct.

Figure 2 shows the obtained prediction by the KNN algorithm for the conjuncts # 1-5. X represents the number of rows in the database, in which everyone has an ORAC value represented in Y. In the graphics, it is possible to notice the correspondence between the predicted and the experimental ORAC values. The prediction resulted better when the method PSO+RST was used.

The results obtained with MLP algorithm (**Figure 3**) showed a less exact prediction. In this graphic it is possible to correlate the real and the predicted values. The ranges of error when algorithms were applied are: MAE (1.7601); RMSE (4.1569).

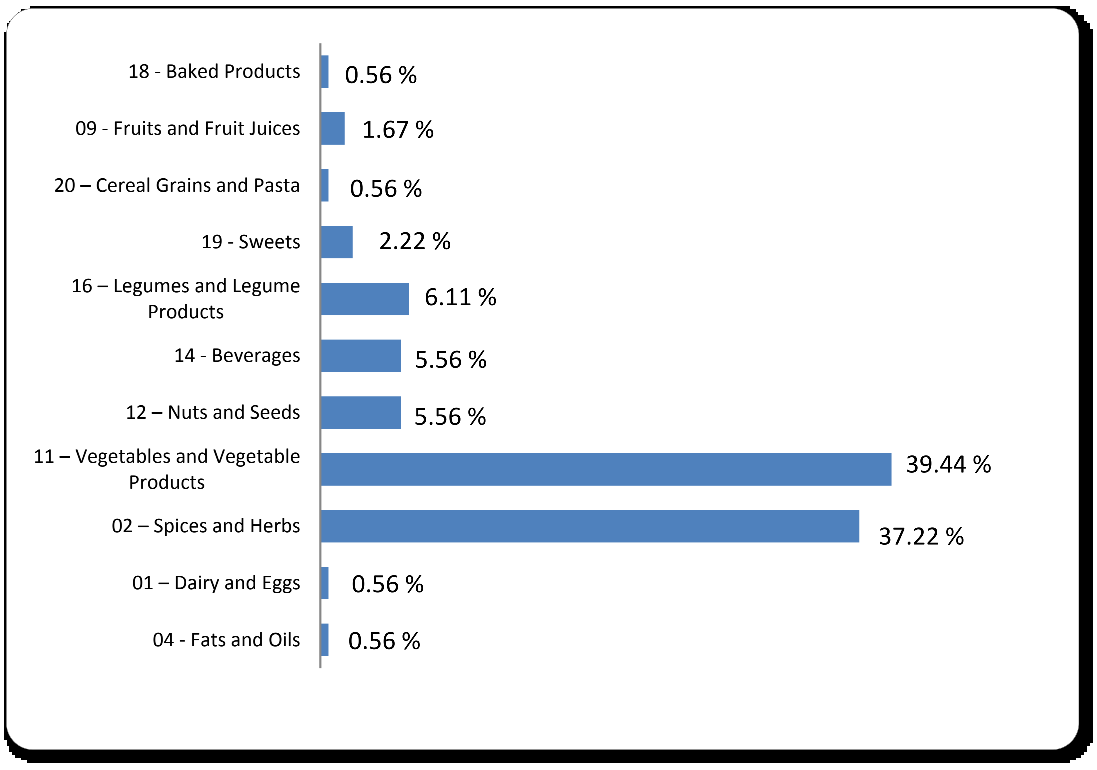


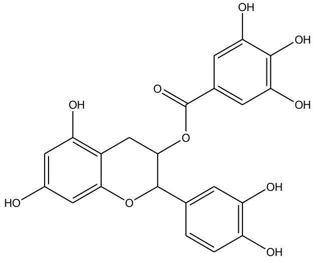
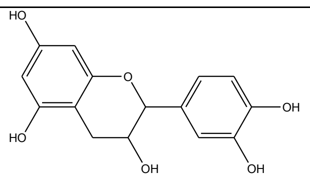
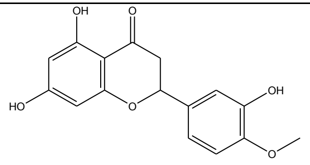
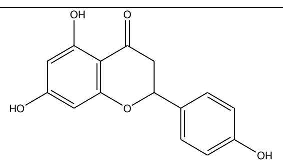
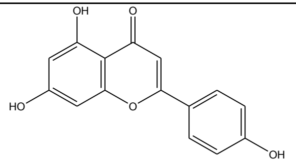
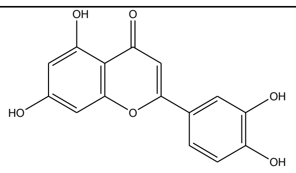
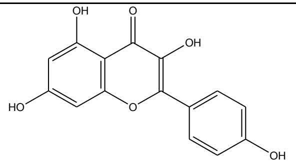
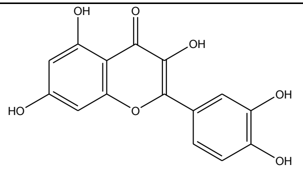
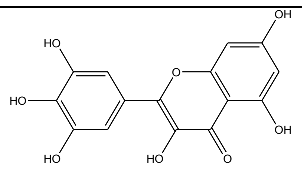
Figure 1. Percentage represented by each alimentary group in the studied data.

Table 1. Examples of the conformation of the data and the respective attributes.

ALIMENTARY GROUP	NDB CODE	FOOD	ATTRIBUTES										CLASS ATTRIBUTE (ORAC EXP)	
			Flavonoid	Class of flavonoid	Amount of flavonoid (mean)	Flavonoid TEAC value		Probability of clastogenicity activity by QSAR methods			TPexp		Mean	Ref
						TEAC exp	Ref	% prob.	Classif.	Ref	mean	Ref		
11 – Vegetables and Vegetable Products	11090	Broccoli, raw (*Brassica oleracea var. italica*)	(+) -Catechin	Flavan-3-ols	0	2.4	[18]	60.5	G2:1	[19]	316	[20]	1510	[20-23]
			(-)-Epigallocatechin 3-gallate	Flavan-3-ols	0	4.93	[18]	84.6	G2:1	*				
			Hesperetin	Flavanones	0	1.37	[18]	85.9	G2:1	[19]				
			Naringenin	Flavanones	0	1.53	[18]	52.6	G2:1	[19]				
			Apigenin	Flavones	0	1.45	[18]	66.2	G1:-1	[19]				
			Luteolin	Flavones	0.8	2.09	[18]	51.3	G2:1	[19]				
			Kaempferol	Flavonols	7.84	1.34	[18]	53.0	G2:1	[19]				
			Myricetin	Flavonols	0.06	3.1	[18]	79.6	G2:1	[19]				
			Quercetin	Flavonols	3.26	4.7	[18]	67.6	G2:1	[19]				
02 – Spices and Herbs	99428	Guava, red-fleshed	Apigenin	Flavones	0	1.45	[18]	66.2	G1:-1	[19]	247	[24]	1990	[24]
			Luteolin	Flavones	0.8	2.09	[18]	51.3	G2:1	[19]				
			Kaempferol	Flavonols	0	1.34	[18]	53.0	G2:1	[19]				
			Myricetin	Flavonols	0	3.1	[18]	79.6	G2:1	[19]				
			Quercetin	Flavonols	1	4.7	[18]	67.6	G2:1	[19]				

*Unpublished; Trolox equivalent antioxidant capacity flavonoid value (TEACexp); Total polyphenol value (TPexp); Ref (Literature reference); prob. (Probability); classif. (Classification by Linear Discriminate Analysis (LDA) method); G2:1 (active); G1:-1 (inactive).

Table 2. Examples of the chemical information of flavonoids and their presence in food contained in the studied database.

FLAVONOIDS	STRUCTURE	SMILE	FOOD (NDB No.)*
(-) -Epicatechin 3-gallate		C1C(C(OC2=CC(=CC(=C21)O)O)C3=CC(=C(C=C3)O)O)OC(=O)C4=CC(=C(C(=C4)O)O)O	Apples, Fuji, raw, with skin (NDB No., 97066)
(+) –Catechin		OC1CC2=C(O)C=C(O)C=C2OC1C3=CC=C(O)C(=C3)O	Bananas, raw (*Musa acuminata colla*) (NDB No., 9040)
Hesperetin		O=C(CC(C3=CC(O)=C(OC)C=C3)O2)C1=C2C=C(O)C=C1O	Juice, orange, raw (NDB No., 9206)
Naringenin		OC1=CC=C(C=C1)C2CC(=O)C3=C(O2)C=C(O)C=C3O	Melons, honeydew, raw (*Cucumis melo*) (NDB No., 9184)
Apigenin		O=C(C=C(C3=CC=C(O)C=C3)O2)C1=C2C=C(O)C=C1O	Pineapple, raw, all varieties (*Ananas comosus*) (NDB No., 9266)
Luteolin		O=C(C=C(C3=CC(O)=C(O)C=C3)O2)C1=C2C=C(O)C=C1O	Pomegranates, raw (*Punica granatum*) (NDB No., 9286)
Kaempferol		O=C(C(O)=C(C3=CC=C(O)C=C3)O2)C1=C2C=C(O)C=C1O	Broccoli, cooked, boiled, drained, without salt (NDB No., 11091)
Quercetin		O=C(C(O)=C(C3=CC(O)=C(O)C=C3)O2)C1=C2C=C(O)C=C1O	Mushrooms, white, raw (*Agaricus bisporus*) (NDB No., 11260)
Myricetin		O=C(C(O)=C(C3=CC(O)=C(O)C(O)=C3)O2)C1=C2C=C(O)C=C1O	Potatoes, red, flesh and skin, raw (*Solanum tuberosum*) (NDB No., 11355)

* Bhagwat S, Haytowitz DB, Holden JM (2012)

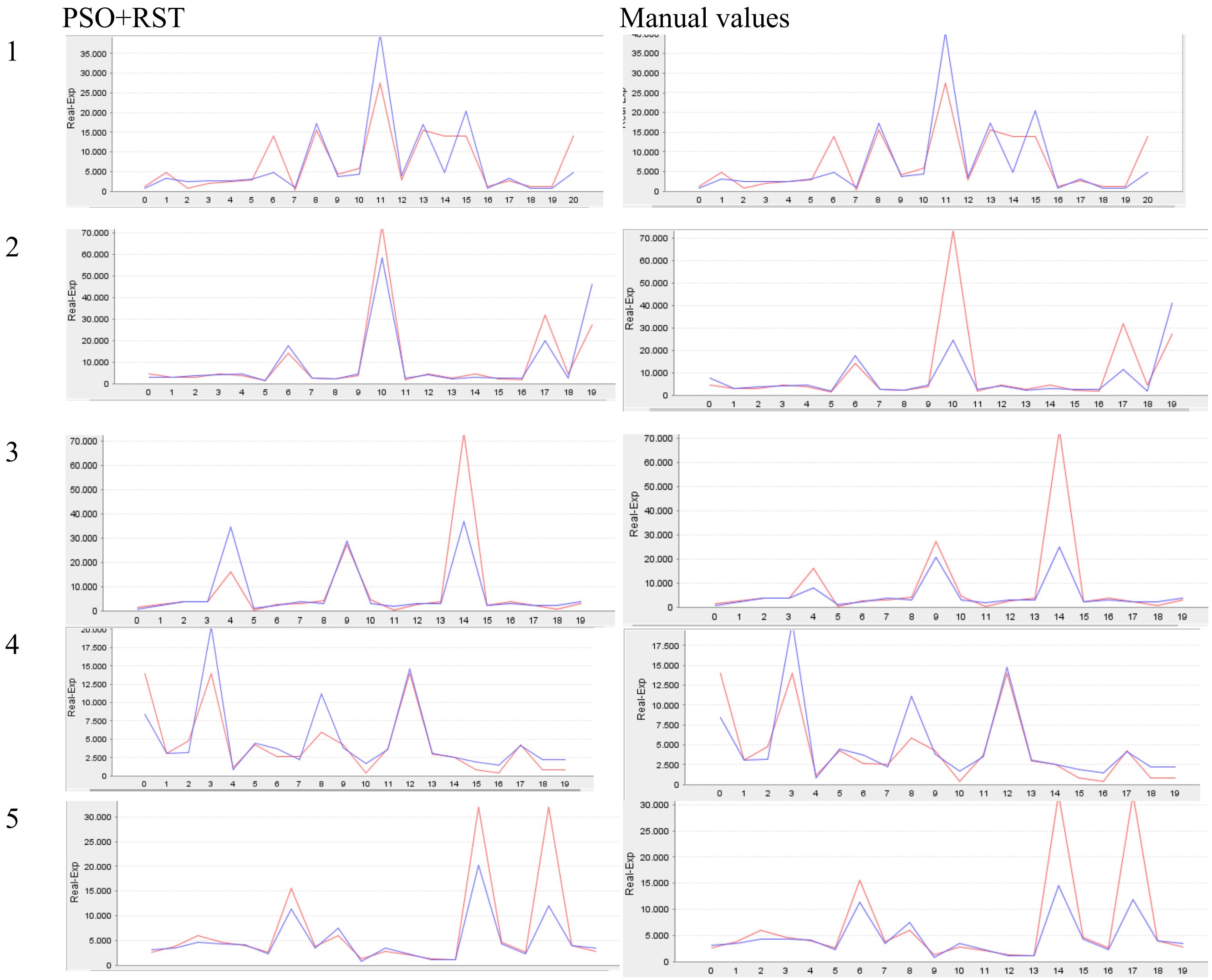


Figure 2. Prediction done using KNN algorithms: based in PSO+RST and manual feature weight calculation. (-) ORAC experimental, (−) ORAC predicted.

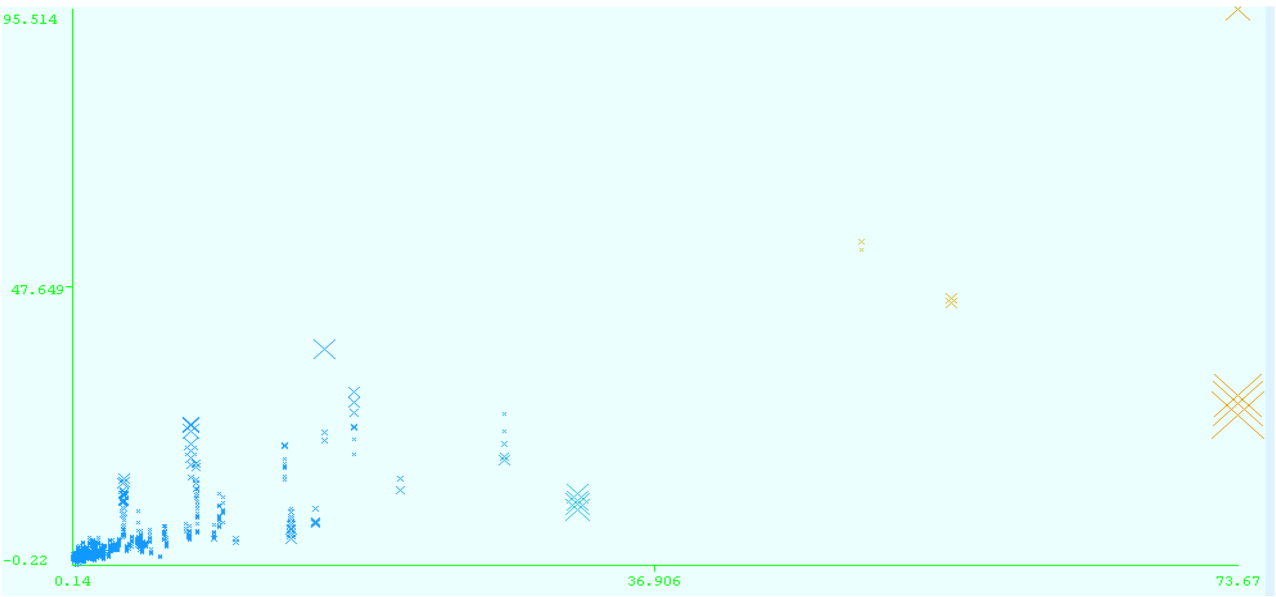


Figure 3. Prediction using MLP.

MAE (1.7601) and RMSE (4.1569)

3. Materials and Methods

1. Conformation of the data related to the food composition

The information was obtained in different food composition database: a) database for the flavonoid content of selected foods, Release 3.1 (FDB 3.1) and b) isoflavones database released by the USDA in 2008 (IDB 2) [2,3]. Therefore, it was used the estimation techniques for calculating unavailable values, and decision making procedure described by Bhagwat S *et al* [15]. This information was used to prepare the register of the data related to the composition of flavonoids in different foods. The Standard Reference (SR) [5] was used to identify each unique food entry if it matches a food in SR.

2. Prediction using AI algorithms

Training set and test set. To obtain the training set and test set it was used k-fold cross validation method of k10 iterations [4].

Attribute selections and weight assignation. To the attributes, different weights were assigned taking into account their influence in the attribute class:

i) Trolox equivalent antioxidant capacity flavonid value ($TEAC_{exp}$),
ii) Class of flavonoid,
iii) Flavonoids,
iv) Amount of flavonoid (mean),
v) Total polyphenol value (TP_{exp}),
vi) Probability of clastogenicity and clastogenicity classification by Quantitative Structure-Activity Relationship (QSAR) method.

These experimental parameters were taken from the scientific literature. The variable selected (attribute class) to predict was the $ORAC_{exp}$ value, expressed in µmolTE/100 g. ORAC was selected because it is considered to be the preferable methodology to evaluate the antioxidant capacity due to the biological relevance to the *in vivo* antioxidant efficacy [25]. The assay has been used to measure the antioxidant activity of foods and measures the degree of inhibition of peroxy-radical-induced oxidation by the compounds of interest in a chemical milieu.

$ORAC_{exp}$ and TP_{exp} (mgGAE/100 g) for each substrate were found in the literature. The analytical method developed by Prior *et al* was used as the reference method for select published sources [26].

A different weight was assigned to each attribute using the measure of the quality of a similarity decision system. Weights were assigned manually and using the Particle Swarm Optimization+Rougt Set Theory method (PSO+RST) [21,22,27]. PSO+RST was implemented in PROCONS software.

AI algorithms. KNN (*K-Nearest Neighbors*) and supervised unidirectional networks MLP, MultiLayer Perceptron algorithms, were used. These algorithms were implemented in the PROCONS software version 4.0 [27] and WEKA version 3.5.7, respectively.

i) *KNN, K-Nearest Neighbors*

KNN method is based in the paradigm that similar entrance values have the same similar exits. They are calculated based on the nearest neighbors.

Using KNN, if the neighbors $k\{e1,...,ek\}$ have values $\{v1,..,vk\}$ for the studied variable, tan the value e' is:

a) if all are similar weights:

$$v = \frac{\sum_{i=1}^{k} v_i}{k} \qquad \text{(I)}$$

b) if they are different weights:

$$v = \frac{\sum_{i=1}^{k} w_i v_i}{\sum_{i=1}^{k} w_i} \qquad \text{(II)}$$

ii) MLP, MultiLayer Perceptron

Units called neurons compose a neuronal network. Each neuron receives a series of entrances related to interconnexions and emits an exit. Furthermore the weights and connexions, each neuron was associated a transference mathematic function. This function generates the exit signal of the neuron based on the entrance signals.

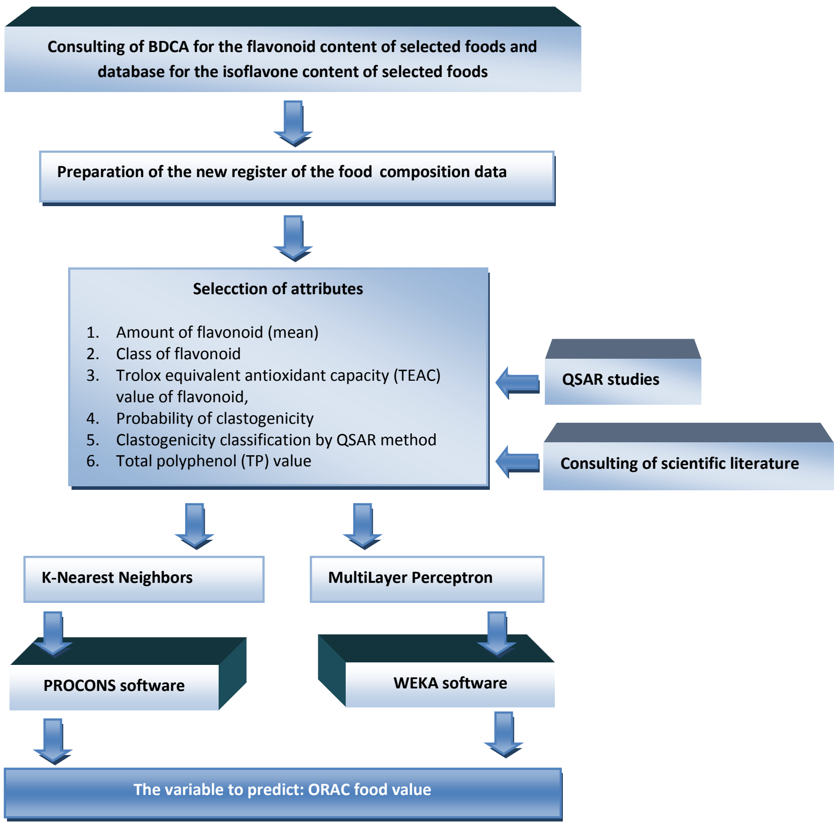
Evaluation of the precision of the algorithms. To evaluate the precision of the results obtained for both methods, they were used [28]: (a) Mean Absolute Error, MAE and (b) Root Mean Square Error, RMSE, (III)-(IV):

$$MAE = \frac{\sum_{i=1}^{N} \left| \frac{ai - yi}{ai} \right|}{N} * 100\ \% \qquad \text{(III)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left| \frac{ai - yi}{ai} \right|^2}{N}} * 100\ \% \qquad \text{(IV)}$$

Where: *ai* is the desirable exit value; *yi* is the value produced by the method and *N* is the numbers of objects.

In **Figure 5** it is shown a general scheme of the methodology used in this study.


Consulting of BDCA for the flavonoid content of selected foods and database for the isoflavone content of selected foods

Preparation of the new register of the food composition data

Selecction of attributes

1. Amount of flavonoid (mean)
2. Class of flavonoid
3. Trolox equivalent antioxidant capacity (TEAC) value of flavonoid,
4. Probability of clastogenicity
5. Clastogenicity classification by QSAR method
6. Total polyphenol (TP) value

Figure 5. Scheme of the applied methodology.

4. Conclusions

The best results were obtained when the calculation of weight and similarity were included in the algorithms. Using KNN, the optimum k value was 3, making evident the importance of the *similarity* between objects for the good predictive results. The results obtained in the predictions show that the weights assigned to the attributes, taking into account their influence in the attribute class (ORACexp) were right.

It was concluded the importance of the use of KNN and MLP technic for the prediction of the antioxidant activity en different alimentary groups. These algorithms can be used, in future work, to identify the responsible features for the relationship between quantity of flavonoids, topologic-structural information and alimentary matrix. It will be further studied the relationship between antioxidant capacity of the food and the composition in flavonoids of a complex alimentary matrix.

Acknowledgments

The authors thank the partial financial support of University of Santiago de Compostela, University of Camagüey Ignacio Agramonte Loynaz and Galician Plan of research, innovation and growth 2011-2015 (Plan I2C). The authors also thank Y. Filiberto *et al.* for the access to the PROCONS software.

Author Contributions

All the authors contributed equally for the execution of the work and writing of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Holdena, J.M.; Bhagwata, S.A.; Haytowitza, D.B.; Gebhardta, S.E.; Dwyerb, J.T.; Petersonb, J.; Beechera, G.R.; Eldridgec, A.L.; Balentined, D. Development of a database of critically evaluated flavonoids data: Application of usda's data quality evaluation system. *Journal of Food Composition and Analysis* **2005**, *18,* 829–844.
2. Bhagwat, S.; Haytowitz, D.B.; Holden, J.M. Usda database for the flavonoid content of selected foods, release 3.1. NDL Web site: http://www.ars.usda.gov/Services/docs.htm?docid=6231, 2012.
3. U.S. Department of Agriculture, A.R.S. Usda database for the isoflavone content of selected foods. Release 2.0. http://www.ars.usda.gov/Services/docs.htm?docid=6382, 2008
4. Bhagwat, S.; Haytowitz, D.B.; Wasswa-Kintu, S.I.; Holden, J.M. USDA develops a database for flavonoids to assess dietary intakes. *Procedia Food Science* **2013**, *2,* 81-86.
5. U.S. Department of agriculture, A.R.S. Usda national nutrient database for standard reference. Available online: http://www.ars.usda.gov/nutrientdata. (April 22/2015).
6. Harnly, J.M.; Doherty, R.F.; Beecher, G.R.; Holden, J.M.; Haytowitz, D.B.; Bhagwar, S.; Gebhardt, S. Flavonoid content of u.S. Fruits, vegetables, and nuts. *J. Agric. Food Chem.* **2006**, *54*, 9966-9977.
7. Kay, C.D. The future of flavonoid research. *Brit. J. Nutr* **2010**, *104*, S91-S95.
8. Robards, K.; Antolovich, M. Analytical chemistry of fruit bioflavonoids.A review. *Analyst* **1997**, *122*, 11R-34R.

9. Halliwell, B. Oxidative stress, nutrition and health. Experimental strategies for optimization of nutritional antioxidant intake in humans. *Free Rad. Res.* **1996**, *25*, 57-74.
10. Chun, O.; Floegel, A.; Chung, S.; Chung, C.; Song, W.; Koo, S. Estimation of antioxidant intakes from diet and supplements in us adults. *J Nutr* **2010**, *140*, 317-324.
11. Schroeter, H.; Heiss, C.; Spencer, J.P.E.; Lupton, J.R.; Schmitz, H.H.; Keen, C.L. Recommending flavanols and procyanidins for cardiovascular health: Current knowledge and future needs. *Molecular Aspects of Medicine* **2010**, *31*, 546-557.
12. Institute of Medicine of the Nation al Academies. *Dietary reference intakes for vitamin c, vitamin e, selenium, and carotenoids*. National Academy Press: Washington, D.C., 2000.
13. FAO. Retos sobre la composicion de alimentos. Available online: http//www.fao.org/infoods/infoods/retos (May 13/2015)
14. Haytowitz, D.B.; Bhagwat, S.; Holden, J.M. Sources of variability in the flavonoid content of foods. *Procedia Food Science* **2013**, *2*, 46-51.
15. Bhagwat, S.A.; Haytowitz, D.B.; Wasswa-Kintu, S.; Pehrsson, P.R. **2015**. Development of USDA's expanded flavonoid database: A Tool for Epidemiological Research. British Journal of Nutrition. DOI: 10.1017/S0007114515001580.
16. Trujillano, J.; March, J.; Sorribas, A. Aproximación metodológica al uso de redes neuronales artificiales para la predicción de resultados en medicina. *Medicina clinica* **2004**, *122*, 22-26.
17. Estrada, E.; Molina, E. Novel local (fragment-based) topological molecular descriptors for qspr/qsar and molecular design. *Journal of Molecular Graphics and Modelling* **2001**, *20*, 54-64.
18. Miller, N.J. The relative antioxidant activities of plant-derived polyphenolic flavonoids. In *Natural antioxidants and food quality in atherosclerosis and cancer prevention*, Kumpulainen, J.T.; Salonen, J.T., Eds. The Royal Society of Chemistry: Cambridge, UK, 1996; pp 256-259.
19. Yordi, E.G.; Molina, E.; Matos, M.J.; Uriarte, E. Structural alerts for predicting clastogenic activity of pro-oxidant flavonoid compounds: Quantitative structure-activity relationship study. *J Biomol Screen* **2012**, *17*, 216-224.
20. Wu, X.; Beecher, G.R.; Holden J. M.; Haytowitz, D.B.; Gebhardt, S.E.; and Prior, R.L. Lipophilic and hydrophilic antioxidant capacities of common foods in the united states. *J. Agric. Food Chem.* **2004**, *52*, 4026-4037.
21. Kevers, C.; Falkowski, M.; Tabart, J.; Defraigne, J.-O.; Dommes, J.; Pincemail, J. Evolution of antioxidant capacity during storage of selected fruits and vegetables. *J. Agric. Food Chem.* **2007**, *55*, 8596-8603.
22. Ou, B.; Huang, D.; Hampsch-Woodill, M.; Flanagan, J.A.; Deemer, E.K. Analysis of antioxidant activities of common vegetables employing oxygen radical absorbance capacity (orac) and ferric reducing antioxidant power (frap) assays: A comparative study. *J. Agric. Food Chem.* **2002**, *50*, 3122-3128.
23. Roy, M.K.; Juneja, L.R.; Isobe, S.; Tsushida, T. Steam processed broccoli (brassica oleracea) has higher antioxidant activity in chemical and cellular assay systems. *Food Chem* **2009**, *114*, 263-269.
24. Thaipong, K.; Boonprakob, U.; Crosby, K.; Cisneros-Zevallos, L.; Byrne, D.H. Comparison of abts, dpph, frap, and orac assays for estimating antioxidant activity from guava fruit extracts. *J. Food Comp. Anal* **2006**, *19*, 669-675.

25. Awika, J.M.; Rooney, L.W.; Wu, X.; Prior, R.L.; Cisneros-Zevallos, L. Screening methods to measure antioxidant activity of sorghum (sorghum bicolor) and sorghum products. *J. Agric. Food Chem.* **2003**, *51*, 6657-6662.
26. Prior, R.L.; Hoang, H.; Gu, L.; Wu, X.; Bacchocca, M.; Howard, L.; Hampsch Woodill, M.; Huang, D.; Ou, B.; Jacob, R. Assays for hydrophilic and lipophilic antioxidant capacity (oxygen radical absorbance capacity (orac) of plasma and other biological and food samples. *J. Agric. Food Chem.* **2003**, *51*, 3273-3279.
27. Filiberto, Y.; Bello, R.; Caballero, Y.; Frias, M. In *A method to build similarity relations into extended rough set theory*, 10 th International Conference on Intelligent Systems and Applications ISDA 2010, Cairo, Egypt., 2010; IEEE Catalog Number CFP 10384 CDR: Cairo, Egypt.
28. Filiberto, Y.; Bello, R.; Caballero, Y.; Larrua, R. Una medida de la teoría de los conjuntos aproximados para sistemas de decisión con rasgosde dominio continuo. *Rev. Fac. Ing. Univ. Antioquia* **2011**, *60* 141-152.

MOL2NET

SciForum
Mol2Net

A Proposal Tool for Manipulation of a Set of Protein Structures from PDB

Vinicius R. Seus [1],*, Adriano V. Werhli [2]and Karina S. Machado [2]

[1] Universidade Federal do Rio Grande—FURG, Campus Carreiros: Av. Itália km 8 Bairro Carreiros

[2] Universidade Federal do Rio Grande—FURG, Campus Carreiros: Av. Itália km 8 Bairro Carreiros; E-Mails: werhli@furg.br (A.V.W.); karina.machado@furg.br (K.S.M.)

* Author to whom correspondence should be addressed; E-Mail: viniciusseus@furg.br;
Tel.: +(55)-(53)-32-33-65-00.

Published: 4 December 2015

Abstract: Protein Data Bank (PDB) is a public web database with more than 100,000 biological macromolecular structures. With this large amount of protein structures available on PDB the use of tools for acquisition and analysis of specific sets of biological macromolecules is a necessity. Hence, in this work we propose the development of a tool for acquiring, storing and analyzing specific sets of proteins from the PDB database. The proposed tool runs on desktop environment allowing the user to acquire the structures from the RESTful web-service provided by PDB server. After the acquisition of a set of interesting PDBs the user can manipulate these data in an off-line environment through a local database that stores the information about the characteristics of the structures, for example, ligands, mutations, residues, sequences and docking results. The protein files are locally stored in the users' computer and can be used, for instance, for molecular docking simulations and alignment of sequences and structures. Having a set of proteins of interest available locally and using our proposed tool the user can perform analysis related to alignments and visualize important proteins characteristics improving the knowledge about specific target. Besides, the user can select PDB files to be visualized on a graphical environment that is integrated in our tool. Other features are related to the exporting of sequence alignments results in csv (comma separated value) format or exporting sequences that have a similar identity in a format that can be easily loaded on graph tools. These alignments allow the user to visualize which proteins are similar and discard those that are not.

Keywords: Protein Data Bank; Sequence alignment; Structural alignment

1. Introduction

With the large growth of protein data stored in the databases available on the web, appears the necessity to create different computer applications that help researchers in knowledge discovery. Protein Data Bank (PDB) [1] is a public web database containing over 100.000 biological macromolecular structures. So, with this large amount of protein structures available in PDB and other global servers that store various information of macromolecular structures, it becomes clear that the use of tools for the acquisition and analysis of specific sets of biological macromolecules is a need to facilitate the search for a specific target protein. In this work, we proposed the development of a tool for acquisition, storage and analysis of specific sets of proteins from PDB in an environment that integrates different functionalities. The proposed tool is executed on the desktop environment allowing the user to perform acquisition of the molecular structures from RESTful web-service provided by the PDB own server.

2. Results and Discussion

In this section we present the proposed tool showing all modules and functionalities. Figure 1 shows all modules of our proposed tool. In the following each module is discussed:

- Proposed tool – This module is the application itself, where the local database and the interface are located in this scheme;
- PDB module - This module represents the public web database PDB (Protein Data Bank). The proposed tool connects with this database for acquiring molecular three-dimensional structures through the web-service RESTfull provided by the PDB. After the data acquisition, the user relates these data to a unique project that can be edited or deleted later;
- Sequence alignment module – This module performs sequence alignments. Having this feature, the user can align the sequences of all structure proteins of one project. As a result we have a matrix nxn where n is the number of proteins of a project and each cell is the identity between two sequence of proteins. Thus, it is possible to visualize proteins that have higher sequence similarity;
- Protein visualization module – This module opens an external protein visualization tool called PyMol[3]. Having this module the user can visualize the proteins of a project in a three-dimensional environment;
- Ligand visualization module – This module presents a list with all ligands presented in a protein structure discretized by chain linked to an user project;
- Structural alignment module - This module is for alignment of the tertiary structures. With respect to this operation, the user selects one of the proteins of his project to be the reference structure for all other structures of one project. Then, an algorithm is performed to align the tertiary structure of all proteins of a project with the reference structure. This functionality is important for Virtual Screening (VS) process, because to consider a set of receptor proteins with an equal grid box for docking all the structures need to have the same cartesian coordinate system.

The proposed tool was developed for the desktop environment. Using its interface, the user starts searching for a target receptor through a keyword in the search field. Next, a set of possible protein structures related to the specific target is listed in the tool showing all their PDB id's. Thus, the user can perform the acquisition of these proteins structures. This acquisition process is possibl using the web-service Restful provided by PDB itself together with the local database provided by our tool. The purpose of this database is to provide access to features of the molecular structures of a specific project in an offline environment. This database is populated after each time the user performs the acquisition of a set of specific data from the proposed tool. Then, with the inclusion of molecular structures in the local database completed, the user can use all the modules presented in the figure 1.

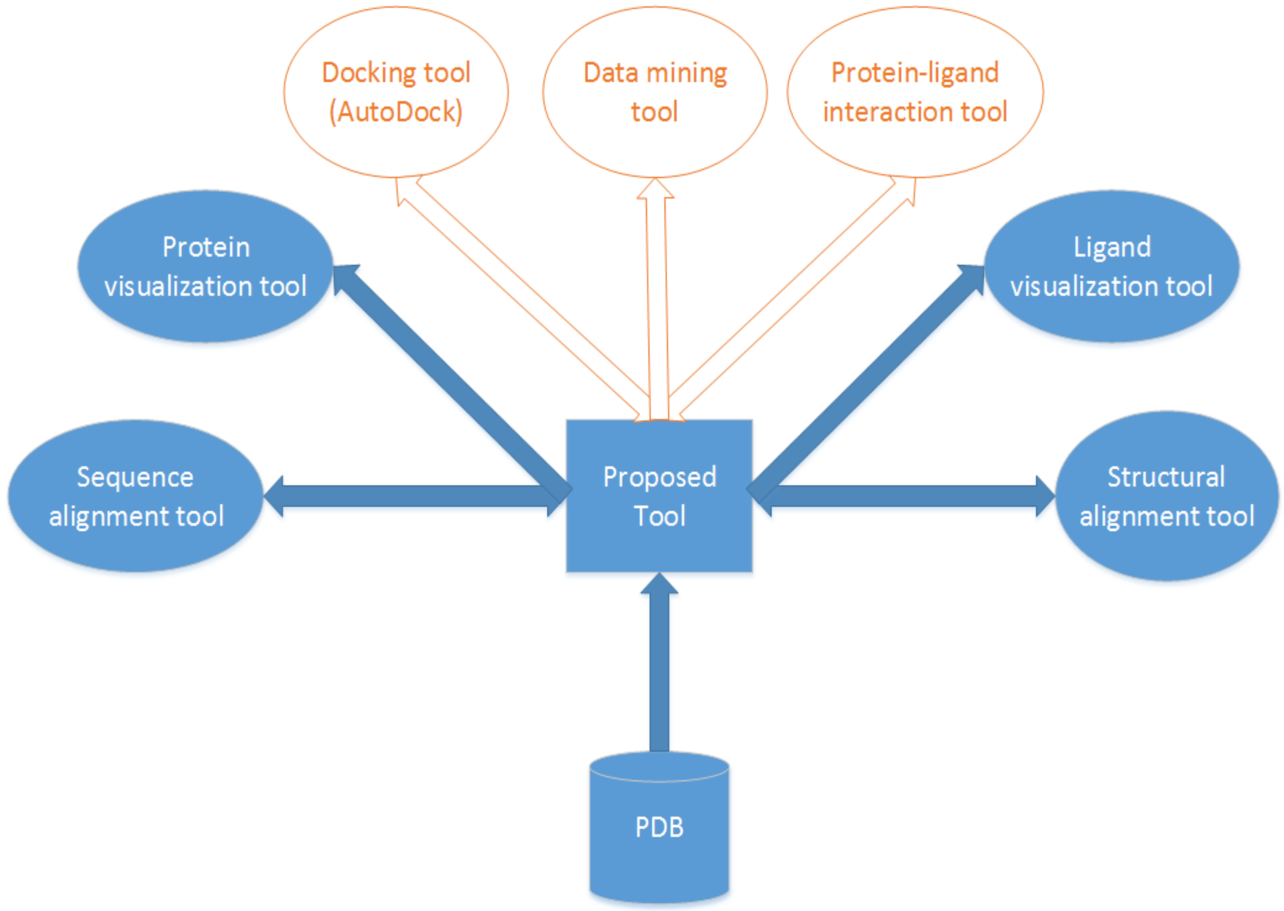

Docking tool (AutoDock)
Data mining tool
Protein-ligand interaction tool
Protein visualization tool
Ligand visualization tool
Sequence alignment tool
Proposed Tool
Structural alignment tool
PDB

Figure 1. System's modules.

3. Materials and Methods

This section presents the materials and methods applied in the development of the proposed tool.

Our proposed tool was developed using Python language programming. Python [5] is a high level interpreted and interactive language that provides to the developers the use of a strong and dynamic typing. Furthermore, the language provides an easy syntax to understand, turning the programming faster and more productive. Besides, Python presents another relevant feature that is the use of the virtual machine bytecode, what makes the code portable. This means that the program can be compiled in one platform to be executed on other platforms.

The local database of the proposed tool was implemented using MySQL. MySQL [4] is an open-source database management system that is the most popular in the world. Its uses SQL (Structured Query Language) as interface allowing an easy handling, excellent performance and stability.

For development of the tool it was used the IDE (Integrated Development Environment) PyCharm [6] and to develop the local database the Navicat [7] software was employed.

The Alignment sequence and alignment structural tools were developeded using Biopython libraries.

According to Cock et al [2], Biopython is a mature and open source tool that provides libraries in Python that help in a wide range of problems commonly found in Bioinformatics. Moreover, Biopython has modules for reading and writing files with different formats and multiple sequence alignments. It also has modules that deal with tertiary macromolecular structures, perform access to the PDB database available on the web and also provide numerical methods for statistical learning. Since its founding in 1999, Biopython has grown to currently having a large collection of modules. This tool is intended for developers of computational biology that need to incorporate in their scripts or their own software modules that help in their specific problems.

For the development of molecular visualization module the Pymol library was used.

According to Schrödinger [3] Pymol is a molecular visualization system that provides high quality three-dimensional images of small and larger macromolecules such as proteins. Pymol is one of the few viewing open source tools available for use in computational biology [3].

4. Conclusions

This paper presents a tool developed to allow the investigation of a set of different macromolecular structures related to a specific target.

Having our proposed tool, the researcher can perform a number of manipulations in a set of protein structures, working in a unique environment with modules for performing the alignment of sequences and structures, for visualization of these structures, to list the associate ligands and so on.

Acknowledgments

K. Machado acknowledges the support from CNPq (Process number 477462/2013-8). Vinicius Rosa Seus receives a CAPES Master Degree fellowship.

Conflicts of Interest

“The authors declare no conflict of interest”.

References and Notes

1. Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov I., Bourne, P.. The protein Data Bank. *Nucleic Acids Research* **2000**, *28*, 235–242.
2. Cock, P. J. A., Antao, T., Chang, J. T., Chapman, A. B., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, T., Wilczynski, B., de Hoon M. J.. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **2009**, *11*, 1422–1423.
3. Schrödinger, L.. The PyMOL Molecular Graphics System, Version 1.7.4, LLC **2010**.
4. MySQL - The world's most popular open source database. Available online: https://www.mysql.com (accessed on 14 October 2015).
5. Python. Available online: https://www.python.org/ (accessed on 14 October 2015).
6. PyCharm – The Most intelligent Python IDE. https://www.jetbrains.com/pycharm (accessed on 15 October 2015).
7. Navicat. http://www.navicat.com (accessed on 15 October 2015).

SciForum
Mol2Net

Multi-Viral Targets Entropy QSAR for Antiviral Drugs

Francisco J. Prado-Prado [1,*] and Xerardo García-Mera [2]

1 Biomedical Sciences Department, Health Sciences Division, University of Quintana Roo, 77019 Chetumal, Mexico

2 Department of Organic Chemistry, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain

Published: 4 December 2015

Abstract: The antiviral QSAR models today have an important limitation. Only they predict the biological activity of drugs against only one viral species. This is determined due the fact that most of the current reported molecular descriptors encode only information about the molecular structure. As a result, predicting the probability with which a drug is active against different viral species with a single unifying model is a goal of major importance. In this we use the Markov Chain theory to calculate new multi-target entropy to fit a QSAR model that predict by the first time a ms-QSAR model for 900 drugs tested in the literature against 40 viral species and other 207 drugs no tested in the literature using entropy QSAR. We used Linear Discriminant Analysis (LDA) to classify drugs into two classes as active or non-active against the different tested viral species whose data we processed. The model correctly classifies 31 188 out of 31 213 non-active compounds (99.92%) and 432 out of 434 active compounds (99.54%). Overall training predictability was 98.56%. Validation of the model was carried out by means of external predicting series, the model classifying, thus, 15 588 out of 15 606 non-active compounds and 213 out of 217 active compounds. Overall validation predictability was 98.54%. The present work report the first attempts to calculate within a unify framework probabilities of antiviral drugs against different virus species based on entropy analysis.

Keywords: Antiviral drugs; QSAR; Entropy; Mutli-tasking Learning; Markov Chain model; Linear Discriminant Analysis

1. Introduction

Examples of diseases caused by viruses include the common cold (produced by any one of a variety of related viruses), AIDS (caused by HIV) and cold sores (caused by herpes simplex); which produced some of the major health problems in the last 30 years. Other relationships are being studied such as the connection of Human Herpesvirus 6 (HHV6), one of the eight known members of the human herpes virus family, with organic neurological diseases such as multiple sclerosis and chronic fatigue syndrome. Recently, it has been shown that cervical cancer is caused, at least partially, by papillomavirus, representing the first significant evidence in humans for a link between cancer and an infective agent. The relative ability of viruses to cause disease is described in terms of virulence.

Consequently, there is an increasing interest on the development of rational approaches for discovery of antifungal drugs. In this sense, a very important role may be played by computer-added drug discovery techniques based on Quantitative-Structure-Activity-Relationship (QSAR) models (1). Unfortunately, almost QSAR studies, including those for antiviral activity and others, use limited databases of structurally parent compounds acting against one single fungus species (2). One important step in the evolution of this field was the introduction of QSAR models for heterogeneous series of antimicrobial compounds; see for instance the works of Cronin, de Julián-Ortiz, Galvéz, Gárcía-Domenech, Gosalbez, Marrero-Ponce, Torrens, *et al.* and others (3-15). As a result, researchers may predict very heterogeneous series of compounds but often need to use/develop as many QSAR equations as microbial species are necessary to be predicted. In any case, if you aim to predict activity against different targets you still need to use one different QSAR model for each target.

An interesting alternative, is the prediction of structurally diverse series of antimicrobial compounds (antiviral in this case) against different targets (mechanisms) using complicated non-linear Artificial Neural Networks with multi-class prediction, *e.g.* the work of Vilar *et al.* (16). We can understand strategies developed in this sense as Multi-Objective Optimization (MOOP) techniques; in this case we pretend to optimize the activity of antiviral drugs against many different objectives or targets (viral species). A very useful strategy related to the MOOP problem use Derringer's desirability function desirability function and many QSAR models for different objectives (17). In this sense, it is of major importance the development of unified but simple linear equations explaining the antimicrobial activity, in the present work antiviral activity, of structurally-heterogeneous series of compounds active against as many targets (viral species) as possible. We call this class of QSAR problem the multi-target QSAR (mt-QSAR) (18, 19).

There are near to 2000 chemical molecular descriptors that may be in principle generalized and used to solve the mt-QSAR problem. Many of these indices are known as Topological Indices (TIs) or simply invariants of a molecular graph G. We can rationalize G as a draw composed of vertices (atoms) weighted with physicochemical properties (mass, polarity, electro negativity, or charge) and edges (chemical bonds) (20). In any case, many of these indices have not been extended yet to encode additional information to chemical structure. One alternative to mt-QSAR is the substitution of classic atomic weights by target specific weights. For instance, we introduced and/or reviewed TIs that use atomic weights for the propensity of the atom to interact with different microbial targets (21) or undergoes partition in a biphasic systems or distribution to biological tissues (22-24). The method, called MARCH-INSIDE approach, Markovian Chemicals In Silico Design, calculates TIs using Markov Chain theory. In fact, MARCH-INSIDE define a Markov matrix to derive matrix invariants such as stochastic spectral moments, mean values, absolute probabilities, or entropy measures, for the study of molecular properties. Applications to macromolecules have extended to RNA, proteins, and blood proteome (25-30). In particular, one of the classes of MARCH-INSIDE descriptors is defined in terms of entropy measures; which have demonstrated flexibility in many bioorganic and medicinal chemistry problems such as: estimation of anticoccidial activity, modelling the interaction between drugs and HIV-packaging-region RNA, and predicting proteins and virus activity (24, 31-33). We give high importance to entropy measures due to it have been largely demonstrate as an excellent function to codify information in molecular systems, see for instance the important works of Graham (34-39). However, have not been studied the proficiency of entropy indices (of MARCH-INSIDE type or not) to solve the mt-QSAR problems in antiviral compounds.

The present study develops the first mt-QSAR model based on entropy indices to predict antiviral activity of drugs against different viral species. The model fits one of the largest datasets used up-to-date in QSAR studies, number of entries 47 000+ cases; which is the result of forming different (antiviral compounds/viral target) pairs.

2. Results and Discussion

One of the main advantages of the present stochastic approach is the possibility of deriving average thermodynamic parameters depending on the probability of the states of the MM. The generalized parameters fit on more clearly physicochemical sense with respect to our previous ones (24, 41, 42). In specific, this work introduces by the first time a linear mt-QSAR equation model useful for prediction and MOOP of the antiviral activity of drugs against different viral target species or objectives. The best model found was:

$$actv = 0.38\cdot\theta_3(s)_{het} - 0.84\cdot\theta_0(s)_{total} - 0.91\cdot\theta_0(s)_{C_{sat}} + 0.89\cdot\theta_1(s)_{C_{sat}} + 2.01\cdot\theta_0(s)_{Csp\&sp2} - 0.32\cdot\theta_5(s)_{h-het} - 4.71 \qquad (8)$$
$$N = 31190 \qquad \lambda = 0.38 \qquad \chi^2 = 377.43 \qquad p < 0.001$$

In the model the coefficient λ is the Wilk's statistics, statistic for the overall discrimination, χ^2 is the Chi-square, and p the error level. In this equation, ${}^k\theta_s$ where calculated for the totality (T) of the atoms in the molecule or for specific collections of atoms. These collections are atoms with a common characteristic as for instance are: heteroatom (*Het*), unsaturated Carbon atoms (C_{unst}), saturated Carbon atoms (C_{sat}) and hydrogen bound to heteroatom (*H-Het*. The model correctly classifies 31 188 out of 31 213 non-active compounds (99.92%) and 432 out of 434 active compounds (99.54%). Overall training predictability was 98.56%. Validation of the model was carried out by means of external predicting series, the model classifying, thus, 15 588 out of 15 606 non-active compounds and 213 out of 217 active compounds. Overall validation predictability was 98.54%.

The more interesting fact is that ${}^k\theta_s$ have the skill of discerning the active/no-active classification of compounds among a large number of viral species. This property is related to the definition of the ${}^k\theta_s$ using species-specific atomic weights (see supplementary material file for method). It allows us to model by the first time a very heterogeneous a diverse data with more than 47 470 cases (one of the largest in QSAR). Another interesting characteristic of the model is that the ${}^k\theta_s$ used as molecular descriptors depend both on the molecular structure of the drug and the viral species against which the drug must act. The codification of the molecular structure is basically due to the use of the adjacent factor α_{ij} to encode atom-atom bonding, molecular connectivity. The other aspect that allows encoding molecular structural changes is that the entropy ${}^k\theta_s$ are atom-class specific. This property is related to the definition of the ${}^k\theta_s$. For example, one change in the molecular structure of, e.g. S by O, necessarily implies a change in the moments of interaction. Moreover, the most interesting fact is that ${}^k\mu_s$ are the molecular descriptors reported for antimicrobial mt-QSAR studies able to distinguish among a large number of viral species. The present work is the first reported mt-QSAR model using entropy $^k\theta_s$ as a molecular descriptor that allow one predicting antiviral activity of any organic compound against a very large diversity of viral pathogens.

3. Materials and Methods

3.1. Markov entropy (θ_k) for drug-target k-th step-by-step interaction

One can consider a hypothetical situation in which a drug molecule is free in the space at an arbitrary initial time (t_0). It is then interesting to develop a simple stochastic model for a step-by-step interaction between the atoms of a drug molecule and a molecular receptor in the time of desencadenation of the pharmacological effect. For the sake of simplicity, we are going to consider from now on a general structure less receptor. Understanding as structure-less molecular receptor a model of receptor which chemical structure and position it is not taken into consideration. Specifically, the molecular descriptors used in the present work are called stochastic entropies θ_k, which are entropies describing th connectivity and the distribution of electrons for each atom in the molecule (40). The initial entropy of interaction a j-th atom of the drug with the target $^0\theta_j(s)$ is considered as a state function so a reversible process of interaction may be came apart on several elemental interactions between the j-th atom and the receptor. The 0 indicates that we refer to the initial interaction, and the argument (s) indicates that this energy depends on the specific viral species. Afterwards, interaction continues and we have to define the interaction probability $^k\theta_{ij}(s)$ between the j-th atom and the receptor for specific viral specie (s) given that i-th atom has been interacted at previous time t_k. In particular, immediately after of the first interaction (t_0 = 0) takes place an interaction $^1p_{ij}(s)$ at time t_1 = 1 and so on. So, one can suppose that, atoms begin its interaction whit the structure-less molecular receptor binding to this receptor in discrete intervals of time t_k. However, there several alternative ways in which such step-by-step binding process may occur (24, 41, 42).

The entropy $^0\theta_j(s)$ will be considered here as a function of the absolute temperature of the system and the equilibrium local constant of interaction between the j-th atom and the receptor $^0\gamma_j(s)$ for a give microbial species. Additionally, the energy $^1\theta_{ij}(s)$ can be defined by analogy as $\gamma_{ij}(s)$ (24, 41, 43):

$$^0\theta_j(s) = -R\cdot T\cdot \log{}^0\Gamma_j(s) \quad (1) \qquad {}^1\theta_{ij}(s) = -R\cdot T\cdot \log{}^1\Gamma_{ij}(s) \quad (1)$$

The present approach to antimicrobial-species-specific-drug-receptor interaction has two main drawbacks. The first is the difficulty on the definition of the constants. In this work, we solve the first question estimating $^0\gamma_j(s)$ as the rate of occurrence $n_j(s)$ of the j-th atom on active molecules against a given specie with respect to the number of atoms of the j-th class in the molecules tested against the same specie $n_t(s)$. With respect to $^1\gamma_{ij}(s)$ we must taking into consideration that once the j-th atom have interacted the preferred candidates for the next interaction are such i-th atoms bound to j by a chemical bond. Both constants can be then written down as (24, 41, 43):

$$^0\Gamma_j(s) = \left(\frac{\mathrm{n}_j(s)}{\mathrm{n}_\mathrm{T}(s)} + 1\right) = e^{\frac{^0\theta_j(s)}{R\cdot T}} \quad (2) \qquad {}^1\Gamma_{ij}(s) = \left(\alpha_{ij}\cdot\frac{\mathrm{n}_j(s)}{\mathrm{n}_\mathrm{T}(s)} + 1\right) = e^{\frac{^1\theta_{ij}(s)}{R\cdot T}} \quad (3)$$

Where, α_{ij} are the elements of the atom adjacency matrix, $n_j(s)$, $n_t(s)$, $^0\theta_j(s)$, and $^1\theta_{ij}(s)$ have been defined in the paragraph above, r is the universal gases constant, and t the absolute temperature. The number 1 is added to avoid scale and logarithmic function´s definition problems. The second problem relates to the description of the interaction process at higher times $t_k > t_l$. Therefore, mm theory enables a simple calculation of the probabilities with which the drug-receptor interaction takes place in the time until the studied effect is achieved. In this work we are going to focus on drugs-microbial structure less target interaction. As depicted in figure 1, this model deals with the calculation of the probabilities ($^kp_{ij}$) with which any arbitrary molecular atom j-th bind to the structure less molecular receptor given that other atom i-th has been bound before; along discrete time periods t_k (k = 1, 2, 3, …); (k = 1 in grey), (k = 2 in blue) and (k = 3 in red) throughout the chemical bonding system. The procedure described here considers as states of the mm the atoms of the molecule. The method arranges all the $^0\theta_j(s)$ values in a vector **θ (s)** and all the $^1\theta_{ij}(s)$ entropies of interaction as a squared table of n x n dimension. After normalization of both the vector and the matrix we can built up the corresponding absolute initial probability vector **φ(s)** and the stochastic matrix ^{1}**Π(s)**, which has the elements $^0p_j(s)$ and $^1p_{ij}(s)$ respectively. The elements $^0p_j(s)$ of the above mentioned vector **φ(s)** constitutes the absolute probabilities with which the j-th atom interact with the molecular target or receptor in the species s at the initial time with respect to any atom in the molecule (24, 41, 43):

$$^0p_j(s) = \frac{^0\theta_j(s)}{\sum_{a=1}^{m} {^0\theta_a(s)}} = \frac{-RT \cdot \log\left(\frac{n_j(s)}{n_T(s)}+1\right)}{\sum_{a=1}^{m} -RT \cdot \log\left(\frac{n_a}{n_T(s)}+1\right)} = \frac{\log\left(\frac{n_j(s)}{n_T(s)}+1\right)}{\sum_{a=1}^{m} \log\left(\frac{n_a}{n_T(s)}+1\right)} \quad (4)$$

Where, m represents all the atoms in the molecule including the j-th, n_a is the rate of occurrence of any atom a including the j-th with value n_j. On the other hand, the matrix is called the 1-step drug-target interaction stochastic matrix. ^{1}**Π(s)** is built too as a squared table of order n, where n represents the number of atoms in the molecule. The elements $^1p_{ij}(s)$ of the 1-step drug-target interaction stochastic matrix are the binding probabilities with which a j-th atom bind to a structure less molecular receptor given that other i-th atoms have been interacted before at time $t_1 = 1$ (considering $t_0 = 0$) (18, 24, 41, 43):

$$^1p_{ij}(s) = \frac{^1\theta_{ij}(s)}{\sum_{a=1}^{n} {^1\theta_{ia}(s)}} = \frac{\alpha_{ij} \cdot (-RT) \cdot \log\left(\frac{n_j(s)}{n(s)}+1\right)}{\sum_{a=1}^{n} \alpha_{ia} \cdot (-RT) \cdot \log\left(\frac{n_a(s)}{n_T(s)}+1\right)} = \frac{\alpha_{ij} \cdot \log\left(\frac{n_j(s)}{n_T}+1\right)}{\sum_{a=1}^{n} \alpha_{ia} \cdot \log\left(\frac{n_j(s)}{n_T(s)}+1\right)} \quad (5)$$

By using, **φ(s),** ^{1}**Π(s)** and chapman-kolgomorov equations one can describe the further evolution of the system.[10-17] summing up all the atomic free energies of interaction $^0\theta_j(s)$ pre-multiplied by the absolute probabilities of drug-target interaction $^ap_k(j,s)$ one can derive the average changes in entropies $^k\theta_s$ of the gradual interaction between the drug and the receptor at a specific time k in a given microbial species (s) (24):

$$^k\theta_s = \boldsymbol{\varphi}(\mathbf{s}) \cdot {^k\mathbf{\Pi}(\mathbf{s})} \cdot {^0\theta(\mathbf{s})} = \boldsymbol{\varphi}(\mathbf{s}) \cdot \left[{^1\mathbf{\Pi}(\mathbf{s})}\right]^k \cdot \theta {^1\mathbf{\Pi}(\mathbf{s})} = \sum_{j=1}^{n} {^k\theta_j(s)} = \sum_{j=1}^{n} {^Ap_k(j,s)} \cdot {^0\theta_j(s)} \quad (6)$$

Such a model is stochastic *per se* (probabilistic step-by-step atom-receptor interaction in time) but also considers molecular connectivity (the step-by-step atom union in space throughout the chemical bonding system).

3.2. Statistical analysis

As a continuation of the previous sections, we can attempt to develop a simple linear QSAR using the MARCH-INSIDE methodology, as defined previously, with the general formula:

$$Actv = a_0 \cdot {^0\theta_s} + a_1 \cdot {^1\theta_s} + a_2 \cdot {^2\theta_s} + a_3 \cdot {^3\theta_s} ..... + a_k \cdot {^k\theta_s} + b_0 \quad (7)$$

Here, $^k\theta_s$ act as the microbial species specific molecule-target interaction descriptors. The calculation of these indices has been explained in supplementary material by space reasons. We selected Linear Discriminant Analysis (LDA) to fit the classification functions. The model deals with the classification of a set of compounds as active or not against different microbial species(43). A dummy variable (Actv) was used to codify the antimicrobial activity. This variable indicates either the presence (Actv = 1) or absence (Actv = –1) of antimicrobial activity of the drug against the specific species. In equation (1), a_k represents the coefficients of the classification function and b_0 the independent term, determined by the least square method as implemented in the LDA module of the STATISTICA 6.0 software package(44). Forward stepwise was fixed as the strategy for variable selection(43). The quality of LDA models was determined by examining Wilk's U statistic, Fisher ratio (F), and the p-level (p). We also inspected the percentage of good classification and the ratios between the cases and variables in the equation and variables to be explored in order to avoid over-fitting or chance correlation. Validation of the model was corroborated by re-substitution of cases in four predicting series (43, 44).

Entropy based mt-QSAR equation is able to predict the biological activity of antiviral drugs in more general situations than the traditional QSAR models; which the major limitation is predict the biological activity of drugs against only one viral species. The present model with a very large data set improves significantly the previous QSAR

3.3. Data set

The data set was formed by a set of marketed and/or very recently reported antiviral drugs which low reported $MIC_{50} < 10$ µM against different virus. The data set was conformed to more of 1100 different drugs experimentally tested against some species of a list of 40 virus. Not all drugs were tested in the literature against all listed species so we were able to collect 47 470 cases (drug/species pairs) instead of 1100 x 40 cases.

4. Conclusions

models and may help to perform MOOP of drug activity against different viral species. This mt-QSAR methodology improves models using entropy as a molecular descriptor that allow predicting antiviral activity of any organic compound against a very large diversity of viral pathogens.

References and Notes

1. Prado-Prado J, Martinez de la Vega O, Uriarte E, Ubeira FM, Chou K-C, González-Díaz H. Unified QSAR approach to antimicrobials. 4. Multi-target QSAR modeling and comparative multi-distance study of the giant components of antiviral
drug–drug complex networks. Bioorg Med Chem. 2008;doi:10.1016/j.bmc.2008.11.075.
2. Fratev F, Benfenati E. 3D-QSAR and molecular mechanics study for the differences in the azole activity against yeastlike and filamentous fungi and their relation to P450DM inhibition. 1. 3-substituted-4(3H)-quinazolinones. Journal of chemical information and modeling. 2005 May-Jun;45(3):634-44.
3. Cronin MT, Aptula AO, Dearden JC, Duffy JC, Netzeva TI, Patel H, et al. Structure-based classification of antibacterial activity. J Chem Inf Comput Sci. 2002 Jul-Aug;42(4):869-78.
4. Marrero-Ponce Y, Castillo-Garit JA, Olazabal E, Serrano HS, Morales A, Castanedo N, et al. Atom, atom-type and total molecular linear indices as a promising approach for bioorganic and medicinal chemistry: theoretical and experimental assessment of a novel method for virtual screening and rational design of new lead anthelmintic. Bioorg Med Chem. 2005 Feb 15;13(4):1005-20.
5. Marrero-Ponce Y, Medina-Marrero R, Torrens F, Martinez Y, Romero-Zaldivar V, Castro EA. Atom, atom-type, and total nonstochastic and stochastic quadratic fingerprints: a promising approach for modeling of antibacterial activity. Bioorg Med Chem. 2005 Apr 15;13(8):2881-99.

6. Marrero-Ponce Y, Meneses-Marcel A, Castillo-Garit JA, Machado-Tugores Y, Escario JA, Barrio AG, et al. Predicting antitrichomonal activity: a computational screening using atom-based bilinear indices and experimental proofs. Bioorg Med Chem. 2006 Oct 1;14(19):6502-24.

7. Montero-Torres A, Vega MC, Marrero-Ponce Y, Rolon M, Gomez-Barrio A, Escario JA, et al. A novel non-stochastic quadratic fingerprints-based approach for the 'in silico' discovery of new antitrypanosomal compounds. Bioorg Med Chem. 2005 Nov 15;13(22):6264-75.

8. Meneses-Marcel A, Marrero-Ponce Y, Machado-Tugores Y, Montero-Torres A, Pereira DM, Escario JA, et al. A linear discrimination analysis based virtual screening of trichomonacidal lead-like compounds: outcomes of in silico studies supported by experimental results. Bioorg Med Chem Lett. 2005 Sep 1;15(17):3838-43.

9. Vega MC, Montero-Torres A, Marrero-Ponce Y, Rolon M, Gomez-Barrio A, Escario JA, et al. New ligand-based approach for the discovery of antitrypanosomal compounds. Bioorg Med Chem Lett. 2006 Apr 1;16(7):1898-904.

10. Marrero-Ponce Y, Meneses-Marcel A, Rivera-Borroto OM, Garcia-Domenech R, De Julian-Ortiz JV, Montero A, et al. Bond-based linear indices in QSAR: computational discovery of novel anti-trichomonal compounds. J Comput Aided Mol Des. 2008 Aug;22(8):523-40.

11. Garcia-Domenech R, Galvez J, de Julian-Ortiz JV, Pogliani L. Some new trends in chemical graph theory. Chem Rev. 2008 Mar;108(3):1127-69.

12. Marrero-Ponce Y, Khan MT, Casanola-Martin GM, Ather A, Sultankhodzhaev MN, Garcia-Domenech R, et al. Bond-based 2D TOMOCOMD-CARDD approach for drug discovery: aiding decision-making in 'in silico' selection of new lead tyrosinase inhibitors. J Comput Aided Mol Des. 2007 Apr;21(4):167-88.

13. Garcia-Garcia A, Galvez J, de Julian-Ortiz JV, Garcia-Domenech R, Munoz C, Guna R, et al. Search of chemical scaffolds for novel antituberculosis agents. J Biomol Screen. 2005 Apr;10(3):206-14.

14. Garcia-Garcia A, Galvez J, de Julian-Ortiz JV, Garcia-Domenech R, Munoz C, Guna R, et al. New agents active against Mycobacterium avium complex selected by molecular topology: a virtual screening method. J Antimicrob Chemother. 2004 Jan;53(1):65-73.

15. Meneses-Marcel A, Rivera-Borroto OM, Marrero-Ponce Y, Montero A, Machado Tugores Y, Escario JA, et al. New antitrichomonal drug-like chemicals selected by bond (edge)-based TOMOCOMD-CARDD descriptors. J Biomol Screen. 2008 Sep;13(8):785-94.

16. Vilar S, Santana L, Uriarte E. Probabilistic neural network model for the in silico evaluation of anti-HIV activity and mechanism of action. J Med Chem. 2006;49(3):1118-24.

17. Cruz-Monteagudo M, Borges F, Cordeiro MN, Cagide Fajin JL, Morell C, Ruiz RM, et al. Desirability-based methods of multiobjective optimization and ranking for global QSAR studies. Filtering safe and potent drug candidates from combinatorial libraries. J Comb Chem. 2008 Nov-Dec;10(6):897-913.

18. González-Díaz H, Prado-Prado FJ, Santana L, Uriarte E. Unify QSAR approach to antimicrobials. Part 1: Predicting antifungal activity against different species. Bioorg Med Chem. 2006 Jun 5;14 5973–80.

19. González-Díaz H, Prado-Prado F. Unified QSAR and Network-Based Computational Chemistry Approach to Antimicrobials, Part 1: Multispecies Activity Models for Antifungals. J Comput Chem. 2008;29:656-7.

20. Todeschini R, Consonni V. Handbook of Molecular Descriptors. Mannhold R, Kubinyi H, Timmerman H, editors: Wiley-VCH; 2002.

21. Gonzalez-Diaz H, Prado-Prado F, Ubeira FM. Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach. Curr Top Med Chem. 2008;8(18):1676-90.

22. González-Díaz H, Cabrera-Pérez MA, Agüero-Chapín G, Cruz-Monteagudo M, Castañedo-Cancio N, del Río MA, et al. Multi-target QSPR assemble of a Complex Network for the distribution of chemicals to biphasic systems and biological tissues. Chemometrics Intellig Lab Syst. 2008;94:160-5.

23. Cruz-Monteagudo M, González-Díaz H, Agüero-Chapin G, Santana L, Borges F, Domínguez RE, et al. Computational Chemistry Development of a Unified Free Energy Markov Model for the Distribution of 1300 Chemicals to 38 Different Environmental or Biological Systems. J Comput Chem. 2007; 28:1909-22.

24. González-Díaz H, Aguero G, Cabrera MA, Molina R, Santana L, Uriarte E, et al. Unified Markov thermodynamics based on stochastic forms to classify drugs considering molecular structure, partition system, and biological species: distribution of the antimicrobial G1 on rat tissues. Bioorg Med Chem Lett. 2005 Feb 1;15(3):551-7.

25. González-Díaz H, Uriarte E. Proteins QSAR with Markov average electrostatic potentials. Bioorg Med Chem Lett. 2005 Nov 15;15(22):5088-94.

26. Saiz-Urra L, González-Díaz H, Uriarte E. Proteins Markovian 3D-QSAR with spherically-truncated average electrostatic potentials. Bioorg Med Chem. 2005 Jun 1;13(11):3641-7.

27. Ferino G, Delogu G, Podda G, Uriarte E, González-Díaz H. Quantitative Proteome-Disease Relationships (QPDRs) in Clinical Chemistry: Prediction of Prostate Cancer with Spectral Moments of PSA/MS Star Networks. In: Mitchem BHaS, Ch.L., editor. Clinical Chemistry Research (ISBN: 978-1-60692-517-1). NY: Nova Science Publisher; 2009.

28. Concu R, Podda G, Uriarte E, González-Díaz H. A New Computational Chemistry & Complex Networks approach to Structure-Function and Similarity Relationships in Protein Enzymes. In: Collett CTaR, C.D., editor. Handbook of Computational Chemistry Research: Nova Science Publishers 2009.

29. González-Díaz H, González-Díaz Y, Santana L, Ubeira FM, Uriarte E. Proteomics, networks and connectivity indices. Proteomics. 2008;8:750-78.

30. González-Díaz H, Vilar S, Santana L, Uriarte E. Medicinal Chemistry and Bioinformatics – Current Trends in Drugs Discovery with Networks Topological Indices. Curr Top Med Chem. 2007;7(10):1025-39.

31. Gonzalez-Diaz H, Saiz-Urra L, Molina R, Santana L, Uriarte E. A model for the recognition of protein kinases based on the entropy of 3D van der Waals interactions. Journal of proteome research. 2007 Feb;6(2):904-8.

32. González-Díaz H, Marrero Y, Hernandez I, Bastida I, Tenorio E, Nasco O, et al. 3D-MEDNEs: an alternative "in silico" technique for chemical research in toxicology. 1. prediction of chemically induced agranulocytosis. Chem Res Toxicol. 2003 Oct;16(10):1318-27.

33. González-Díaz H, Molina R, Uriarte E. Markov entropy backbone electrostatic descriptors for predicting proteins biological activity. Bioorg Med Chem Lett. 2004 Sep 20;14(18):4691-5.
34. Graham DJ. Information Content in Organic Molecules: Brownian Processing at Low Levels. Journal of chemical information and modeling. 2007;47(2):376-89.
35. Graham DJ, Schacht D. Base Information Content in Organic Molecular Formulae. J Chem Inf Comput Sci. 2000;40:942.
36. Graham DJ. Information Content in Organic Molecules: Structure Considerations Based on Integer Statistics. J Chem Inf Comput Sci. 2002;42:215.
37. Graham DJ, Malarkey C, Schulmerich MV. Information Content in Organic Molecules: Quantification and Statistical Structure via Brownian Processing. . J Chem Inf Comput Sci. 2004;44(1601).
38. Graham DJ, Schulmerich MV. Information Content in Organic Molecules: Reaction Pathway Analysis via Brownian Processing. J Chem Inf Comput Sci. 2004;44(1612).
39. Graham DJ. Information Content and Organic Molecules: Aggregation States and Solvent Effects. Journal of chemical information and modeling. 2005;45(1223).
40. Gonzalez-Diaz H, Tenorio E, Castanedo N, Santana L, Uriarte E. 3D QSAR Markov model for drug-induced eosinophilia--theoretical prediction and preliminary experimental assay of the antimicrobial drug G1. Bioorg Med Chem. 2005 Mar 1;13(5):1523-30.
41. González-Díaz H, Cruz-Monteagudo M, Molina R, Tenorio E, Uriarte E. Predicting multiple drugs side effects with a general drug-target interaction thermodynamic Markov model. Bioorg Med Chem. 2005 Feb 15;13(4):1119-29.
42. Cruz-Monteagudo M, González-Díaz H. Unified drug-target interaction thermodynamic Markov model using stochastic entropies to predict multiple drugs side effects. Eur J Med Chem. 2005 Oct;40(10):1030-41.
43. Van Waterbeemd H. Discriminant Analysis for Activity Prediction. In: Van Waterbeemd H, editor. Chemometric methods in molecular design. New York: Wiley-VCH; 1995. p. 265-82.
44. StatSoft.Inc. STATISTICA (data analysis software system), version 6.0, www.statsoft.com.Statsoft, Inc. 6.0 ed2002. p. STATISTICA (data analysis software system), version 6.0, www.statsoft.com.Statsoft.

SciForum
Mol2Net

Prognostic Value of Affective Symptomatology in First-Admitted Psychotic Patients

Marta Arrasate [1,*], Itxaso González-Ortega [2], Adriana García-Alocén [2], Susana Alberich [2], Iñaki Zorrilla [2] and Ana González-Pinto [2]

[1] RSMB- Algorta-Bizkaia, EHU/UPV University, Basque Country, Spain
[2] CIBERSAM, Department of Psychiatry. Araba University Hospital - Santiago, EHU/UPV, Vitoria-Gasteiz, Basque Country, Spain

* Author to whom correspondence should be addressed; E-Mail: marta.arrasategil@osakidetza.net.

Published: 4 December 2015

Abstract: Objective: to analyze the predictive value of affective symptomatology in a first psychotic episode sample followed up during three and five years, regarding to hospitalization, relapses, suicidal behaviour, working level, social activity and global functioning. **Method:** 112 inpatients with a first psychotic episode were included in a longitudinal-prospective study followed up during three (N=91) and five-year (N=82). Assessments included the YMRS and HRDS-21, the GAF, the Strauss-Carpenter prognostic scale, the PANSS and the Phillips pre-morbid adjustment scale. We used descriptive and logistic analysis to determine the predictive factors associated to the number of relapses, hospitalizations and suicide attempts; depressive, manic, activation and dysphoric dimensions as covariables. **Results:** 91.46% of relapses and 21% of suicide attempts at fifth year. The GAF discriminated among prognostic groups from the third year (p 0.020), with the poorest prognosis in the schizophrenia group, while bipolar disorders and the rest of the diagnoses achieved an intermediate prognosis. The Strauss-Carpenter scale, specifically working, social activity and global functioning items, discriminated among three diagnostic groups and between affective and non-affective psychosis ($p<0.05$); while schizophrenia scored the poorest outcome, bipolar disorder scored the highest. Depressive dimension was significantly associated with a lower number of relapses and hospitalizations ($p= 0.045$ and $p= 0.012$) and manic dimension with more relapses ($p= 0.023$). **Conclusion:** The depressive dimension presents the best prognosis. On the contrary, the activation dimension, in general, gives a more favourable prognosis with regards to functionality (social) and unfavourable with respect to relapses. Finally, the manic dimension is associated with a worse

Prof. Luis Lezama, Department of Inorganic Chemistry, University of Basque Country (UPV/EHU), Leioa, Sarriena w/n, Bizkaia.

Prof. Ramón J Estévez, Department of Organic Chemistry, University of Santiago de Compostela (USC), Coordinator of PhD Program in Chemical Science and Technology , and MSc Program Organic Chemistry.

Assoc. Prof. Ana Gonzalez-Pinto Arrillaga, M.D., Ph.D. Head of Stanley Center Category Research Group, Department of Neurosciences, University of the Basque Country (UPV/EHU), Head of Psychiatry Research of Osakidetza (Basque Public Health System), Head of Medical Psychiatry Service, University Hospital Santiago Apostol de Vitoria-Gasteiz, Vitoria.

Prof. Carmen Cadarso-Suarez, Unit of Biostatistics, Department of Statistics and Operations Research, School of Medicine, University of Santiago de Compostela (USC), Spain.

Prof. Javier Meana, M.D., Ph.D. Department of Pharmacology, Faculty of Medicine, University of the Basque Country (UPV/EHU), Bizkaia.

Prof. Eugenio Uriarte, Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, USC, Spain.

Prof. Jose Angel Irabien Gulias, Department of Chemical and Biomolecular Engineering ETSIIT, University of Cantabria, Spain.

Prof. Fernando Martin Sanchez, PhD. Foundational Chair of Health Informatics at Melbourne Medical School, Professor Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Australia.

Prof. Victor Maojo, M.D. Ph.D., Biomedical Informatics Group, Polytechnic University of Madrid, Spain.

Prof. Mario Piris, Ikerbasque Professor, Faculty of Chemistry, University of the Basque Country UPV/EHU, Donostia International Physics Center (DIPC), P.K. 1072, 20080 Donostia, Euskadi, Spain.

CEO, Dean, & Directors OF Science (Advisory Honor Committee)

Prof. Fernando Cossío, President of IKERBASQUE, Basque Foundation for Science, Prof. Department of Organic Chemistry I, University of Basque Country (UPV/EHU), Donostia - San Sebastián Campus, Gipuzkoa.

Prof. Allen B. Reitz, Ph.D., CEO Fox Chase Chemical Diversity Center, Inc., Doylestown, PA, USA. Editor-in-Chief of the journal Current Topics in Medicinal Chemistry, Adjunct Professor at Drexel University College of Medicine, Moore Fellow in the Management of Technology University of Pennsylvania (Wharton, Penn Engineering), Founder CEO of ALS Biopharma, LLC.

Prof. Danail Bonchev, Director of Research on Bioinformatics, Center for the Study of Biological Complexity. Professor, Department of Mathematics and Applied Mathematics, Virginia Commonwealth University (VCU), USA.

Prof. Jose María Pitarke, Full Professor of Condesed Matter Physics, UPV/EHU, Director of Nanomaterials Cooperative Research Center (CICNanoGune), Tolosa Hiribidea, 76, E-20018 Donostia – San Sebastian, Gipuzkoa.

Porf. Jesús Jimenez Barbero, Ikerbasque Professor, Scientific Director of Center for Cooperative Research in Biosciences (CICBiogune), Bizkaia.

Prof. Luis M Liz-Marzán, Ikerbasque Senior Professor, Scientific Director of Center for Cooperative Research in Biomaterials (CICbiomaGUNE), Gipuzkoa.

Prof. Francesc Illas Riera, Director of Institute of Theoretical and Computational Chemistry (IQTCUB), Physical Chemistry Department, Faculty of Chemistry, University of Barcelona.

Full. Prof. Yiyu Cheng, Director of Pharmaceutical Informatics Institute, Zhejiang University (ZJU), China.

Prof. Dr. Peter Langer, Full Professor (C4) of Organic Chemistry, Vice Director Institute of Chemistry, University of Rostock, Head of the Department of Organic Chemistry. Universität Rostock Institut für Chemie Abteilung für Organische Chemie Albert-Einstein-Straße 3a 18059 Rostock.

Heads of Hospital Medical Service (Advisory Honor Committee)

Prof. Mariano Provencio, Ph.D., D.M., Head of Medical Oncology Service, Universitary Hospital Puerta de Hierro (HUPH), Autonomous University of Madrid (UAM), Madrid, Spain.

Dr. A Rodríguez-Antigüedad Zarrans, Head of Medical Service of Neurology Hospital of Basurto, Bilbao, President Spain Society of Neurology (SEN). Prof. Department of Neuroscience, Faculty of Medicine UPV/EHU, Leioa, Sarriena w/n, Bizkaia.

Assoc. Prof. Ana Gonzalez-Pinto Arrillaga, M.D., Ph.D. Head of Stanley Center Category Research Group, Department of

evolution regarding relapses. Only the dysphoric dimension is not associated with syndromic and/or functional prognosis.

Keywords: prognostic, affective, dimension, first psychosis.

1. Introduction

First episode psychosis includes a heterogeneous population which represents an extensive number of diagnoses. Today's classifications systems are every time more focused in the inclusion of dimensions versus categories in psychiatry, and the clinical definition of psychosis may involve only one part of the total psychosis phenotype[1].

Little is studied about the influence of affective symptomatology in functional psychosis and results are frequently controversial. Moreover, these studies are nearly non-existent in first psychotic episode, and only a few of them used a dimensional approach. Therefore, dimensional representations would be useful to predict the clinical course and treatment needs in first episode psychosis.

Crow [2] and van Os [3] suggested the hypothesis of the psychopathological continuum where different diagnostic categories share dimensional factors which could refer to similar neurobiological mechanisms for each of the dimensions regardless of the type of psychosis. Dimensions are not diagnostic-specific and have been reasonably replicable in psychosis, stable solutions in a variety of settings, diagnostic groups and patient samples [4]. Initial work was done on schizophrenia, finding a three-factor solution, with positive, negative and disorganized dimensions [5]. Afterwards, Cassidy [6], Serretti [7] and Disalver [8] examined the factor structure of the bipolar disorder. González-Pinto et al. [9] obtained a five-factor solution in a 103 bipolar disorder sample. Later, samples included the full spectrum of psychosis, and five-factor solutions were found, including manic and depressive dimensions [10-15]. Finally, factor structure analyses were targeted to first psychotic episode samples [4,16-21].

Regarding to the influence of the affective symptomatology in psychosis, some authors found that affective symptomatology associates good prognosis [17,22-25]; some of them associated the better prognosis specifically with the depressive dimension [26]; others, like Paillére-Martinot [27] associated the better prognosis to a higher score on the GAF (Global Assessment of Functioning) [28]. Both van Os [17] and Allardyce et al.[4] associated the manic dimension with a good outcome; the first one specified fewer symptoms and their lesser severity, while the latter associated it with being married and working; McIntosh et al.[19] also found a good outcome related to depression dimension.

However, others researchers found a negative association between depression and outcome: Geddes et al [29] found early relapse and more time in hospital; Birchwood [30] found early relapse; Meng et al. [31] also associated it with a poor prognosis. Thara et al.[32] associated longer time with symptoms with manic descompensation. Power et al.[33] associated affective symptoms with more hospitalizations. Finally Sipos et al.[34] also associated it with a poor outcome.

In conclusion, our **objective** was to study the predictive value of affective symptomatology in a first psychotic episode sample followed up during three and five years, using a dimensional approach. We studied outcome in terms of hospitalization, relapses, suicidal behaviour, working level, social activity and global functioning.

.

2. Results and Discussion

Patient Sociodemographic and Clinical Characteristics

A total of 112 patients with a first psychotic episode were included in the study at baseline. Of these 112 patients, 91 (81.25%) and 82 (73.2%) patients were available for analysis at 3 and 5 years' follow-up. At baseline, the mean age of the total sample was 28.8 years (SD = 10.3) and 75 (67%) were men. Initial DSM-IV diagnosis at baseline included bipolar disorder (23.2%), schizophrenia (15.2%) and other diagnosis (61.6%). Sociodemographic and clinical baseline characteristics are describe in a previous work[50.]

There were no differences between patients followed or not followed with respect to the following baseline variables: age (U= 1023, p=0.62), sex (χ2 =0.30, p=0.58), marital status (Fisher, p= 0.69), socioeconomic level (Fisher, p=0.27) and tobacco use (Fisher, p= 0.53).

Diagnostic Categories

The patient sample was classified both into three diagnostics groups: (1) those with schizophrenia diagnosed; (2) those with bipolar disorder diagnosed; and (3) those with other psychosis, and two diagnostic groups: affective psychosis (bipolar disorder, depressive disorder) and non-affective psychosis (the rest of the psychosis).

Of the 91 patients at 3-year follow-up, 25 (27.47%), had a diagnosis of schizophrenia, 34 (37.36%) bipolar disorder and 32 (35.17%) were classified as other psychosis. Final diagnosis at fifth year were: 34.14% of the patients have schizophrenia, 37% bipolar disorder and 29.26% other psychosis.

Prognostic Groups

Of the 91 patients, 20.9%had a good prognosis (GAF ≥71), 51.6% intermediate prognosis (GAF 51-70) and 27.5% (GAF ≤50) had a bad prognosis at 3-year. And of the 82 patients at fifth year, 23.7% had a good prognosis, 51.3% intermediate prognosis and 25% bad prognosis. See table 1 for Strauss-Carpenter.

Affective Dimensions

As previously reported, factor structure analysis[9] produced a five-factor solution explaining 60.8% of the total variance in a sample of patients with bipolar disorder. In the present study, we analysed four of these affective dimensions. The *depressive dimension* at baseline included symptoms of depressed mood, suicidal thoughts, feeling of guilt, obsessive and compulsive symptoms, and anxiety, and had a mean score of 3.92 (SD = 3.65). The *dysphoric dimension* at baseline included disruptive-aggressive behaviour, irritability, and lack of insight, and had a mean score of 8.55 (SD = 4.89). The *manic dimension* at baseline included appearance, sexual interest, elevated mood and reduced sleeping, and had a mean score of 5.27 (SD = 3.44). Finally, the *activation dimension* at baseline included speech difficult to understand, increased motor activity-energy and language-thought disorder, and had a mean score of 5.37 (SD = 4.78).

Clinical Characteristics at Follow-up

Of the 91 patients at third year: 80.2% had relapses, 61.5% hospitalizations and 19.8% suicide attempts during the follow up. Of the 82 at fifth year: 91.46% have relapses, 73.17% hospitalizations and 21% suicide attempts along the total follow-up period.

Outcome by GAF and Diagnostic Categories

The GAF discriminated among prognostic groups from the third year of the follow up (X2 11.725; p 0.020): the poorest prognosis in the schizophrenia group, while bipolar disorders and the rest of the diagnoses achieved an intermediate prognosis, with the bipolar disorder group as having a slightly better prognosis. Figure 1.

Outcome by Strauss-Carpenter and Diagnostic Categories

The Strauss-Carpenter scale, specifically working item (X2=10.551; p 0.032 / X2=8.661; p 0.013), social activity item (X2= 16.231; p 0.003 / X2=6.237; p 0.044) and global functioning item (X2=12.742; p 0.013 / X2=11.443; p 0.003) discriminated among three diagnostic groups and between affective and non-affective psychosis (X2=8.611; p 0.013 for hospitalization item; X2=6.237; p 0.044 for working activity item and X2=11.443; p 0.003 for social activity item) at fifth year. At work functioning: in schizophrenia, 53.6% have a bad prognosis, 28.6% intermediate prognosis and 17.9% a good one; in bipolar disorder, 41.2% bad prognosis, 5.9% intermediate and 52.9% good prognosis; for the rest of psychosis, 45% bad, 15% intermediate and 40% a good prognosis. At social functioning: in schizophrenia, 35.7% bad, 35.7% intermediate and 28.6% good prognosis; in bipolar disorder, 20.6% bad, 11.8% intermediate and 67.6% good prognosis; and for the rest of psychosis, 15% bad, 5% intermediate and 80% good prognosis. At global functioning: in schizophrenia, 42.9% have bad prognosis, 50% intermediate and 7.1% good prognosis; in bipolar disorder, 17.6% bad, 35.3% intermediate and 47.1% good prognosis; and finally, for the rest of psychosis, 25% bad, 45% intermediate and 30% good prognosis. Therefore, while schizophrenia scored the poorest outcome at work functioning, social activity and global functioning, bipolar disorder scored the highest. Figures 2, 3, 4

Diagnostic Predictive Value of Affective Dimensions

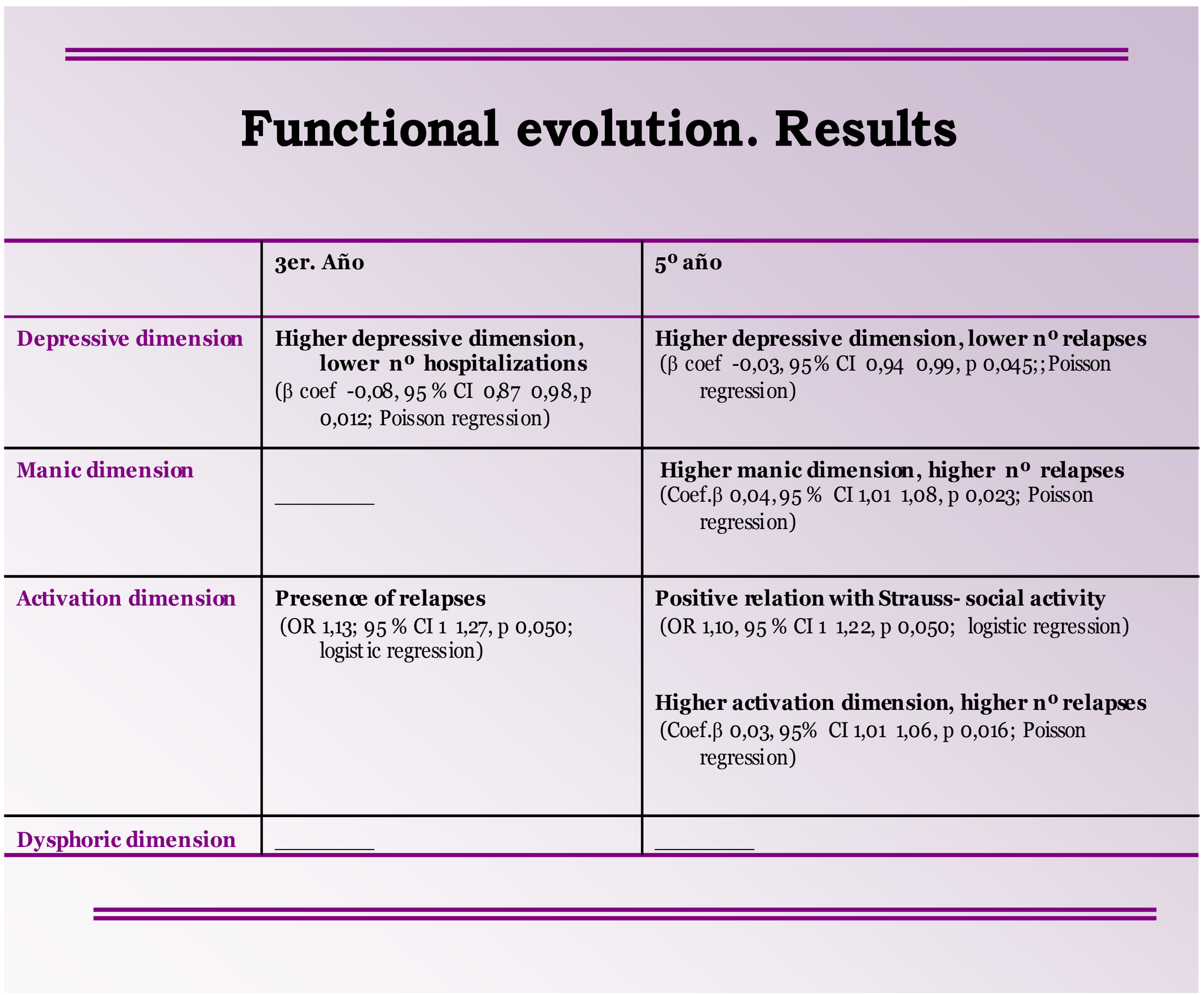
The predictive value of affective symptomatology was also determined by analysing the influence of dimensions on hospitalizations, relapses, suicidal behaviour, working activity, social activity and global functioning, using regression models.

With respect to the depressive dimension, we observed that it significantly associated with a lower number of relapses at fifth year and hospitalizations at 3-year (β coef -0,03, 95 % CI 0,94 0,99, p 0.045 and β coef -0,08, 95 % CI 0,87 0,98, p 0.012), meanwhile manic dimension was significantly associated with more relapses (Coef.β 0,04, 95 % CI 1,01 1,08, p 0.023) at fifth year. Finally, activation dimension was significantly associated with the presence (OR 1,13; 95 % CI 1 1,27, p 0.050) and higher number of relapses (OR 1,10, 95 % CI 1 1,22, p 0,050) and with a more benign illness in terms of social activity in Strauss-Carpenter (Coef.β 0,03, 95% CI 1,01 1,06, p 0.016) at fifth year. However, dysphoric dimension was the unique dimension not significantly associated with any of the tested variables. Table 2.

Table 1. Frequencies in % in respect to Strauss-Carpenter at third and fifth years by prognostic groups.

Strauss-Carpenter	**Prognosttic groups**	**Third year**	**Fifth year**
Hospitalization	**Good prognosis** (punctuation: 4)	62,6 %	92,7 %
	Intermediate prognosis (punctuation:2 and 3)	36,3 %	4,9 %
	Bad prognosis (punctuation:0 and 1)	1,1 %	2,4 %
Work activity	**Good prognosis** (punctuation:4)	31,9 %	37,8 %
	Intermediate prognosis (punctuation:2 and 3)	36,3 %	15,9 %
	Bad prognosis (punctuation: 0 y 1)	31,9 %	46,3 %
Social activity	**Good prognosis** (punctuation: 4)	35,2 %	57,3 %
	Intermediate prognosis (punctuation: 2 and 3)	38,5 %	18,3 %
	Bad prognosis (punctuation:0 and 1)	26,4 %	24,4 %
Global functioning	**Good prognosis** (punctuation:4)	16,5 %	29,3 %
	Intermediate prognosis (punctuation:2 and 3)	62,6 %	42,7 %
	Bad prognosis (punctuation:0 and 1)	20,9 %	28 %

Table 2. Results of functional evolution.



Functional evolution. Results

	3er. Año	5º año
Depressive dimension	**Higher depressive dimension, lower nº hospitalizations** (β coef -0,08, 95 % CI 0,87 0,98, p 0,012; Poisson regression)	**Higher depressive dimension, lower nº relapses** (β coef -0,03, 95% CI 0,94 0,99, p 0,045;; Poisson regression)
Manic dimension	________	**Higher manic dimension, higher nº relapses** (Coef.β 0,04, 95 % CI 1,01 1,08, p 0,023; Poisson regression)
Activation dimension	**Presence of relapses** (OR 1,13; 95 % CI 1 1,27, p 0,050; logist ic regression)	**Positive relation with Strauss- social activity** (OR 1,10, 95 % CI 1 1,22, p 0,050; logistic regression) **Higher activation dimension, higher nº relapses** (Coef.β 0,03, 95% CI 1,01 1,06, p 0,016; Poisson regression)
Dysphoric dimension	________	________

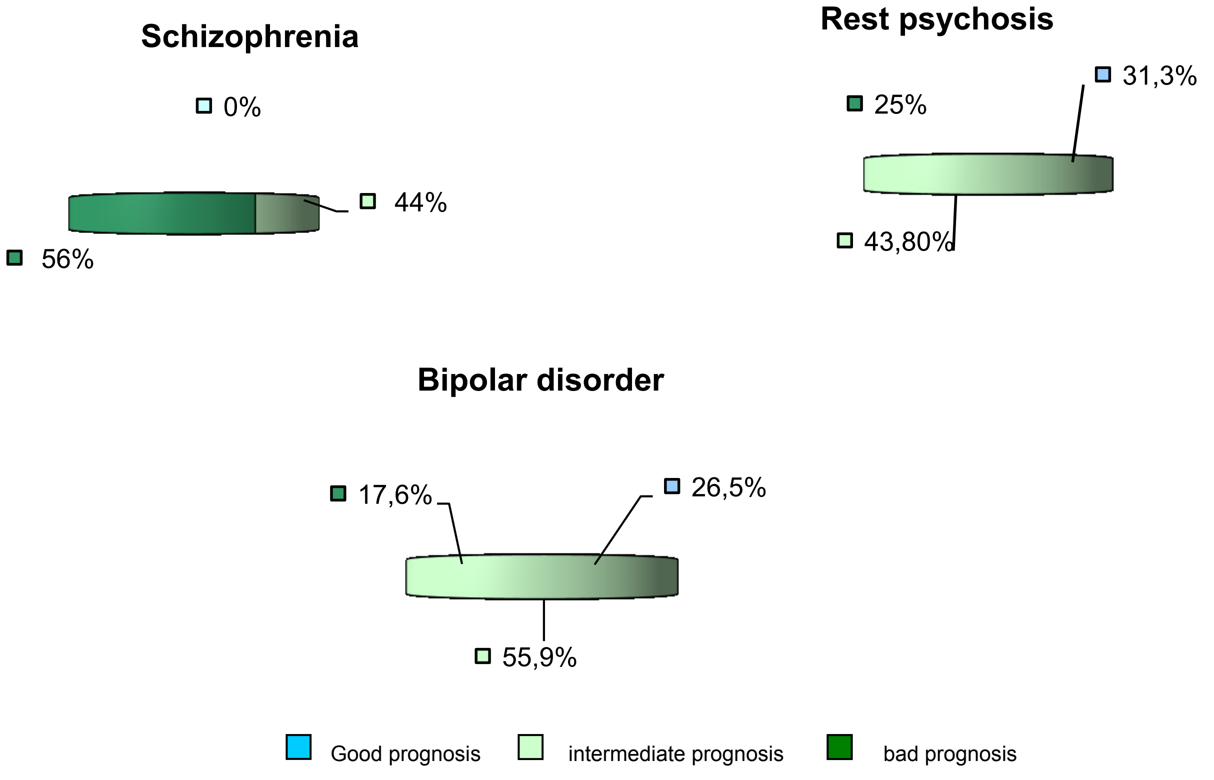


Figure 1. Prognostic by GAF and by diagnostic groups, at 3rd year.

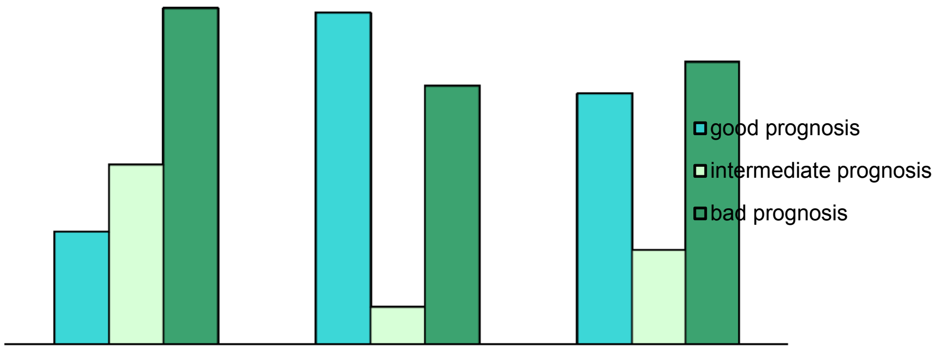


Figure 2. Working activity prognosis by diagnostic groups, at 5th year.

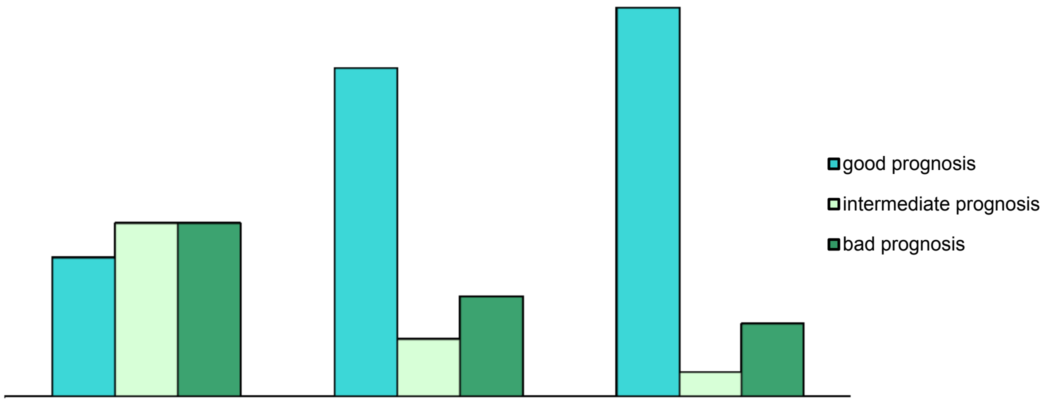


Figure 3. Social activity prognosis by diagnostic groups, at 5th year.

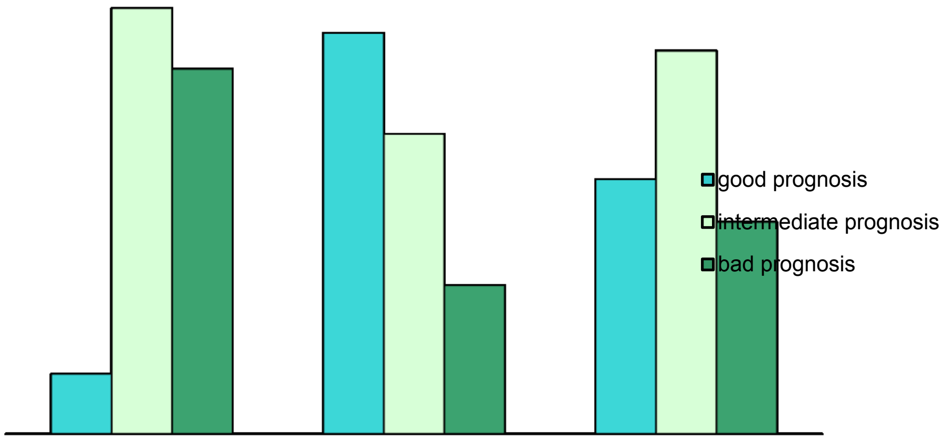


Figure 4. Global functioning prognosis by diagnostic groups, at 5th year.

3. Materials and Methods

Study Design and Participants

This was a prospective, longitudinal study of 112 patients presenting with a first episode of psychosis between January 1996 and December 1997, and who were admitted to the only psychiatric inpatient unit in the Vitoria-Gasteiz region of Spain. First episode psychosis was defined as the first time a patient presented with psychotic symptomatology, consisting of the presence of one or more of the following symptoms: delusions, hallucinations, grossly disorganized behaviour and marked thought disorder.

Patients, aged 16-65 years, were included in the study if they met the diagnostic criteria of the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders[35] (DSM-IV) for schizophreniform disorder, schizoaffective disorder, schizophrenia, delusional disorder, brief psychotic disorder, atypical psychosis or psychotic disorder not otherwise specified, bipolar I or II disorder, or major depressive disorder with psychotic symptoms (American Psychiatric Association, 1994). The DSM-IV axis I diagnosis was made using the Structured Clinical Interview for DSM-IV[36] (SCID-I) (Spitzer et al., 1996); the same interviewers for baseline and follow-up assessments. Subjects with mental retardation, organic brain disorders and substance-induced psychotic disorders as their main diagnosis were excluded from the study.

The study was approved by the ethics committee of the hospital and all participants provided informed consent.

Assessments

Assessments were made at baseline and at 3 and 5 years of follow-up. The baseline assessment was performed within 24 hours of hospitalization for the first psychotic episode and reflected the patient's clinical status during the previous week. After hospital discharge, subjects attended their corresponding mental health care centre.

Data collected included patient sociodemographics and clinical characteristics. Patients were assessed by different raters from those who assessed the diagnosis, using the following scales: Young Mania Rating Scale (YMRS) (Young et al., 1978)[37], Hamilton Depression Rating Scale (HDRS-21) (Hamilton, 1960) [38,39], Global Assessment of Functioning (GAF) (American Psychiatric Association, 1987) [40], Phillips Rating Scale of Premorbid Adjustment in Schizophrenia (Phillips) (Phillips, 1953) [41], Strauss-Carpenter Scale (Strauss and Carpenter, 1972) [42] and the Positive and Negative Syndrome Scale (PANSS) (Kay et al., 1986) [43]. Additional information provided by family informants and from staff observations was incorporated into the rating process. All interviews were carried out independently by one psychiatrist and one psychologist who demonstrated good inter-rater reliability for SCID diagnoses ($\kappa = 0.88$), YMRS ($\kappa = 0.90$), HDRS-21 ($\kappa = 0.93$), GAF ($\kappa = 0.94$), Phillips ($\kappa = 0.80$), Strauss-Carpenter ($\kappa = 0.81$) and PANSS ($\kappa = 0.82$).

The affective dimensions used in the present study were based on a previous factor structure analysis using the YMRS and HDRS-21 in 103 patients with bipolar disorder[9.] This gave a five-factor solution and the component symptom loadings obtained for each of the affective dimensions (depressive, dysphoric, manic, psychosis and activation) is summarised in a previous work (González-Pinto et al., 2003) [9]. Factor structure analysis has been widely used for research purposes and in clinical trials for studying the symptom dimensions of psychosis [4,11,14, 16-19, 34, 44]. In the present study, we analysed four affective dimensions (depressive, dysphoric, manic and activation; baseline scores); the psychosis factor was not used because all patients presented with psychosis symptoms.

The patient sample was classified both into three diagnostics groups: (1) those with schizophrenia diagnosed; (2) those with bipolar disorder diagnosed; and (3) those with other psychosis, and into two diagnostic groups: affective psychosis (bipolar disorder, depressive disorder) and non-affective psychosis (the rest of the psychosis).

In respect to the GAF Scale, the followings groups were considered to describe outcome among diagnostic categories (schizophrenia, bipolar disorder and the rest of the psychosis): good prognostic for the punctuation ≥71, intermediate prognostic for 51-70 and bad prognostic for ≤50.

Likewise, for the Strauss-Carpenter Prognostic Scale, a good prognostic group when 4 punctuation was scored in all the items evaluated, an intermediate prognostic group for 2 and 3, and finally a bad prognostic group for 1 and 0.

Statistical Analysis

Statistical packages used for the analyses were SAS, SPSS and R 2.5.1.

Baseline characteristics of the total study sample were described using summary statistics (means and standard deviations (SD) or median and range, as appropriate, for continuous variables, and frequencies for categorical variables). Statistical comparisons between groups were performed using the χ^2 test (or Fisher's test where n≤5) for categorical variables and the Student's *t* test or Mann-Whitney *U* test (depending on the distribution of the sample) for continuous variables.

The prognostic value of affective dimensions was examined using regression models, with number of hospitalizations, relapses, suicidal behaviour, working level, social activity and global functioning as the dependent variable. A logistic regression model including all four affective dimensions as independent variables was used to identify which dimensions were predictive of the evolution of first-admitted psychotic patients. Logistic regressions were adjusted by age and gender, negative symptoms (PANSS-N) and premorbid state (Phillips Rating Scale of Premorbid Adjustment) according to the method used by other researchers since it is known these variables influence the outcome. Effect sizes are expressed as odds ratios (ORs) and 95% confidence intervals (CIs) with *P* values. Poisson regressions effect sizes are expressed as β coefficient, 95% confidence intervals (CIs) with *P* values. Associations were considered significant when $P \leq .05$.

We established three cut-points for GAF for statistical purposes: 70, which, in our opinion, divided the sample in two groups, related to a complete recovering or not; 60 [25,34,45]; and finally, 50, following criterions of other researchers [26].

In the case of the Strauss-Carpenter scale, the cut-points were the followings: 4 vs the rest of the values for the hospitalization item [25,46,47], working activity item [48,49]and the global functioning item[25]; we considered 0 and 1 vs the rest of the values for the social activity item, considering that this cut-point divided patients in two completely different groups [49].

4. Discussion

This prospective, longitudinal study of the predictive diagnostic value of affective symptomatology in a sample of hospitalized first-episode psychosis patients followed-up over 5 years shows that affective dimensions (manic, activation, dysphoric and depressive) have different kind of influence in the prognostic of psychosis.

Regarding number of relapses, our percentage is high, 80.2%-91.46% . While Robinson et al.[51] also found a high percentage of relapse (86.2% at fifth year), most authors [27,30,52-54] find 58-78%. Diverse definitions of the "relapse term" may be considered; besides, our patients are hospitalized and their severity is higher. In our study, manic and activation dimensions are associated with higher number of relapses, while depressive dimension protects against them.

In respect to the number of hospitalizations, while 61.5% of the total samples were hospitalized sometime in the first three years, 73.17% were hospitalized at the end of the following period; Power et al.[33] confirmed this percentage. Means of both periods are similar and identical to Sipos et al.[34]. Some authors find higher number. This point depends on a variety of factors: organization of both intra and extra mental services and accessibility. In our study, depressive dimension protects against hospitalizations.

With regard to the number of suicides, 19.8% at third year and 21% at fifth year, our percentages are identical to Birchwood et al.[30], van Os et al.[17], Verdoux et al.[55] and Robinson et al.[51], and the mean is similar in both periods. Two patients committed suicide in the last two years (2.4%); unfortunately, not for being the first years of the illness, suicide risk is disminished [30].

Additionally, and with respect to the outcome assessed by the Strauss-Carpenter Prognostic Scale: this scale discriminates among the three-diagnostic groups, schizophrenia, bipolar disorder and other psychotic disorders, for working and social activity at third and fifth year and for global functioning at fifth year; also discriminates among affective and non-affective psychosis. Prognosis gets better within time of evolution. While schizophrenia scored the poorest outcome at work functioning, social activity and global functioning, bipolar disorder scored the highest.

Furthermore, the GAF discriminates among prognostic groups from the third year of the follow-up: while the schizophrenia has the poorer prognosis [26], the bipolar disorder has the best [24-25] ; the rest of the psychosis have an intermediate prognosis in the outcome. Considering the three diagnostic groups, the majority of the patients are in the group of intermediate prognosis.

In summary, prognosis improved along time of evolution. Although the percentage of relapses is high in our sample, many patients maintained a good level of functioning. Tohen et al.[24] and Swaran et al.[25] pointed out the importance of both sindromic and functional outcome, separatedly.

Additionally, and concerning the prognostic value of affective dimensions, the depressive dimension is significantly associated with fewer relapses and hospitalization at fifth and third years respectively; therefore, it conferres a good prognosis. Many authors confirm a better outcome [14,19,25-27,56] in the presence of depressive symptomatology. Lindenmayer and Kay [57] nevertheless, question themselves about the influence of negative symptoms in that result. We obviously took this problem into account, since our statistical analyses were adjusted by baseline negative symptomatology. Also Peralta et al.[58] found that depressive dimension was associated to negative factors. So, we used assessment tools which are specifically designed for rating affective rather than negative symptomatology. There are also both authors who do not find an association between depressive dimension and outcome[17,32,47,53,59] and some who describe a worse course [10,29,31,60].

The manic dimension is significantly associated with a higher number of relapses at the end of the follow-up period. The activation dimension is also associated with the presence of relapse at the third year and a higher number of relapses at the fifth year. It is also significantly associated with better social functioning. Therefore, the activation dimension is related to the outcome in two ways: better social adjustment, but increased relapse risk. Consequently, both manic and activation dimensions are related to a poorer symptomatic outcome; activation dimension, nevertheless, confers a good functional prognostic. Tohen et al.[24] agree with this afirmation.

Sipos et al.[34] and Gift et al.[59] also find a major risk for hospitalization and Erickson et al.[22] and Allardyce et al.[61] confirmed the better social outcome for manic dimension. On the contrary, Murray et al.[14], McIntosh et al.[19] and van Os et al.[17] described a better symptomatic outcome.

Besides, manic dimension was associated with the absence of suicide attempts as a tendency. In the opinion of the majority of the researchers the depressive dimension is the one which is associated with poorer outcome regarding this subject [14,62,63].

The activation dimension was also nearly significantly associated with a better work level at the third year, which agrees with Allardyce et al.[61].

Finally, the dysphoric dimension was not associated with any of the variables described above and it do not discriminate among all groups.

The fact that these results have been adjusted by negative symptomatology and premorbid adjustment make the results consistent.

In summary, only one of the dimensions is not associated with syndromic and/or functional prognosis, the dysphoric dimension. The depressive dimension presents the best prognosis. On the contrary, the activation dimension, in general, gives a more favourable prognosis with regards to functionality (social) and unfavourable with respect to relapses. Finally, the manic dimension is associated with a worse evolution regarding relapses.

Our results suggest that the affective symptomatology gives a determined prognosis to the evolution of the psychotic illness. Therefore, the systematic evaluation of affectivity will permit us to reach important conclusions regarding the prognosis. The intervention on the patients with manic and activation syndrome could be beneficial in decreasing relapses in the first episodes.

It is of maximum interest to point out that our original contribution is the using of affective dimensions obtained from a bipolar disorder sample and their application to a sample with functional psychosis.

We also would like to mark the representativeness of the sample as our unit is the unique one for acute inpatients in our region. Besides, our study is longitudinal and includes an heterogeneous sample. It also includes a large time of follow-up.

Nevertheless, some limitations must be considered. First, a number of patients were taking medication; we tried to overcome this limitation assessing them within 24-48 hours of hospitalization. Secondly, patients with more severe conditions are probably overrepresented; thus, the results generalization is limited to patients who are hospitalized. Nevertheless, more than 80% of first psychotic episodes are hospitalized. Also, a few of the assessments had been done by telephone when coming was not possible for them. Finally, the main limitation is that we have not adjusted results by drugs; cannabis use is frequent in this kind of patients and we know its influence in psychotic episodes. Therefore, we will choose this issue for future studies. It have not been possible to introduce one more variable for statistical reasons; we adjusted by age,sex, negative symptomatology and premorbid adjustment following the method of most of the authors.

Despite these limitations, definitively our results suggest that affective symptomatology confers a certain prognosis to the course of the illness, so that systematic evaluation of affectivity will make possible conclusions to be obtained in regard to prognosis. Also, intervention in patients with manic and activation syndrome could be benefitial to disminish relapses in first psychotic episodes. The fact that these results were obtained after controlling the analyses by the presence of negative symptoms and premorbid adjustment and, therefore, basal functionality, makes the data be consistent.

Of course, the evolutions of determined variables do not have any reason to reflect the general evolution as was clarified through an evolution study by the World Health Organization [64]; the variables that determine the global evolution are different and varied. This affirmation is in harmony with that mentioned previously with respect to the need to differentiate between the syndromic and functional recovery. One must take into account that this differentiation has its value when proposing the prevention and improving the prognosis of the patients with real possibilities of recovery.

References and Notes

1. McGlashan TH. The Chestnut Lodge follow-up study. II. Long-term outcome of schizophrenia and affective disorders. *Archieves of General Psychiatry.*1984;41:586-601.
2. Crow TJ. The continuum of psychosis and its genetic origins. The sixty-fifth Maudsley lecture. *British Journal of Psychiatry.* 1990;156:788-97.
3. Van Os J, Gilvarry C, Bale R et al. Diagnostic value of the DSM and ICD categories of psicosis: an evidence-based approach. UK700 Group. *Social Psychiatry and Psychiatric Epidemiology.* 2000;35:305-311.
4. Allardyce J, Suppes T, Van Os J. Dimensions and the psychosis phenotype. *Int. Journal Methods Psychiatric Research.* 2007a;16(S1):34-40.
5. Peralta V, Cuesta MJ, Farre C. Factor structure of symptoms in functional psychoses. *Biological Psychiatry.* 1997;42:806-815.
6. Cassidy F, Forest K, Murry E, & Carrol BJ. A factor analysis of the sings and symptoms of mania. *Archieves of General Psychiatry.* 1998;55:27-32.
7. Serreti A, Rietschel M, Lattuada E, Kraub H, Nothen MM, & Smeraldi E. Factor analysis of mania. (Letter to the editor). *Archieves of General Psychiatry.* 1999;56:671-672.
8. Disalver SC, Chen YR, Shoaib AM, & Susan AC. Phenomenology of mania: Evidence for distinc

depressed, dysphoric, and euphoric presentations. *American Journal of Psychiatry.* 1999;15:426-430.

9. González-Pinto A, Ballesteros J, Aldama A et al. Principal components of mania. *Journal of Affective Disorder.* 2003;76(1-3):95-102.
10. Van Os J. To what extent does symptomatic improvement result in better outcome in psychotic illness?. *Psychological Medicine.* 1999;29:1183-1195.
11. Rosenman S, Korten A, Medway J, Evans M. Characterising psychosis in the Australian National Survey of Mental Health and Wellbeing Study on low-prevalence disorders. *Australian and New Zealand Journal of Psychiatry*. 2000;34:792-800.
12. Ventura J, Nuechterlein KH, Subotnik KL et al. Symptom dimensions in recent-onset schizophrenia and mania: a principal components análisis of the 24-item Brief Psychiatric Rating Scale. *Psychiatry Research.* 2000;97:129-135.
13. Drake RJ, Dunn G, Tarrier N et al. The evolution of symptoms in the early course of non-affective psychosis. *Schizophrenia Research*. 2003;63:171-179.
14. Murray V, McKee I, Millar PM. Dimensions and classes of psychosis in a population cohort: a four-class, four-dimension model of schizophrenia and affective psychosis. *Psychological Medicine*2005; 35(49): 499-510.
15. Dikeos DG, Wicham H, McDonald C et al. Distribution of symptom dimensions across Kraepelinian divisions. *British Journal of Psychiatry.* 2006;189:346-353.
16. Kitamura T, Okazaki Y, Fujinawa A, Yoshino M, Kasahara Y. Symptoms of psychoses; a factor-analytic study. *British Journal of Psychiatry*. 1995;166:236-240.
17. Van Os J, Fahy TA, Jones P, Harvey I, Sham P, Lewis S, Bebbington P, Toone B, Williams M, Murray R. Psychopathological syndromes in the functional psychoses: associations with course and outcome. *Psychological Medicine*. 1996;26:161-176.
18. McGorry PD, Bell RC, Dudgeon PL, Jackson HJ. The dimensional structure of first episode psychosis: an exploratory factor analysis. *Psychological Medicine.* 1998;28:935-947.
19. McIntosh AM, Forrester A, Lawrie SM et al. A factor model of the functional psychoses and the relationship of factors to clinical variables and brain morphology. *Psychological Medicine.* 2001;31:159-71.
20. Thakur A, Jagadheesan K, Sinha VK. Psychopathological dimensions in childhood and adolescent psychoses: a confirmatory factor analytical study. *Psychopathology.* 2003; 36(4):190-4.
21. Salvatore P, Khalsa HMK, Hennen J, Tohen M, Yurgelun-Todd D, Casolari F, De Panfilis C, Maggini C, Baldessarini RJ. Psychopathology factors in first-episode affective and non-affective psychotic disorders. *Journal of Psychiatric Research.* 2007;41: 724-736.
22. Erickson DH, Beiser M, Iacono WG, Fleming JAE, Tsung-yi L. The role of social relationships in the course of first-episode schizophrenia and affective psychosis. *American Journal of Psychiatry.* 1989; 146(11):1456-1461.
23. Jonsson H, Nyman AK. Predicting long-term outcome in schizophrenia. *Acta Psychiatrica Scandinavica.* 1991; 83:342-346.
24. Tohen, M., Stoll, A.L., Strakowski, S.M., Faedda, G.L., Myer, P.V., Goodwin, D.C., Kolbrener, M.L., Madigan, A.M. (). The McLean first-episode psychosis project: six-month recovery and

recurrence outcome. *Schizophrenia Bulletin,* 1992; 18(2), 273-282.

25. Swaran P, Singh SP, Croudace T, Amin S, Kwiecinski R, Medley I, Jones PB, Harrison G. Three-year outcome of first-episode psychoses in an established community psychiatric service. *British Journal of Medicine.* 2000;176:210-216.
26. Möller HJ, SCHMID-Bode W et al. Psychopathological and social outcome in schizophrenia versus affective/schizoaffective psychoses and prediction of poor outcome in schizophrenia. *Acta Psychiatrica Scandinavica.* 1988;77:379-389.
27. Paillere-Martinot ML, Aubin F et al. A prognostic study of clinical dimensions in adolescent-onset psychoses. *Schizophrenia Bulletin.* 2000;26(4):789-799.
28. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Revised 3 rd (DSM-III-R). APA: Washington, DC. 1987.
29. Geddes J, Mercer G, Frith CD, Macmillan F, Owens DGC, Johnstone EC. Prediction of outcome following a first episode schizophrenia ; a follow-up study of Northwick Park first episode study subjects. *British Journal of Psychiatry.* 1994;165:664-668.
30. Birchwood M, Todd P, Jackson C. Early intervention in psychosis. *British Journal of Psychiatry.* 1998*;*172(suppl.33):53-59.
31. Meng H, Schimmelmann BG, Mohler B. et al. Pretreatment social functioning predicts 1-year outcome in early onset psychosis. *Acta Psychiatrica Scandinavica.* 2006; 114(4):249-256.
32. Thara R, Henrietta M, Rajkumar S, Eaton WW. Ten-year course of schizophrenia- the Madras longitudinal study. *Acta Psychiatrica Scandinavica.*1994;90:329-336.
33. Power P, Elkins K, Adlar S, Curry C, McGorry P, Harrigan S. Analysis of the initial treatment phase in firs-episode psychosis. *British Journal of Psychiatry.* 1998; 172 (suppl. 33):71-76.
34. Sipos A, Harrison G, Gunnell D, Amin S, Singh SP. Patterns and predictors of hospitalization in first-episode psychosis. *British Journal of Psychiatry.* 2001;178*:*518-523.
35. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV). APA: Washington, DC. 1994.
36. Spitzer RL, Williams JBW, Gibbon M, & First, M B. SCID I. Version 2.0 for DSM IV. Indiana: Lilly Research Laboratories. 1996.
37. Young, R. C., Biggs, T., Ziegler, E., & Meyer, D. A. (1978). A rating Scale for mania: reability, validity and sensivity. British Journal of Psychiatry, 133, 429-435.
38. Hamilton, M. (1960). A Rating Scale for Depression. *Journal of Neurology, Neurosurgery and Psychiatry, 23*, 56-62.
39. Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology, 6,* 278-296.
40. American Psychiatric Association. (1987). Diagnostic and Statistical Manual of Mental Disorders, Revised 3 rd (DSM-III-R). APA: Washington, DC.
41. Phillips L. Case history data and prognosis in schizophrenia. J Nervous and Mental Disease. 1953;117:515-525.
42. Strauss JS, Carpenter WT Jr. The prediction outcome in schizophrenia. II. Relationships between predictor and outcome variables: a report from the WHO International Pilot Studt of Schizophrenia. *Archieves of General Psichiatry.* 1974);31*:*37-42.

43. Kay SR, Fiszbein, Opler AL. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophrenia Bulletin.* 1987;13:261-76.
44. Ventura J, Nuechterlein KH, Subotnik KL et al. Symptom dimensions in recent-onset schizophrenia and mania: a principal components análisis of the 24-item Brief Psychiatric Rating Scale. *Psychiatry Research.* 2000*;*97:129-135.
45. Strakowski SM, Keck PE, McElroy SL, West SA, Sax KW, Hawkins JM, Kmetz GF, Upadhyaya VH, Tugrul KC, Bourne ML. Twelve-month outcome after a first hospitalization for affective psychosis. *Archieves of General Psychiatry.* 1998;55(1):49-55.
46. Sands JR, Harrow M. Depression during the longitudinal course of schizophrenia. *Schizophrenia Bulletin.* 1999;25(1):157-171.
47. Siegel SJ, Irani F, Brensinger CM. Prognostic variables at intake and long-term level of functioning in schizophrenia. *American Journal of Psychiatry.* 2006;*163(3):*433-441.
48. Harrow M, Goldberg JF, Grossman LS, Meltzer HY. Outcome in manic disorders. *Archieves of General Psychiatry.* 1990;47: 665-671.
49. Wieselgren IM, Lindström E, Lindström LH. Symptoms at index admission as predictor for 1-5 year outcome in schizophrenia. *Acta Psychiatrica Scandinavica.* 1996; 94(5):311-319.
50. M. Arrasate, I. Gonzalez-Ortega, S. Alberich, M. Gutierrez, M. Martinez-Cengotitabengoa, F. Mosquera, N. Cruz, M.A. Gonzalez-Torres, C. Henry, A. González-Pinto. Affective dimensions as a diagnostic tool for bipolar disorder in first psychotic episodes. *European Psychiatry*. 2014; 29: 424-430.
51. Robinson DG, Woerner MG, Alvir JMJ, Bilder R, Goldman R, Geisler S, Koreen A, Sheitman B, Chakos M, Mayerhoff D, Lieberman JA. Predictors of relapse following response from a first episode of schizophrenia or schizoaffective disorder. *Archieves of General Psychiatry.* 1999;56: 241-247.
52. Johnstone EC, Macmillan JF, Frith CD, Benn DK, Crow TJ. Further investigation of the predictors of outcome following first schizophrenic episodes. *British Journal of Psychiatry.* 1990;157:182-189.
53. Sheperd M, Watt D, Fallon I, Smeeton N. The natural history of schizophrenia: a five-year follow-up study of outcome and prediction in a representative sample of schizophrenics. *Psychological Medicine.* 1989;monograph supplement 15:1-46.
54. Vázquez-Barquero JL, Cuesta MJ, Herrera S, Lastra I, Herrán A, Dunn G. Cantabria first-episode schizophrenia study: three-year follow-up. *British Journal of Psychiatry.* 1999;174:141-149.
55. Verdoux H, Liraud F, Gonzales B, Assens F, Abalan F, van Os J. Predictors and outcome characteristics associated with suicidal behaviour in early psychosis: a two-year follow-up first-admitted subjects. *Acta Psychiatrica Scandinavica.* 2001;103:347-354.
56. Eaton WW, Thara R, Federman E, Tien A. Remission and relapse in Schizophrenia: the Madras longitudinal study. *The Journal of Nervous and Mental Disease.* 1998;186(6):357-363.
57. Lindenmayer JP, Kay, SR. Affective impairment in young acute schizophrenics: its structure, course and prognostic significance. *Acta Psychiatrica Scandinavica.* 1986;175: 287-296.
58. Peralta V, Cuesta MJ, Farre C. Factor structure of symptoms in functional psychoses. *Biological Psychiatry.* 1997:*42*:806-815.

59. Gift TE, Strauss JS, Kokes RF, Harder DW, Ritzler BA. Schizophrenia: affect and outcome. *American Journal of Psychiatry.* 1980;137(5):580-585.
60. Sim K, Mahendran R. et al. Subjective quality of life in first episode schizophrenia spectrum disorders with comorbid depresión. *Psychiatry Research.* 2004;129:141-147.
61. Allardyce J, McCreadie RG, Morrison G, van Os J. Do symptoms dimensions or categorical diagnoses best discriminate between known risk factors for psychosis?. *Social Psychiatry and Psychiatric Epidemiology*. 2007b;42:429-437.
62. González-Pinto A, Aldama A, González C, Mosquera F, Arrasate M, Vieta E. Predictors of suicide in first-episode affective and nonaffective psychotic inpatients: five-year follow-up of patients from a catchment area in Vitoria. *Journal of Clinical Psychiatry.* 2007;68 (2): 24-27.
63. Robinson J, Harris MG, Harrigan SM, Henry LP, Farrely S, Prosser A, Schwartz O, Jackson H, McGorry PD. Suicide attempt in first-episode psychosis: a 7.4 year follow-up study. *Schizophrenia Research.* 2010;116:1-8.
64. Jablensky A, Sartorius N, et al. Schizophrenia: manifestations, incidente and course in different cultures. *Psychological Medicine Monograph supplement.* 1992;20:1-97.

SciForum
Mol2Net

Enhancement of Photovoltage Generation, Storage Capacity and Energy Conversion Efficiency of Photoelectrochemial Cell of Mixed Dye System: Role of Oxidized Multi-Walled Carbon Nanotubes

Poonam Bandyopadhyay [1,2], Ruma Basu [1,3], Sukhen Das [1,2,4], Papiya Nandy [1]

[1] Centre for Interdisciplinary Research and Education, 404B Jodhpur Park, Kolkata 700068, India;
[2] Physics Department, Jadavpur University. Kolkata 700032, India
[3] Physics department, Jogamaya Devi College. Kolkata 700026, India
[4] Indian Institute of Engineering Science & technology, Shibpur, Howrah – 711103, India

Published: 4 December 2015

Abstract: In the crisis of fast diminishing fossil fuels, looking for alternative energy sources by utilizing solar energy has the highest research priority and engineered nanoparticles play a very important role here. Using a specially designed photoelectrochemical cell and a mixture of two dyes, Phenosaffranine (PSF) and Azure C (AZC) conjugated with oxidized multi-walled carbon nanotubes (OMWCNTs) we have been able to generate photovoltage of reasonably high magnitude (763 mV). The storage time is also quite high ~ 66 hrs. The photovoltage cycle was reproducible upon further illumination. Energy conversion efficiency (η%) of the cell has been calculated for the system (η% = 4.33). Spectral studies show that with addition of OMWCNTs to a fixed concentration of PSF and AZC solution, the absorbance increases throughout the spectral range. FTIR spectrum reveals that there is no chemical change in the mixed system indicating that only the dyes got adsorbed on the side walls on the OMWCNTs. From fluorescence vspectral study, it has been seen that while fluorescence intensity of PSF quenches with addition of AZC, which does not have any fluorescence of its own, the intensity regains with addition of OMWCNTs. This emphasizes the role of oxidized multi-walled carbon nanotubes in the performance of photoelectrochemical cell of mixed dye system.

.

.

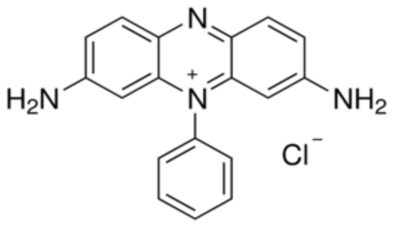
Keywords: mixed dye system; oxidized multi-walled carbon nanotubes; photovoltage; storage capacity; energy conversion efficiency

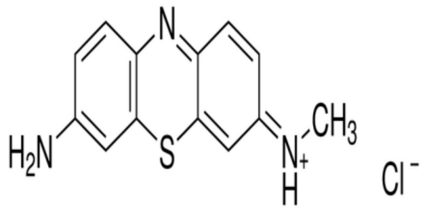
1. Introduction

The depletion of fossil fuels at an alarming rate and their hazardous combustion products pose a serious challenge to the scientists for developing alternative energy sources. In this endeavor utilization of solar energy has become the most promising one. Owing to the limitations of conventional silicon technology based photovoltaics, other cost effective options for harvesting solar energy more efficiently are being explored during the last couple of decades. Use of compound semiconductors caused high optical absorbance and hence good conversion efficiency [1-3]. The invent of nanoparticles, which possess the unique assets of altered chemical and physical properties with size and shape variations, has a strong impact in widening the of possibility in photovoltaic technology [4]. The important factor which limits the performance of conventional solar cells is the mismatch between incident solar spectrum and the spectral absorption properties of the material [5,6]. In the present study we have used photosensitive dyes conjugated with carbon nanotubes as the light energy harvester in photoelectrochemical (PEC) cell. Two dyes having absorption bands in two different spectral regions have been chosen in order to overcome band absorption limits of each dye.

Materials used:

Multiwall Carbon Nanotubes of diameter 2–3 nm were purchased from Arry International, Germany with 60% purity. The dyes Phenosaffranine [$C_{18}H_{15}ClN_3$] (M.W. 322.79) and Azure C [$C_{13}H_{12}ClN_3S$] (M.W. 277.77) were purchased from Sigma- Aldrich, India and were used after recrystallization. Cholesterol (E. Merck, Germany) was oxidized and recrystallized from n-octane (E. Merck, Germany). N-decane and iodine were purchased from E. Merck, Germany and before use iodine was purified by resublimation.





Phenosafranine

Azure C

Preparation of Oxidised Multi-wall carbon nanotubes (OMWCNTs):

For oxidation Multi-wall carbon nanotubes (MWCNTs) were at first refluxed with 2M nitric acid at 150°C for 18 hrs and then sonicated for 6 hrs using an ultrasonic bath sonicator (Model 229; Imeco Ultrasonic, India) in the same acid. Resulting material was collected by filtration and then washed with water and ethanol for several times until the pH of the solution became neutral i.e, 7. After drying the solution in hot air oven at 60°C for 8 hrs, we got oxidized MWCNTs (OMWCNTs) and those were well dispersed in water owing to presence of hydrophilic groups, such as carboxyl (-COOH), hydroxyl (-OH) and carbonyl groups (>C=O) along their side walls [7].

Preparation of mixed dye-OMWCNTs system:

The OMWCNTs were dispersed in water (concentration 1 mg mL^{-1}) and concentration of aqueous stock solution of both dyes was $1x10^{-4}$ M. The required concentrations of dye solutions for photovoltage (PV) studies were prepared after further dilution of the stock solutions. Individual dye solution was then mixed with OMWCNTs solution and sonicated in an ultrasonic bath for 4 hrs at room temperature and then different dye-OMWCNTs solutions were mixed together in a desired specific ratio by diffusion method.

Absorption spectra of dye-OMWCNT conjugates were recorded in a PERKIN ELMER Lambda 25 UV/VIS Spectrometer (Shelton, CT064844794).

Details of the experimental set-up:.

.

Table 1. The characteristics of the PEC cell.

Different parameters	Barrier: Planar lipid membrane	
	Sample Used	
	PSF & AZC	PSF & AZC adsorbed CNTs
Open circuit voltage (Voc) in mV	254	763
Short circuit current in µA	3.9	54.2
Time taken for Voc in hrs	0.8	1.5
Decay time	5.7	66.5
Storage duration in hrs	5	66
Fill factor	0.09	0.38
Energy conversion efficiency (η%)	0.005	4.33

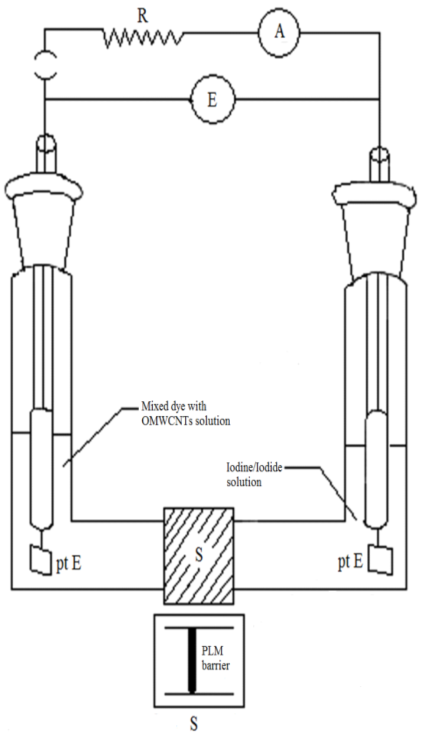


Figure 1. Schematic diagram of photoelectrochemical cell. S: PLM barrier, pt E - Electrode, E - electrometer, R - variable discrete resistance, A – ammeter.

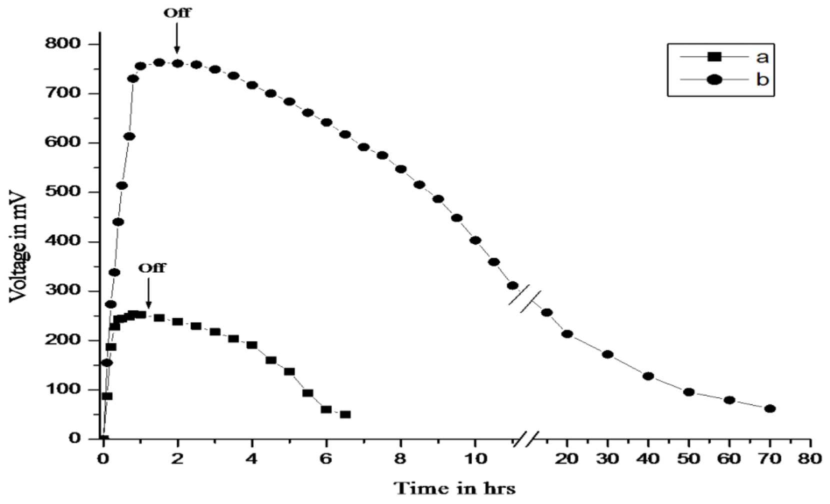


Fig 2. Growth and decay curve of photovoltage (V_{oc}) generation using planar lipid membrane as barrier for (a)PSF & AZC, (b)PSF & AZC adsorbed CNTs system.

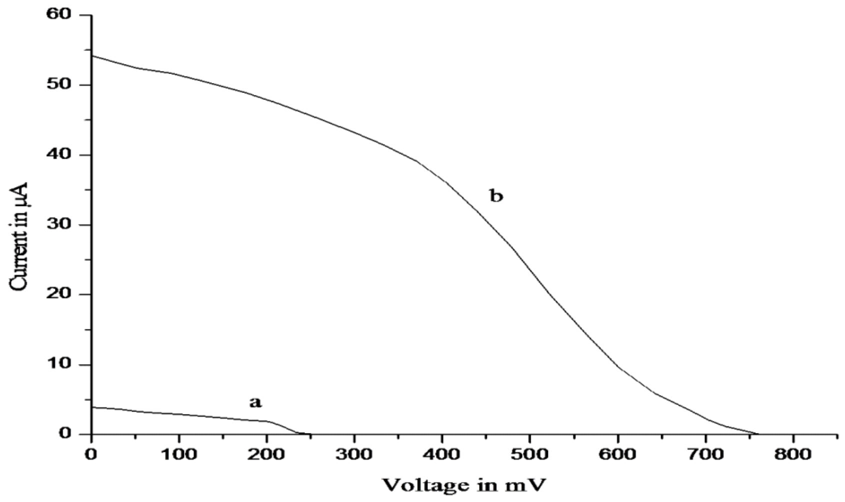


Fig 3. Current vs. voltage characteristic curves of the PEC cells for (a) PSF & AZC and (b) OMWCNTs adsorbed PSF & AZC.

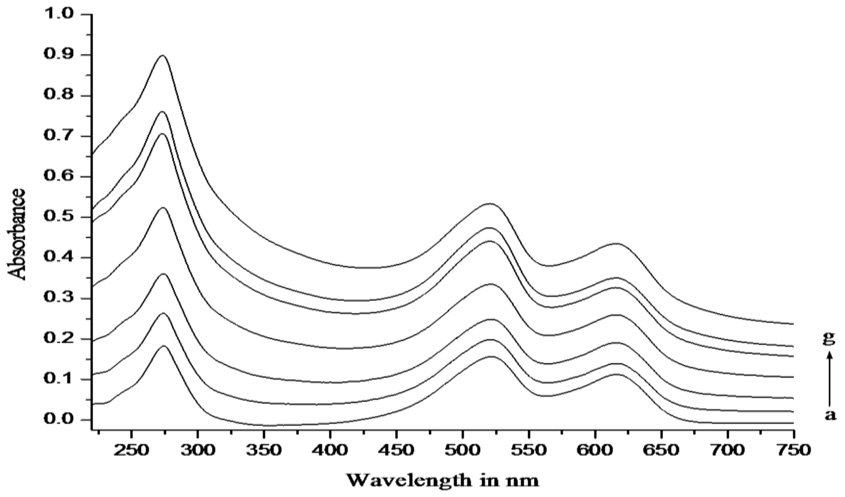


Fig 4. Absorption spectra of (a) PSF & AZC and (b to g) OMWCNTs adsorbed PSF & AZC systems. Concentration of PSF (a to g) 3 x 10^{-5}M, AZC (a to g) 3 x 10^{-5}M and OMWCNTs (x10^{-7} Kg/L) (a) 0.0, (b) 2.2, (c) 4.1, (c) 6.5, (d) 8.3, (e) 10.4, (f)11.1 and (g) 12.2.

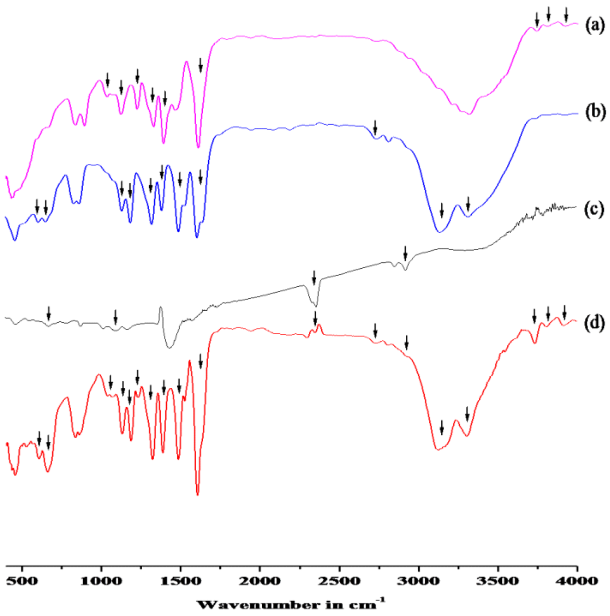


Fig 4. FITR spectra for (a) PSF, (b) AZC, (c) OMWCNTs, (d) Mixed system of PSF, AZC and OMWCNTs.

The PEC cell consisted of two L-shaped glass tubes (fig 1). A barrier was mounted on one tube which was fitted into the other by means of a standard joint. Here we have used a planar lipid membrane of oxidized cholesterol as barrier. One side of the barrier was bathed with Iodine Iodide (I^-/I^{3-}) solution and thve other side with aqueous solution in OMWCNT conjugated with PSF and AZC, in a 1:1 concentration ratio. A pair of platinum electrode was placed symmetrically across the barrier. A 60W lamp was used for illumination and the light intensity was measured by using a Luxmeter with photodetector (D & L Instrument, MS6610). Photovoltages and currents were measured by using Keithly digital multimeters (DM196). The energy conversion efficiency of the cell was calculated by using standard methods [8-10]. Keeping the cell under illumination, we recorded the photovoltage (Voc) generation.

Results and Discussion:

Upon illumination the photovoltage started increasing, attained a saturation value and remained constant till the illumination was there. When the light was switched off voltage started decreasing slowly (Fig 2). The storage time was quite high nearly 66 hrs. The photovoltage cycle was reproducible upon further illumination. From the current-voltage curve (Fig 3) we calculated the fill factor (FF) and energy conversion efficiency (η%) of the cells using equations (1) and (2).

FF = (V_{pp} x Ipp) / (V_{oc} x I_{sc}) ------------ (1)

η% = (Voc x Isc x FF x 100) / Incidient light power -------------- (2)

The values of open circuit voltage (Voc), short circuit current (Isc), storage duration, fill factor (FF), energy conversion efficiency (η%) are summarized in the table . It is evident that the presence of OMWCNTs causes radical increase in the values of V_{oc}, I_{sc}, storage time, FF and η%.

. Spectral studies showed that with addition of OMWCNTs to a fixed concentration of PSF & AZC solution, the absorbance of PSF & AZC increased throughout the spectral range without any shift in the absorption peaks. The results clearly indicated that the dye molecules got adsorbed on OMWCNT side walls without any change in chemical as well as photochemical properties. Increase in the absorbance was a definite indication to the enhancement of the absorption of the incident photons which resulted in the amplification of efficiency of the cells [7]. One dimensional carbon nanotubes have excellent electron-storage capacity (one electron for every 32 carbon atoms) and exhibit metallic conductivity similar to metals [11].We propose that the presence of OMWCNT in the conjugate resulted in further improving the efficiency by improved charge transfer, charge transport and efficient charge collection [12].

Moreover, the hydrphobic barrier offered by the planar lipid barrier planar lipid membrane due to its low dielectric constant hindered back recombination of the photodissociated charges and as a result of which the efficiency as well as the storage duration of the cell increased further.

. We got maximum PV for OMWCNTs conc. (10.4 x10^{-7} Kg/L).

For further support to our proposed idea, we will perform FTIR of all sample. From Infrared Spectra, it can be suggested that the basic structure of the dyes remain changed as all the peaks from two dyes and OMWCNTs are found in the same position as previous, for the mixed system FTIR spectrum.

Conclusions

Our studies showed that the light-harvesting performance of such PEC cells was much better in presence of dye-CNT mixed systems. As a particular dye absorbed only a certain fraction of the incident light spectrum, being transparent to the rest, we used two different dyes to overcome the band absorption limits of each dye. The presence of OMWCNT boosted the absorption of incident photons. The use of lipid membrane barrier played an important role in enhancing the storage duration of the photovoltage and the overall efficiency of the cell. These low cost cells were easy to assemble compared to the intricate device fabrication techniques of other conventional solar cells

Conflicts of Interest

State any potential conflicts of interest here or “The authors declare no conflict of interest”.

References and Notes

1. J. S. Ward, K. Ramanathan, F. S. Hasoon, T. J. Coutts, J. Keane, M. A. Contreras, T. Moriarty, R. Noufi, Prog. Photovoltaics 2002, 10, 41.
2. M. Afzaal, P. O’Brien, J. Mater. Chem. 2006, 16, 1597.
3. H. W. Schock, Appl. Surf. Sci. 1996, 92, 606.
4. P. V. Kamat . J. Phys. Chem. C 2007, 111, 2834.

5. W.G.J.H.M Van Sark , A. Meijerink, R.E.I. Schropp. In: Fthenakis V, editor. Third generation photovoltaics. Utrecht, The Netherlands: E-Publishing Inc.; 2012. p. 1-28.
6. T.P. Chou , Q. Zhang, B. Russo, G. Cao. J Nanophot, 2008, 2, 023511 (1-11).
7. A. Dey, R. Basu, S. Das, P. Nandy. Photochem. Photobiol. 2010, 86, 1000.
8. A. Mondal, R. Basu, S. Das, P. Nandy. J. Photochem. Photobiol. A 2010, 11, 143.
9. P. Bandyopadhyay, A. Dey, R. Basu, S. Das, P. Nandy. Curr. Appl. Phys. 2014, 1149.
10. R. Basu, S. Das, P. Nandy. J. Photochem. Photobiol. B 1993, 18, 155.
11. A. Kongkanand, P.V.Kamat, ACS Nano 2007, 1, 13.
12. Z. Peining, A.S. Nair, Y. Shengyuan, P. Shengjie, N.K. Elumalai, S. Ramakrishna, J. Photochem. Photobiol. A: Chem. 2012, 231, 9.

SciForum
Mol2Net

Application of Triturated Copper Nanoparticles as an Agent for Remediation of an Azo dye, Methyl Orange

Monalisa Chakraborty [1,2], Ruma Basu [1,3], Sukhen Das [1,2,4], Papiya Nandy [1]

[1] Centre for Interdisciplinary Research and Education, Kolkata 700068, India
[2] Physics Department, Jadavpur University. Kolkata 700032, India
[3] Physics department, Jogamaya Devi College. Kolkata 700026, India
[4] Indian Institute of Engineering Science & technology, Shibpur—711103, India

Published: 4 December 2015

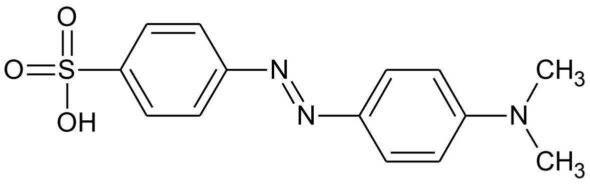
Abstract: Use of Azo dyes, having one or more azo bond (-N=N-), are more than 70% among all textile dyes used. The stability and the xenobiotic nature of reactive of these azo dyes make them recalcitrant and hence they are not totally degraded by conventional wastewater treatment processes that involve physical, chemicals or activated sludge methods. The dyes are therefore released into the environment, in the form of colored waste water. Color of the waste effluent is an important parameter for a long time. So the main criteria of removing the waste and degrade is to first remove the color of the effluent. Some common techniques which are used for degrading the chemically complex dye have several harmful side effects and they are not very cost effective and some require long retention time. So extensive research is going on to find a simple method which should obviously be a low cost process, will need less time to complete the process and will have minimum harmful side effects.

.

Keywords: Cuprum metallicum, methyl orange, UV-Vis spectroscopy, dye degradation, antimicrobial property

1. Introduction

Synthetically prepared colorful azo dyes are mostly used in textile, leather, pharmaceuticals, paper, and cosmetic industries. These dyes are marked as waste water pollutants. The water effluent which is coming from industrial waste, gets into the water bodies and cause extreme pollution.[1]Effluent containing compounds are chemically complexed and are not easily degradable for example with the help of bacteria or common chemical degradation process.



Methyl orange dye

Due to their complex structure they sometimes prevent the sunlight to penetrate inside the water bodies, because of which the life cycle of aquatic flora and fauna get hampered. Use of azo dyes, having one or more azo bond (-N=N-), are more than 70% among all textile dyes. **2.** Color of the waste effluent is an important parameter since long time. So the main criteria of removing the waste and degrade is to first remove the color of the effluent. The stability and their xenobiotic nature of reactive azo dyes makes them recalcitrant hence they are not totally degraded by conventional wastewater treatment processes that involve physical, chemicals or activated sludge methods 3. The dyes are therefore released into the environment, in the form of colored waste water. Some common techniques which are used for degrading the chemically complex dye, are precipitation, coagulation, adsorption, flocculation, flotation, electrokinetic coagulation, silica gel, membrane filtration, ion exchange, activated carbon, NaOCl, photochemical ,electrochemical destruction, fenton's reagent ,ozonation. However, every possible side effects obtained like some techniques produce sludge and harmful by-products. Another important factor is that these methods are not very cost effective and some requires long reaction time. **4 and 5** So scientist are trying to find a simple method which should obviously be a low cost process and will need less time to complete the process. **4 and 5.**

.

2. Results and Discussion

Methyl orange, 4-[4-(dimethylamino) phenylazo] benzenesulfonic acid, is an azo dye that forms orange crystals and is also used in laboratories and industries as a coloring agent having a negative impact on natural resources. **6**

Cuprum metallicum is a triturated medicine, which contains copper as a main constituent and by increasing the potency their size become in nanoform.

Conventional drugs available in the market, having antibacterial activity can lead to cause resistivity among different strains of bacteria. So research has been focusing on those materials which are new to the strains and do not have bad impact on the environment. So scientists are opting for alternative methods as antimicrobial agent. Copper nanoparticles are often used as an antibacterial agent7

Here cuprum metallicum can be utilized on different aspects with varied application. Different potencies of the drug (Cuprum metallicum) were used against the azo dye methyl orange and tried to degrade the effluent and function as a remediating material. Nanoparticles are very helpful as antimicrobial agent and the drug behaved as nanomedicine at higher potency **5** so we tried to find that whether at higher potency , the drug can show better result or not.

Parallelly they have been used here on pharmaceutical ground against two bacterial strains of gram positive and gram negative bacteria.

Material and methods:

Methyl orange dye was obtained from Merck, Bacterial strains i e, gram positive *Staphylococcus aureus* and gram negative

Escherichia.coli was from Microbial type culture collection and gene bank (MTCC) Institute of microbial technology, Chandigarh, India. Peptone, beef extract, yeast extract, sodium chloride, agar-agar was purchased from HiMedia. Cuprum metallicum of three different potencies were obtained as a gift from HAPCO. All the chemicals were used without purification as are all analytical grades. Throughout the experiment double distilled water was used.

Methods:

10^{-3}M Methyl orange dye was weighted and mixed with water so total solution is up to 300µl and 2ml of Cuprum metallicum of three different potencies (6C,30C,200C) were added individually, distilled water was added and make up to 4ml by volume.

Degradation of Methyl orange:

The kinetic study of the degradation of methyl orange (azo dye) with respect to time was thoroughly studied with the help of UV-Vis spectrophotometer .The time dependant degradation of the dye was investigated through change in the color and the absorbance of the dye. Just after addition of water with dye, the absorbance was measured and the sample was named as sample I. Then sample containing dye was placed in dark condition which was named as sample II. Then sample named as sample III which contained dye and water was placed under light condition only. Sample IV, V, VI contained dye and water .Then cuprum metallicum of 6C, 30C, 200C was added to IV, V, and VI and placed under exposure of visible light source. Finally absorbance was recorded at 464nm. And comparative study was carried out and checked whether cuprum metallicum would help in degradation process if the answer is yes then next query was that is 200C potency of the drug showed better result than 6C and 30C.

Antibacterial activity:

The second application was Cuprum metallicum of three different potencies were used against gram positive *S.aureus* sub species aureus (MTCC no.96) and gram negative *E.coli* DH5 alpha(MTCC no.1652) bacterial strain . The most common method of spread plate technique was used and both the strains were individually treated with Cuprum metallicum of three potencies (6C, 30C, 200C).Bacterial strains were grown in recommended liquid media which contains peptone (0.5%) beef extract (0.1%), yeast extract (0.2%), Nacl (0.5%) and distilled water.10µl of subcultured bacterial strains were inoculated in each test tube along with drug of different potencies. After overnight growth with antibacterial agent they were plated next day in an agar plate (Liquid media with 1.8% agar-agar) and colonies were counted with reference to control plate where no antibacterial agents were given..

.

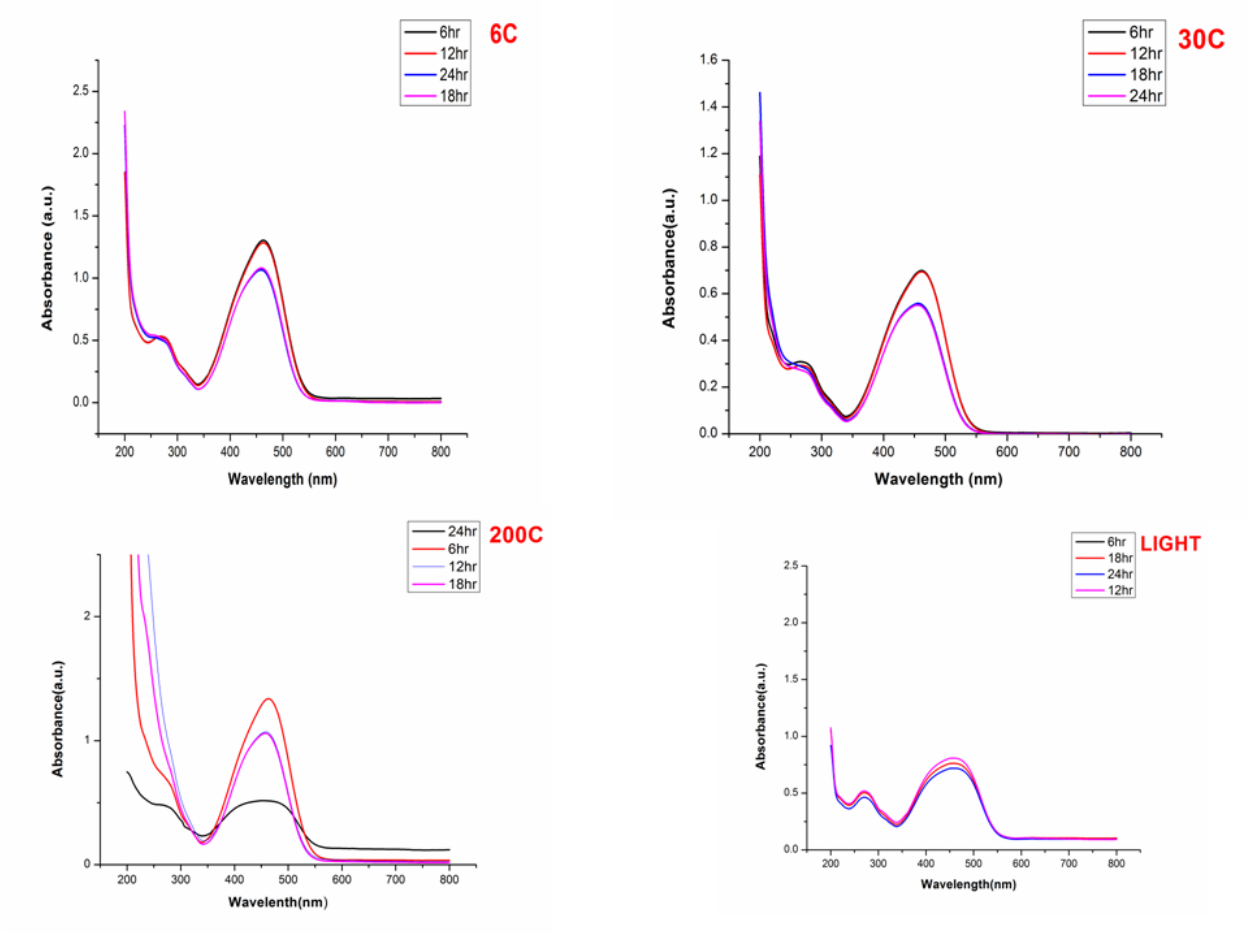


Figure 2. UV- Vis spectroscopic absorbance of dye treated with 6C, 30C, 200C and light.

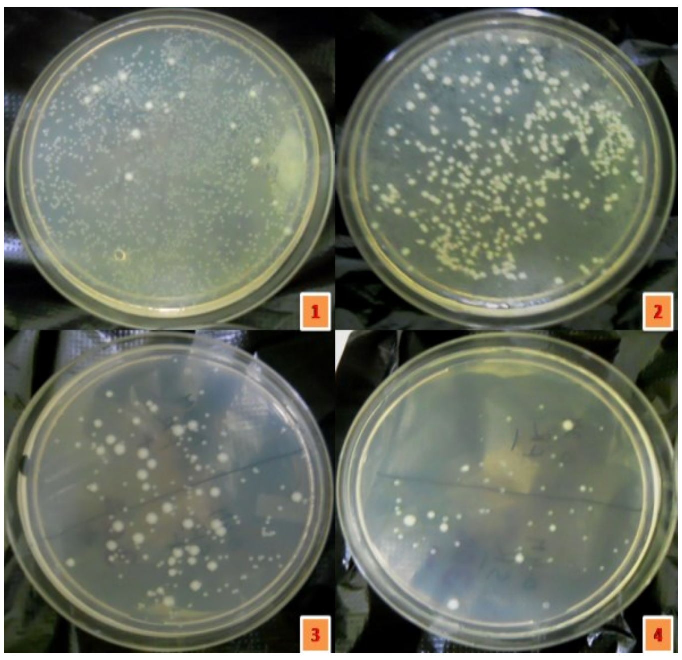


Figure 3. Antibacterial activity of cuprum metallicum for Gram negative bacteria.(*E.coli*).

Result and discussion:

Methyl orange dye treated at various conditions. Sample I showed image just after addition of dye in water. Sample II was the image taken after it was kept at dark condition for 24 hrs. Image named sample III was taken after it was placed at light condition only for 24 hrs. Sample IV, V, VI was the image of dye treated with cuprum metallicum 6C, 30C 200C respectively. The visual observation was clearly explained that color intensity of dye when treated with 200C was maximum detoriated. It means 200C work most effectively than other samples. In figure 2 we could see UV-Vis absorbance of dye individually treated with cuprum metallicum at a particular interval of time which was around 6 hrs gapping and it was for 24 hrs. Absorbance was taken at 464 nm 8The graph named 6C,30C,200C was the graph where four absorbance of the sample contained 6C,30C,200C cuprum metallicum along with dye and also was kept under light condition were taken. Absorbance was compared with respect to time. The graph named light was the absorbance of dye when kept at light condition only. The result was when dye treated with 6C+light, and then with respect to time the absorbance was decreased at 24 hr. It means when absorbance decreased that means chemical complex also degraded with time. The degradation was maximum observed in case of 200C when compared with initial time (6hr).So it can be proved that 200C showed a very impressive result.

In figure 3 we could see four images named 1,2,3,4. Image 1 showed growth of bacteria when nothing was treated. In Image 2, 3, 4 plates were treated with 6C,30C, 200C drug of cuprum metallicum where bacterial growth was maximum restricted in case of 200C treated plate. It means 200C could easily kill bacteria of gram negative strain.

The above experiment showed a quite positive result regarding the remediation and detoxification of azo dye also the nanomedicine cuprum metallicum worked as a potent source against gram strains. According to the previous reports it has already been proved by different characterization methods that higher potencies of the drug particle act as nanoparticles as size of the particles were reduced with dilution.9 So from the above experiment It has been found that three different potencies of cuprum metallicum were able to degrade the azo dye as with respect to time more the dye degrade more the colour intensity get lowered which was further confirmed by taking the absorbance of the dye. Previous reports were their where visible light can help to degrade dye samples through photo degradation. 10 Here it was observed that Sample VI showed maximum decrease in color intensity. Copper has always been a good and stable photo catalyst which can work better in presence of visible light.11 It is because of the excitation of surface Plasmon resonance(SPR) which is actually oscillation of charge density that promote at the interface between metal and dielectric medium.12 Here cuprum metallicum of three different potencies (6C, 30C, 200C) were used against two different strains of bacteria (Gram positive and gram negative).The best result was obtained against gram negative bacteria ie, *E.coli* and among different potencies of the drug, 200C of cuprum metallicum showed the best result. They can able to kill maximum no. of bacteria and few numbers of colonies were observed in the agar plate. The more the potency of the drug were used, more the antibacterial activity was shown by the drug. That means 200C showed maximum antibacterial effect and 6C showed minimum effect as compared to control where no drugs were used for treatment. The reason behind this result was copper nanoparticles is the prime component of cuprum metallicum and copper nanoparticles has multi toxicological effect against gram negative bacteria. They can generate reactive oxygen species (ROS) which cause DNA degradation, lipid peroxidation, protein oxidation in *E. coli* cells. 13

Nanoparticle can attached to the membrane of bacterial cell by electrostatic interaction and disrupt the integrity of the membrane of the cell. At higher potency means in case of 200C the individual particles have less chances of aggregation and so each individual particle can have its own separate surface area to interact with the bacterial membrane. More the surface area more chance of interaction which leads to more toxicity. So that is why higher potency of nanomedicine showed more antibacterial activity and they simply destroy bacterial cells.14

.

References and Notes

1. Yang-Hsin Shih , Chih-Ping Tso and Li-Yuan Tung ,2010, Rapid Degradation of Methyl Orange with nanoscale zerovalent iron particle, Journal of Environmental Engineering and Management, 20(3), pp-137-143.
2. A.Tripathi, S.K. Srivastava ,2011, Ecofriendly Treatment of Azo Dyes: Biodecolorization using Bacterial Strains, International Journal of Bioscience, Biochemistry and Bioinformatics, 1(1).pp-37-40.
3. Water Reuse: Conventional waste water treatment process. URL http://www.sheffy6marketing.com/index.php?page=test-child-page.
4. 4. Joshni. T. Chacko , Kalidass Subramaniam ,2011, Enzymatic Degradation of Azo Dyes – A Review, Internatinal journal of environmental sciences. 1(6),
5. 5.- M.Sudha, A.Saranya , G. Selvakumar and N. Sivakumar ,2014, Microbial degradation of Azo Dyes: A review, Internatinal journal of current research in microbial applied sciences. 3(2),pp-670-690
6. A.Shyamala , J.Hemapriya , Kayeen Vadakkan and S.Vijayanand ,2014, Bioremediation of Methyl Orange, a synthetic textile azo dye by a halotolerant bacterial strain, Internatinal journal of current research and academic review. 2(8),pp-373-381
7. Maqusood Ahamed, Hisham A. Alhadlaq, M. A. Majeed Khan, Ponmurugan Karuppiah, and Naif A. Al-Dhabi, 2014, Synthesis, Characterization, and Antimicrobial Activity of Copper Oxide Nanoparticles Journal of Nanomaterial. 2014, Article ID 637858, 4 pages
8. Henam Sylvia Devi, Thiyam David Singh, 2014, Synthesis of copper oxide nanoparticles by a novel method and its application in the degradation of methyl orange, Advance in Electronic and Electric Engineering, 4(1) pp-83-88
9. Subhajit Ghosh, Monalisa Chakraborty, Sukhen Das, Ruma Basu , Papiya Nandy. 2014,Effect of Different Potencies of Nanomedicine Cuprum metallicum on Membrane Fluidity – a Biophysical Study, American journal of Homeopathic medicine. 107(4) pp- 161-169.
10. R . Saravanan, Vinod kr Gupta, Edgar Mosquera, F Gracia, V Narayanam, A Stephen. Visible light induced degradation of methyl orange using beta –AgO .333 V_2O_5 nanorod catalysts by facile thermal decomposition method. Journal of Saudi Chemical Society.9(5) ,pp- 521-527.

11. Hongyan Liu, Tingting Wang, Heping Zeng, 2015, Copper nanoparticles: CuNPs for efficient photocatalytic hydrogen evolution. Particle and particle system characterization ,32(9) pp-857
12. P.Kumar, M.Govindaraju, S. Senthamilselvi, K. Premkumar, 2013 ,Photocatalytic degradation of methyl orange dye using silver (Ag) nanoparticle synthesized from *Ulva lactuca,* Colloids and surface B:Biointerface, 103 (1) pp-658-661
13. Arijit Kumar Chatterjee, Ruchira Chakraborty and Tarakdas Basu,2014,) Mechanism of antibacterial activity of copper nanoparticles, Nanotechnology. 25, 135101, pp-12-14.
14. Mohammad J. Hajipour[1],Ali Akbar Ashkarran,Dorleta Jimenez de Aberasturi,Idoia Ruiz de Larramendi,Teofilo Rojo,Vahid Serpooshan,Wolfgang J parak,Morteza Mahmoudi. 2012, Antibacterial properties of nanoparticles, Trends in Biotechnology, 30(10) pp-499-511.

SciForum
Mol2Net

Cobalt Alumino Silicate Ceramic(CASC) Nanocomposite, a Material with Moderately High Dielectric Constant and Low Tangent Loss at a Critical Concentration in High Frequency Range

Biplab Kumar Paul [1], Smarajit Manna [2], Debasis Roy [1], Papiya Nandy [1,2] and Sukhen Das [1,5,*]

[1] Department of Physics, Jadavpur University. Kolkata-700 032, India
[2] Jagadis Bose National Science Talent Search, Kolkata-700 107, India
[3] Central Glass and Ceramic Research Institute, Kolkata-700 032, India
[4] West Bengal State University, Kolkata, India
[5] Indian Institute of Engineering Science and Technology, Shibpur, India
* Author to whom correspondence should be addressed; E-Mail: sukhendasju@gmail.com.

Published: 4 December 2015

Abstract: Cobalt Alumino Silicate Ceramic (CASC) **c**omposites, with different molar weight concentration (i.e. G_0=0, G_1=0.4, G_2=0.6, G_3=0.8, G_4=1.0, and G_5=1.2 (M.W.)) of cobaltus acetate are prepared via sol-gel route. XRD shows mullite and cobalt aluminate phase which is found to depend on the concentrations of Co^{+2} ions. Field emission scanning electron microscope (FESEM) images show for all samples of CASC nanocomposites after sintering at 1400ºC, nano sized cobalt aluminate grains are embedded in evenly spread mullite grains; but in case of G_0, there are only evenly spread mullite grains. The study of dielectric property of the composite samples at room temperature shows that at all concentration (G_1, G_2, G_3, G_4 and G_5) the dielectric constant is higher than pure mullite (G_0) and there is a critical concentration of cobaltus acetate (G_3) where there is maximum enhancement of dielectric constant in the higher frequency range from 40 KHz to 2 MHz. The dielectric constant varies from 44.77 to 37.75 for G_3 and from 29.8 to 24.12 for G_0 respectively. The tangent loss of composite with G_3 has the lowest value than that of other concentrations including pure mullite in the frequency range 40 KHz to 2MHz. Due to high dielectric constant and low tangent loss, the composite with specific concentration and in the high frequency range has great importance as an electronic material.

.

Keywords: Mullite; dielectric constant; tangent loss; electronic material; sensor; capacitor

1. Introduction

During past decades Alumino silicate ceramic composite was used as an advanced structural ceramics and now it is well known as a promising electronic material with some great properties such as high melting point, anti-erosion, better chemical stability to oxidation as well as good resistance to most chemical attack, great coercivity, good electrical resistance, good mechanical strength, low thermal expansion coefficient, very good high temperature strength and also a low cost material [1-4].

So this composite have been increasingly gaining importance for high frequency circuit packaging and electronic substrate application due to their low dielectric constant[1,5-7] since last few decades. But recently researchers have been trying to enhance their dielectric constant by making composites by doping metal ions so that it can maintain its physical properties as earlier and also enhance the electrical and electronic properties. These composites with metal ions have wide range of application in electrical and electronic industries[8,9]. So CASC composite is a leading candidate for high transmitting IR windows, electronic material, humidity sensors, protective coatings with pigments, electrical insulators, and turbine engine components etc[10-13].

Cobalt aluminate composite behaves like an outstanding dielectric material due to so cobalt aluminate mullite ceramic composites have been studied widely because of its potential applications such as in high charge storage multilayer ceramics capacitors (MLCC), transducer because of its high polarization, high permittivity and very low tangent loss and can be promising candidate in Microelectronic industries, interconnect technology[8,14-17]..

2. Experimental

Chemicals used in the preparation of precursor gels, $C_9H_{21}O_3Al_{11,}$ puriss (Spectrochem Pvt.Ltd., India.), and Si $(OC_2H_5)_4$,(MERCK, Germany 99.9%),), were simultaneously added to 0.5 M solution of Al $(NO_3)_3$, $9H_2O$, extra pure(MERCK, India,99.9%), in double distilled water. The molar ratio of $C_9H_{21}O_3Al$ and Al $(NO_3)_3$, $9H_2O$ is 7:2. In the resulting sol 0.4 (G_1), 0.6 (G_2), 0.8 (G_3), 1.0 (G_4) and 1.2 (G_5) molar weighted (CH_3COO) $_2Co$, $4H_2O$, extra pure (MERCK, India 99.9.%), salt were added to the mixture separately. The amount of given salts for the increasing doping concentration of cobaltus acetate in alumino silicate ceramic system is shown in Table 1.

After stirring the solution for 3 hours, gel formation was completed and was kept overnight at 60°C for ageing the solution. The gel was then dried at 120°C and after grinding. the samples were then pelletized in cylindrical pellet form of 1 mm thickness by pressing powder to 70 MPa in a 10 mm diameter stainless steel die using hydraulic press and sintered in air environment at 1400°C for 2 hours at the heating rate of 5°C/minute[18-20]. Sintering was performed in air environment only to maintain the neutrality of the sample. Silver paste was painted on both surfaces of the pellet and then dried using hot air flow for 10 to 15 minutes.

.

Table 1. Different amount of salts given to prepare CASC composites, with different molar weight concentration.

Salts	G_0	G_1	G_2	G_3	G_4	G_5
$Al(NO_3)_3$, $9H_2O$ (g)	3.75	3.75	3.75	3.75	3.75	3.75
$C_9H_{21}O_3Al_{11}$ (g)	7.30	7.30	7.30	7.30	7.30	7.30
$Si(OC_2H_5)_4$ (g)	3.40	3.40	3.40	3.40	3.40	3.40
$(CH_3COO)_2Co$, $4H_2O$ (g)	0.00	1.922	2.988	3.984	4.981	5.977

Table 2. Elementary distribution (norm.wt.%) for all CASC composites.

Elementary Distribution. (In norm.[wt.%])	G_0	G_1	G_2	G_3	G_4	G_5
Aluminum	42.56	43.34	42.81	30.87	32.57	19.12
Silicon	9.22	3.96	2.58	7.82	3.51	5.66
Cobalt	0.00	9.64	13.60	24.93	30.95	51.75
Oxygen	48.22	43.06	41.01	36.37	32.97	23.46

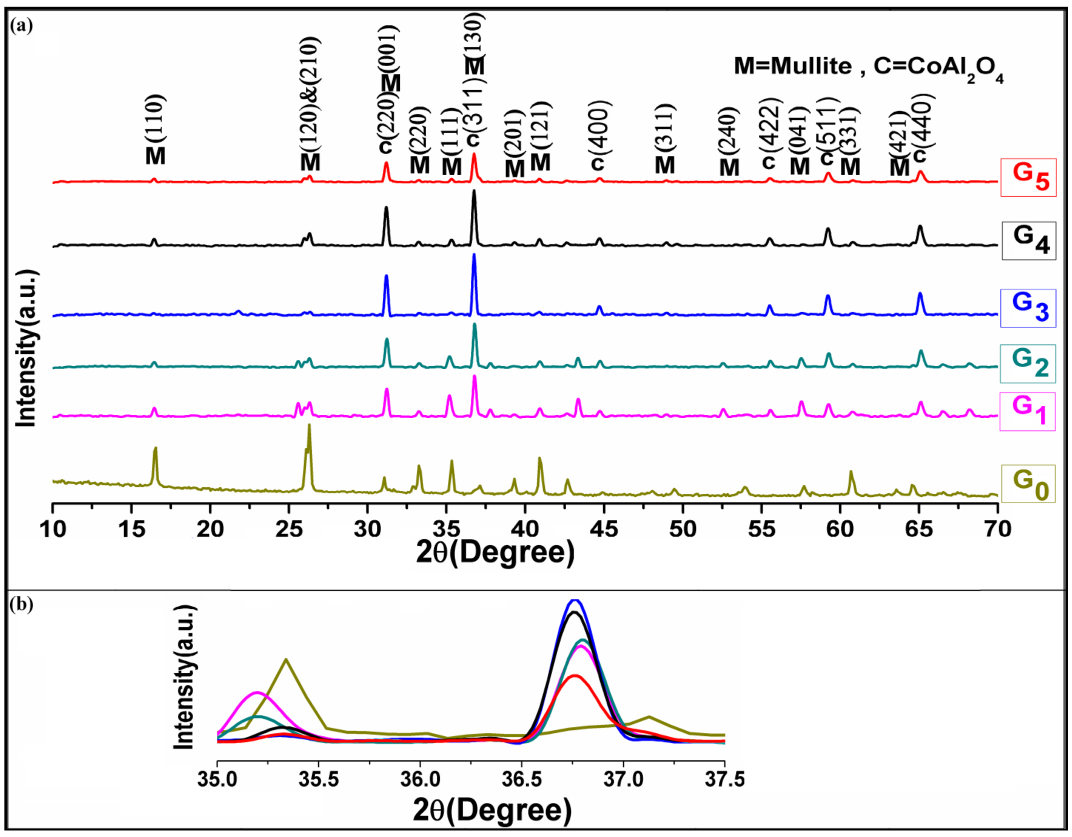


Fig.1 (a).X-ray diffraction pattern of pure mullite and all CASC composites containing increasing doping concentration of Cobaltus Acetate. (b).XRD patterns of composites with different concentration between 2θ from 35° to 37.5°.

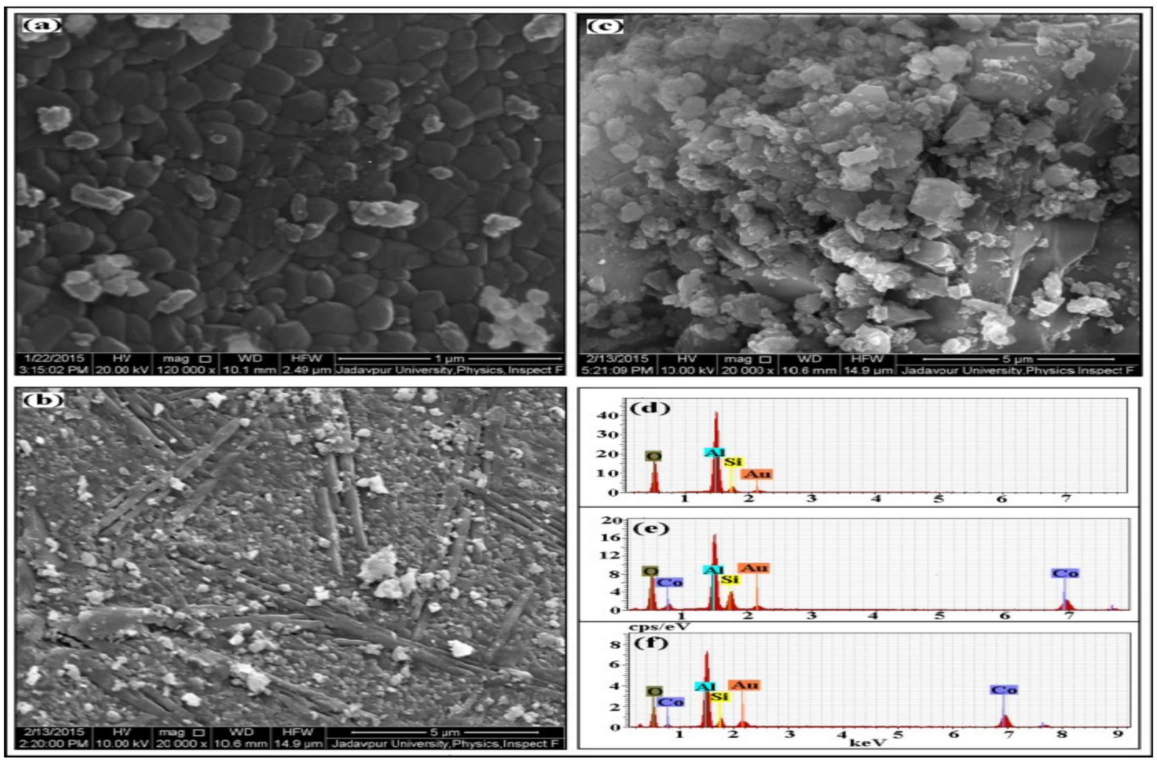

(a)
1/22/2015 3:15:02 PM HV 20.00 kV mag 120 000 x WD 10.1 mm HFW 2.49 µm 1 µm Jadavpur University,Physics,Inspect F
(c)
2/13/2015 5:21:09 PM HV 13.00 kV mag 20 000 x WD 10.6 mm HFW 14.9 µm 5 µm Jadavpur University,Physics,Inspect F
(b)
2/13/2015 2:20:00 PM HV 10.00 kV mag 20 000 x WD 10.6 mm HFW 14.9 µm 5 µm Jadavpur University,Physics,Inspect F
(d)
O
Al
Si
Au
(e)
O
Co
Al
Si
Au
Co
cps/eV
(f)
O
Co
Al
Si
Au
Co
keV

Fig. 2(a),(b) and (c). FESEM micrograph for G_0,G_3 and G_5 CASC composites respectively. Fig.2(d),(e)and(f). Energy-dispersive X-ray spectroscopy (EDX) image of G_0, G_3 and G_5 CASC composites respectively.

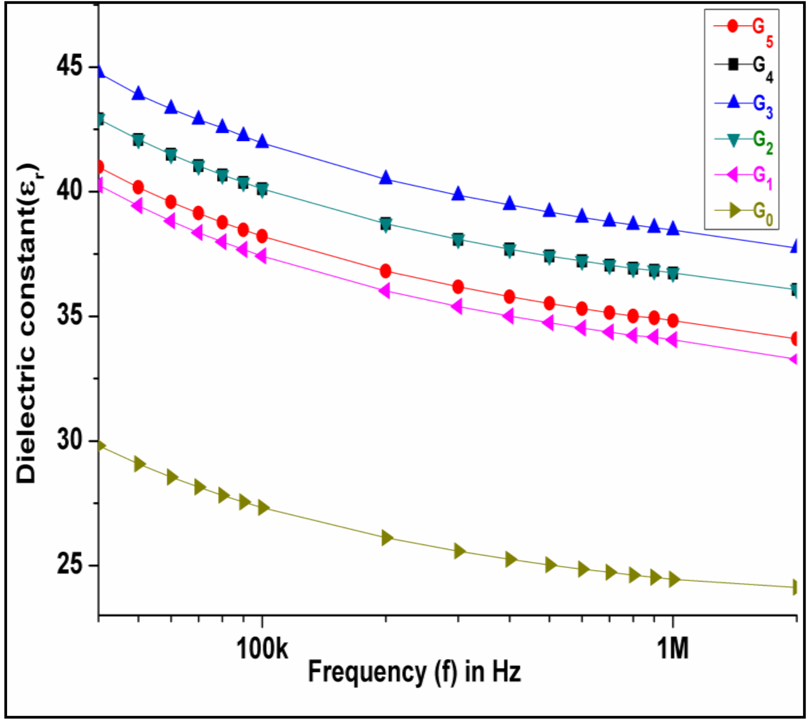


Fig.3. Frequency response Dielectric Constant (ε_r) behavior for all CASC composites.

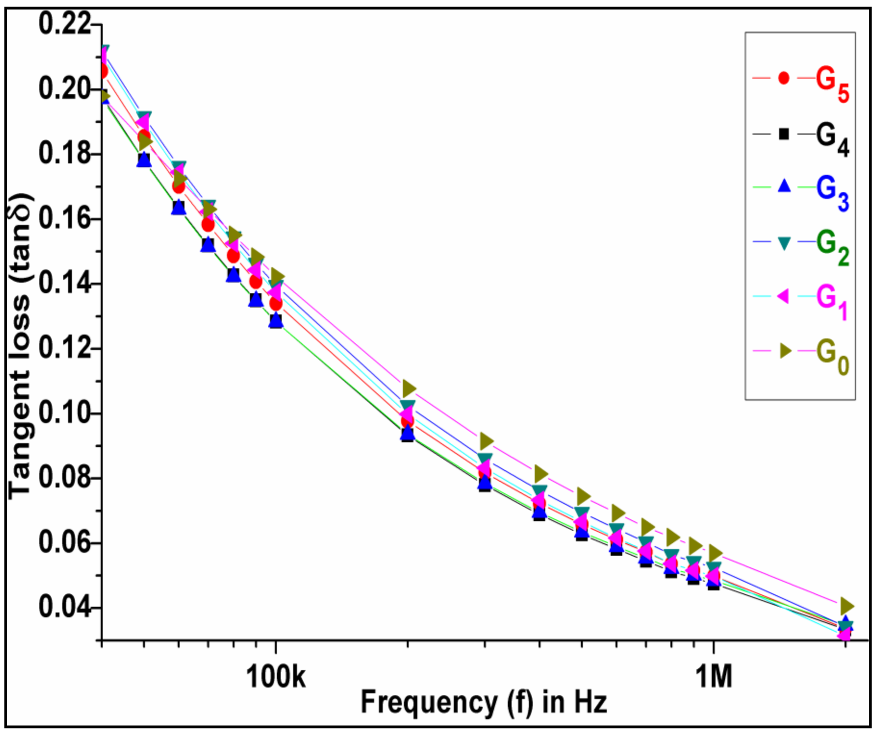


Fig.4. Frequency response Tangent loss (Tanδ) behavior for all CASC composites.

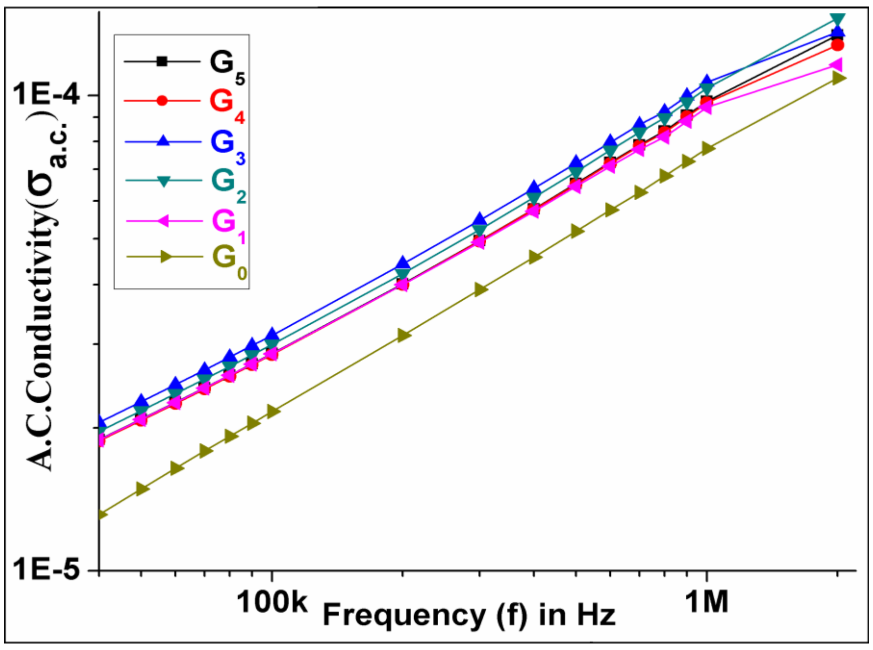


Fig.4. Frequency response of A.C. Conductivity ($\sigma_{a.c}$) for all CASC composites.

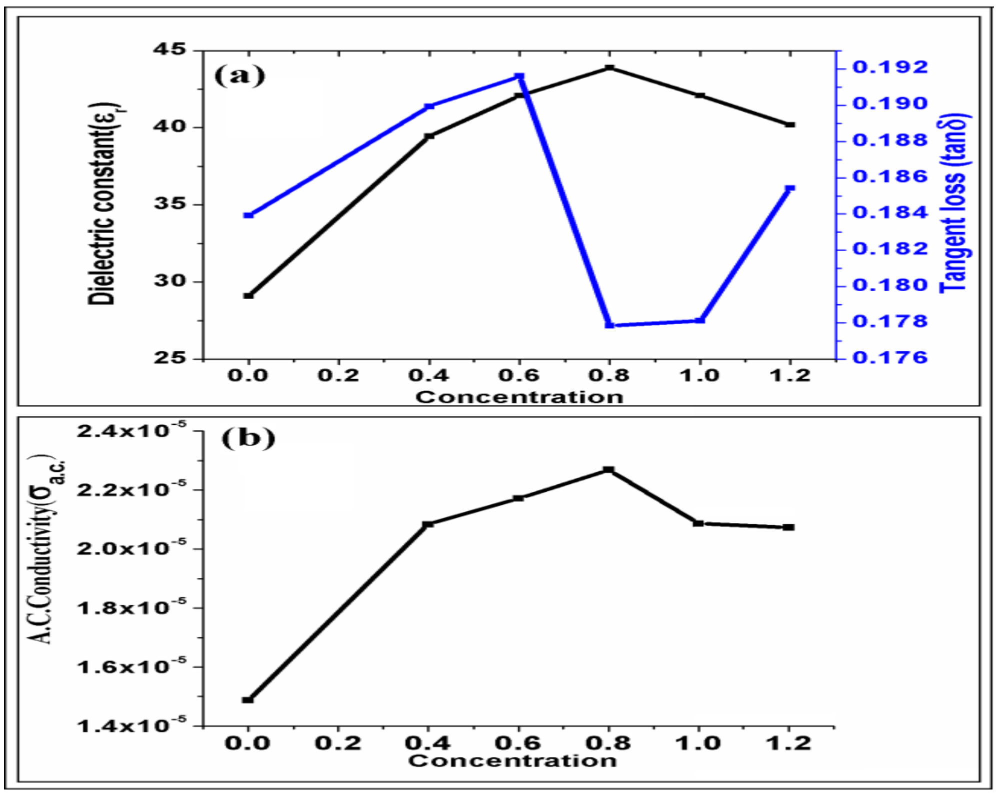


Fig.6.(a) Concentration dependent Dielectric Constant (ε_r) and corresponding Tangent loss (Tanδ) and Fig.6.(b) A.C. Conductivity ($\sigma_{a.c}$) behavior for all CASC composites at 50KHz.

The micrographs of mullite composites G_0 and G_3 and G_5 are shown in Figure 2(a) and 2(b) and 2(c) respectively. Micrograph of G_0 shows sphere shaped morphology of mullite grains with 100-300 nm size which is evenly distributed in the whole matrix. On the otherhand micrograph of G_3 shows crystalline plate like grains of cobalt aluminate of average size 3-6 μm embedded in uniformly spread numerous smaller mullite particles along with amorphous aggregates [9,18,19,24]. For further increased concentration of cobalt salt (i.e. for G_5) large amount of agglomerated particles along with bulk cobalt aluminate grains has been observed.

3. Used Instruments

The electrical properties of the pellets as a function of frequency were studied at room temperature by using LCR meter (Agilent 4294A precision impedance analyzer 40Hz-110MHz) in the frequency range of 40 KHz to 2 MHz. X-ray diffractometer (XRD) (Bruker D8 Advanced) was used to identify the crystalline structure and phases of the samples with different concentrations after sintering. The observations were made at angle between 10°-70°. The morphology and energy-dispersive X-ray spectroscopy (EDX) of the fracture surface of mortared dust was observed by using field emission scanning electron microscope (FESEM) (FEI Inspect F50).

4. Results and Discussion

4.1. XRD Analysis

Figure 1(a) and 1(b) shows the XRD spectrum of all CASC composite. Sample G_0 shows the prominent peaks of a single phase mullite that can be indexed with the standard JCPDS file number (JCPDS Card No-150776). The peaks of $CoAl_2O_4$ emerge when it is added to the mullite. Intensity of the $CoAl_2O_4$ peaks increase with increasing of its concentration in the composites upto G_3 and then decrease for higher concentration [peak C (440)]. Intensity of the individual mullite peaks decrease with increasing concentration of cobalt aluminate in the composites [peak M (111)]..

Intensity of the peaks of $CoAl_2O_4$ increases with increasing concentration of cobalt aluminate in the composites upto G_3 and then decrease for higher concentration [peak C (311),and C (220)]. Figure 1(b) illustrates the XRD patterns of composites with different concentration between 2θ of 35° -37.5° showing the strong mullite peaks M (111) and common peaks M (130) and C (311). Decrement of mullite phase with the increasing concentration of cobalt acetate is due to increment of cobalt aluminate phases formed by Co^{+2} ions. In earlier papers [18, 19, 21, 22, 24], it has been reported that mullite phase increases with the increment of metal ions upto certain concentration and then decreases, same characteristics is reported here. It has been observed here that due to the formation of cobalt the catalytic action of the metal decreases because of the developed strain within the composite as a result mullite phase decreases with increasing concentration of Co^{+2} ions. The phase variation in mullite was due to the change of concentration of metal ion and was due to John–Teller distortion. The observed difference in mullite formation of metal ions was due to weak ligand field. For weak ligand field the metal ions will be in high spin configuration. The unpaired electron in the t_{2g} orbital of Co^{2+} (d^7) will cause John– Teller distortion of CoO_6 octahedra in the gel matrix.

4.2. FESEM and EDX Analysis

Figure 2(d), 2(e) and 2(f) stands for elemental distribution (in normality weight %) analysis for the sample G_0 and G_3 and G_5 CASC composites respectively. Figure (2)d confirms that there is no cobalt content. Fig 2(e) and 2(f) confirms increasing concentration of cobalt content. As the given cobalt content increases peak intensity of cobalt content also increases. Here Au content arises from the gold coating unit.

Table 2 shows elementary distribution (norm.wt.%) for all CASC composites which confirms that with the increasing concentration of cobalt acetate normality weight % of cobalt also increases.

5. Electrical Properties Analysis

5.1. Frequency Dependent Analysis

5.1.1. Dielectric Constant Behavior

The variations of dielectric constant of all CASC composites with frequency are shown in Figure 3. From the figure it is clearly seen that throughout the frequency ranges 40 KHz to 2 MHz, dielectric constant continuously decreases with increasing frequency for all concentration of CASC composites which can be explained by the electron hopping model of Heikes and Johnston[23]. The effect of polarization is to reduce the field inside the medium. Therefore, the dielectric constant of a substance decreases substantially due to the limited dipole response. As the frequency is increases dipole response is limited and the dielectric constant attained a saturation tendency[24]. In this case the internal individual dipoles contribute the dielectric constant. At comparatively lower frequency range the dipoles can orient easily with electric field and so they can contribute improved polarization which is mainly responsible for the enhancement of the dielectric constant [23-26.]

5.1.2. Tangent Loss Behavior

Figure 4 shows the variation of tangent loss with frequency for G_0 to G_5 in the high frequency range 40 KHz to 2 MHz. From the figure it is clearly seen that throughout the whole frequency ranges, tangent loss continuously decreases with increasing frequency for all concentration of CASC composites. At comparatively lower frequency range the dipoles can orient easily with external electric field due to more relaxation time. This phenomenon is mainly responsible for intermolecular friction or vibration which contributes the higher tangent loss. As frequency increases, less polarization effect continues due to less relaxation time. So intermolecular friction or vibration diminishes which is responsible for decreasing tangent loss [23-26].

5.1.3. Frequency Dependent A.C. Conductivity ($\sigma_{a.c}$) Behavior

A.C. conductivity ($\sigma_{a.c}$) is calculated using the formula,

$\sigma_{a.c.} = 2\pi\ f \tan\delta\ \epsilon_r\ \epsilon_o$ Where,

$\sigma_{a.c}$ = AC conductivity,

f = frequency in Hz,

tanδ = tangent loss factor,

ϵ_r = dielectric constant of the material and

ϵ_o = vacuum permittivity respectively.

The variation of A.C conductivity with frequency is shown in figure 5. It shows A.C. conductivity increases with frequency. Increase of frequency increased a.c. conductivity by increasing hopping of conducting electrons present in CASC composite. At higher frequency range, rapid increase of conductivity with increasing frequency is referred to electronic polarization effect [23-26].

5.2. Concentration Dependent Dielectric Constant (ε_r) and Corresponding Tangent loss (Tanδ) and A.C. Conductivity ($\sigma_{a.c}$) Behabior.

From fig.6(a) it is clearly seen that throughout the frequency ranges 40 KHz to 2 MHz, dielectric constant has substantially higher value in case of G_1, G_2, G_3, G_4 and G_5 than G_0 and it also increases almost sharply with increasing concentration of cobalt acetate upto (0.8M.W.%) concentration i.e. (G_3), above which it decreases. This phenomenon can be explained by Maxwell–Wagner–Sillars (MWS) interfacial polarization effect which appears in heterogeneous medium consisting of different phases with different permittivity and conductivity due to accumulation of the charges at the interfaces. At low cobalt concentration (i.e.≤0.8M.W.%), the well crystalline nano mullite particles are well separated from each other with no such effective interaction between them. On the otherhand well crystalline $CoAl_2O_4$ grains is embedded in good and homogeneously distributed nano crystalline mullite matrix. So the number of nanoparticles and their interfacial area per unit volume increases while the interparticle distance decreases. This improves the average polarization associated with the particles and the coupling between neighboring grains, resulting in the significant enhancement of dielectric constant as well as significant decrement of tangent loss. This phenomenon observed upto G_3 i.e.(0.8M.W.%).

For further increment of cobalt content (i.e.>0.8M.W.%), the grain size of $CoAl_2O_4$ transformed into bulk form which is dissolved into agglomerated alumino silicate matrix. So the interfacial area per unit volume decreases while the interparticle distance decreases. This decreases the average polarization associated with the particles resulting in the further decrement of dielectric constant and ac conductivity as well as increment of tangent loss. This phenomenon is also clearly observed from their microstructures (FESEM) [Fig. 5(a),(b) and (c)].

It is also seen that the value of A.C conductivity is higher for all CASC composite than G_0. The increase of $\sigma_{a.c}$ is due to the increase of Co^{+2} ions and their mobility. The electrical conductivity fully depends on the mobility of the metal ions, so for the increasing number of metal ions electrical conductivity is also increased substantially [6, 24,26]. Presence of cobalt ions in the mullite matrix increases the mobility of ions, so a.c. conductivity increases with the increasing concentration of cobalt. It has been also observed upto G_3 (i.e.≤0.8M.W.%) concentration ac conductivity increases and for further increasing (i.e.>0.8M.W.%) concentration ac conductivity decreases. This may be due to more agglomerated glassy phase and amorphous matrix[(27-30)].

6. Conclusions

CASC nanocomposites with different concentrations of cobalt acetate have been synthesized by sol-gel technique and their phase evolution and dielectric properties have been investigated. The dielectric constant of all CASC nanocomposites is higher than alumino-silicate composite throughout the frequency range 40 KHz to 2 MHz and there is a critical concentration i.e. G_3 (0.8M.W.%) of metal ions (Co^{+2}) where the dielectric constant is maximum. The tangent loss of G_3 (0.8M.W.%) composites is lower than pure alumino-silicate composite in the frequency range 50 KHz and sample G_3 has minimum tangent loss in that frequency range. The A.C conductivity increases with frequency for all samples and is higher for metal doped mullite composite than pure alumino-silicate composite for presence of mobile metal ions in the composites. Thus pure alumino-silicate composite which has comparatively low dielectric constant can be modified into materials with high dielectric constant and low tangent loss by making a composite with cobalt acetate with molar weight concentration 0. 8M.W.%. These metal doped mullite composites may be used as dielectric material for the fabrication of high charge storing capacitors and also as ceramic capacitors and can be promising candidate for electronic industries.

Acknowledgments

We are grateful to UGC, Government of India, for financial assistance. We are also thankful to Dr. Abhijit Kar, Scientific Officer JBNSTS, Kolkata for fruitful discussion.

References and Notes

1. K. Maex, M.R. Baklanov, D. Shamiryan, F. Lacopi, S.H. Brongersma, Z.S. Yanovitskaya, Appl. Phys., 2003, 93, 8793.
2. H. Schneider, J. Schreuer, B. Hildmann,, Journal of the European Ceramic Society, 2008, 28,329.
3. M. I. Osendi & C. Baudin, Journal of the European Ceramic Society, 1996, 16,21l.
4. M.A. Camerucci, G. Urretavizcaya, M.S. Castro, A.L. Cavalieri, J. Eu. Ceram. Soc., 2001, 21,2917.
5. T. Homma, Materials Science and Engineering,1998, R23, 243.

6. T. Kurmara, M. Horiuchi, Y. Takeuchi, S.I. Wakabayashi, Electronic Components and Technology Conference, 1990,1, 68.
7. V. Viswabaskarana, F.D. Gnanama, M. Balasubramanian, App. Clay Sci., 2004, 25, 29.
8. M. M. S. Sanad ,M. M. Rashad ,E. A. Abdel-Aal ,M. F. El-Shahat ,K. Powers, J Mater Sci: Mater. Electron., 2014,25,2487.
9. B. K. Paul, K. Halder, D. Roy, B. Bagchi, A. Bhattacharya, S. Das, J Mater Sci: Mater Electron .DOI 10.1007/s10854-014-2291-6
10. M. Llusar, A.Fores, J.A. Badenes, J. Calbo, M.A. Tena, G. Monros, J. Eur. Ceram. Soc., 2007, 21,1121.
11. G. Carta, M. Casarin, N. El Habra, M. Natali, G. Rossetto, C. Sada, E .Tondello, P. Zanella, Electrochim. Acta., 2005, 50, 4592.
12. R. R. Turnma ,J.Am. Ceram. Soc., 1991,74,5, 895.
13. I.A. Aksay, D.M. Dabbs And M.Sarikaya, J. Am. Ceram. Soc., 1991,74, 2343.
14. F.Wen, Z. Xu, W. Xia, X. Wei and Z. Zhang, J. Adv.Dielect., 2013, 3, 1350010.
15. B.Bagchi, S.Das, A.Bhattacharya, R.Basu, and P.Nandy, J. Am. Ceram. Soc., 2009, 92,748.
16. S.P.Radhika, K. J. Sreeram, B.U. Nair,J. Adv.Ceram., 2012,1,301.
17. X. Hao, J.Adv.Dielect., 2013, 3,1330001.
18. B.Bagchi, S. Das, A. Bhattacharya, R. Basu, P. Nandy, J Sol-Gel Sci. Technol., 2010, 55,135.
19. D. Roy, B.Bagchi, A.Bhattacharya, S.Das, P.Nandy, J.Wu.Univ.Technol.-Mater. Sci. Ed ., 2012,27,836.
20. K.C.Song, Materials Letters, 1998, 35,290.
21. J.P. Tkalcec, B. G.Eta, S. Kurajica, J. Schmauch, American Mineralogist, 2007,92, 408.
22. W.Lv, Q.Qiu, F.Wang, S.Wei, B.Liu, Z.Luo, Ultrasonics Sonochemistry, 2010, 17,793.
23. R.R. Heikes and W.D. Johnston, Journal of Chemical Physics, 1957, 26, 582.
24. D.Roy, B.Bagchi, S.Das, P. Nandy, Materials Chemistry and Physics, 2013, 138, 375.
25. M.S. Sanad, M.M. Rashad, E.A. Abdel-Aal, M.F. El-Shahat, J. Eur. Ceram. Soc., 2012,32 ,4249.
26. D.Roy, B.Bagchi, S.Das, P.Nandy, J. Electroceram., 2012, 28,261.
27. V. K. Thakur, E. J. Tan, M.F. Linb and P. S. Lee, Polym. Chem.,2011,2,2000.
28. P. Thakur, A. Kool, B. Bagchi, S. Das and P. Nandy, PCCP, DOI: 10.1039/c4cp04006f.
29. C. C. Wang, J. F. Song, H. M. Bao, Q. D. Shen and C. Z. Yang, Adv. Funct. Mater., 2008, 18(8), 1299
30. P. Lunkenheimer, V. Bobnar, A. V. Pronin, A. I. Ritus, A. A. Volkov and A. Loidl, Phys. Rev. B: Condens. Mater. Phys., 2002, 66, 052105.

SciForum
Mol2Net

Evaluation of Computational Tools for Thermodynamics and Structural Analysis of Protein Stability upon Point Mutation Prediction

Alex D. Camargo [1],*, Adriano V. Werhli [1] and Karina S. Machado [1]

[1] Universidade Federal o Rio Grande, Av. Itália – KM 8; E-mails: alexcamargo@furg.br, werhli@furg.br, karina.machado@furg.br;

* Author to whom correspondence should be addressed; E-mail: alexcamargo@furg.br; Tel.: +55 53-3233-6623; Fax: +55 53-3233-6652.

Published: 4 December 2015

Abstract: In Bioinformatics, review of the state of the art about computational tools, including the interpretation of generated outputs and the restrictions of each software, contributes for choosing the best application to a specific problem. This way, an important research topic is the study of the impact of mutations in the treatment of complex diseases. Mutations have fundamental roles in evolution by introducing diversity into genomes, however, they can affect protein stability. Actually, researchers need accurate computational tools for prediction of how single point amino acid mutations affect the stability of a protein structure. Recent works show significant advances in predicting stability upon point mutation. This paper presents an evaluation of computational tools for thermodynamics and structural analysis of protein stability upon point mutation prediction. We choose to evaluate for thermodynamic analysis the software CUPSAT (Cologne University Protein Stability Analysis Tool) and mCSM (mutation Cutoff Scanning Matrix), and for structural analysis the software FoldX and Modeller. These software were chosen these software due to their popularity in this type of analyzes. In our proposed evaluation we verified the software outputs and evaluated the proximity to experimental results. As a case study we selected a set of 25 proteins extracted from: (i) MutaProt, which analyses pairs of PDB files whose members differ in one, or two, amino acids; (ii) ProTherm, database that contain experimentally determined thermodynamic parameters of protein stability. Each mutation in the datasets has attributes, as: PDB code, mutation, solvent accessibility, pH value, temperature and energy change (ΔΔG). A stability prediction model was successfully created, and the majority of the point mutations were predicted successfully having a high correlation and low standard error.

Keywords: point mutation; protein structure; computational biology

1. Introduction

Mutations have fundamental roles in evolution of organisms by introducing diversity into genomes [7]. Methods for protein structure prediction have advanced rapidly in recent years. There are a wide range of strategies for estimating protein energy, most of the methods are based on the statistical analysis of known protein structures [2]. The core functionality of these computational methods is an energy function that calculates the free energy of the system [8].

The understanding about the mutations that affect protein stability often resulting in diease is an important subject. Accurate prediction of point mutations effect on protein stability that can appear upon mutagenesis is fundamental when it is necessary to understand the structure-function relationship of a protein or in the cases where a new protein needs to be designed [4].

In addition to the natural variations in single mutations on proteins among organisms, bioinformaticians frequently introduce single amino acid residue replacements by site-directed mutagenesis in the laboratory to explore structural and functional features of proteins [4].

Several recent papers focused on testing sophisticated potential functions for conformational search and development of new scoring functions for side-chain modeling, reporting improved accuracy compared to earlier approaches [8, 10].

This paper aims at evaluating computational tools for thermodynamics and structural analysis of protein stability upon point mutation prediction. We choose to evaluate for thermodynamic analysis the softwares CUPSAT (Cologne University Protein Stability Analysis Tool) [6] and mCSM (mutation Cutoff Scanning Matrix) [7]. For structural analysis the software FoldX [9] and Modeller [3] were chosen. Our choice for these tools is because they are commonly used in these kind of analyzes.

As a case study we selected a set of 50 proteins extracted from: (i) MutaProt, which analyses pairs of PDB files whose members differ in one, or two, amino acids [2]; (ii) ProTherm, a database that contain experimentally determined thermodynamic parameters of protein stability [5].

This paper is organized as follows: the obtained results and discussion are described in Section 2. Section 3 introduces the protein stability predictors. Finally, Section 4 presents the conclusion.

2. Results and Discussion

A good computational biology method to predict stability changes upon mutation will help in designing new or altered proteins with specific levels of stability, enzymatic activity and binding to other molecules [8]. However, the number of false positives and false negatives returned by the programs, is generally substantial [4].

In this study, the prediction performances were evaluated based on accuracy measure. Accuracy is defined as a percentage of correctly identified mutations on the total number of mutations.

For evaluation of computational tools for thermodynamics analysis we use the difference in the calculated free energies ($\Delta\Delta G$) between the mutant and the wild-type, well with it was observed structural changes by RMSD (Root-Mean-Square Deviation) also mutant and the wild-type. RMSD values are considered as reliable indicators of variability when utilized to very similar proteins, like alternative conformations of the same protein [1].

For the set of selected proteins, none of the methods was able to accurately predict ΔΔGs for all mutations, as there is a significant deviation between experimental and calculated values. As seen in Table 1, often we are more interested to know whether a mutation is stabilizing or destabilizing, than to obtain the exact ΔΔG value.

In thermal experiments, it was observed that for the 25 mutations 88% were correctly predicted by mCSM to be either stabilizing or destabilizing. However, with a lower score, CUPSAT hit 72% of the predictions.

Table 2 shows the accuracy of the methods used for the three-dimensional structure prediction upon mutation point. The testing set consists of 25 pairs of known protein structures differing by a single mutation. The RMSD average between experimental and predicted structures values were 0,4Å for both methods, FoldX and Modeller.

Table 1. Accuracy of predicted change in ΔΔG upon point mutation.

Method	Accuracy
CUPSAT	72%
mCSM	88%

Table 2. Accuracy of predicted change in RMSD upon point mutation.

Method	Average	Standard deviation
FoldX	0,402Å	0.322
Modeller	0,404Å	0.323

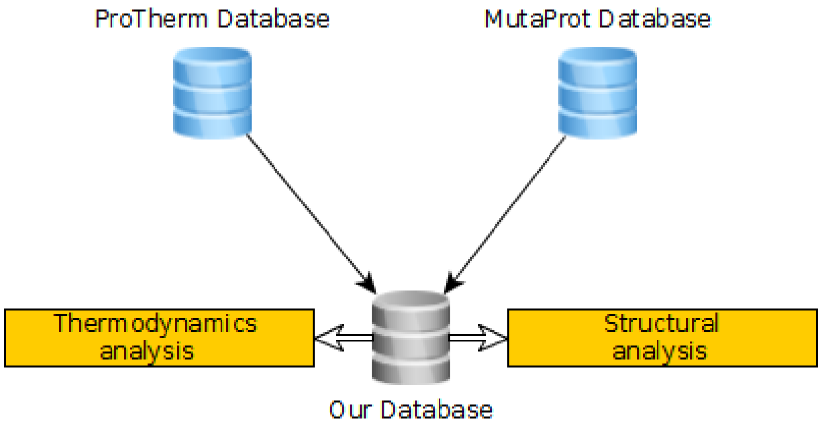


Figure 1. After collecting information from the used two datasets, we integrated the results into a non redundant dataset of protein stability change effects data.

3. Materials and Methods

In the current study, we chose two different methods that were previously reported as being able to predict the effects in protein stability (ΔΔG) upon mutation: CUPSAT [6] and mCSM [7]. Also, we tested two other methods for the modeling of point mutation in protein structures: FoldX [9] and Modeller [3].

CUPSAT (Cologne University Protein Stability Analysis Tool) is a web tool to analyse and predict protein stability changes upon single amino acid point mutations [6].

The approach, called mutation Cutoff Scanning Matrix (henceforth called mCSM), encodes distance patterns between atoms to represent protein residue environments [7].

FoldX is an empirical force field that was developed for evaluation of the effect of mutations on the stability, folding and dynamics of proteins and nucleic acids [9].

Modeller is a method to model point mutations in protein structures with two cycles of the conjugate gradients: molecular dynamics with simulated annealing, and conjugate gradients phases [3].

The effects of mutations on protein stabilities were predicted using the default parameters of the analyzed tools. As shown in Figure 1, for the purpose of this job a data set of mutations was compiled from two sources. The first, called MutaProt, was a list of single mutations that was published previously by Eyal et al. [2]. The second set of single mutations was obtained from the ProTherm database [5]. Some of the mutations in the datasets were listed several times. Therefore, in such situations, it was filtered to exclude any mutation that is listed more than once, or has incomplete information.

4. Conclusions

This paper evaluated the accuracy of common methods used to predict stability changes in proteins upon mutation. In our proposed evaluation we have verified the software outputs and analyzed the proximity to experimental results. In general there was good agreement between the methods in predicting the direction of change when compared with the experimental data.

The validation tests with mCSM showed that 90% of the mutations were correctly predicted for thermal stability. To evaluate the structural rearrangements upon mutations we calculated the RMSD of backbone movements upon single mutation. The results from both methods were identical and, in addition, Modeller is relatively faster than many of the FoldX method.

All the tested computational methods showed a correct trend in their predictions, but failed in providing the precise values. In summary, the current computational methods are clearly good enough for most of the tasks they are used for.

Acknowledgments

This work was supported by CAPES (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*). *Edital Biologia Computacional* - nº 051/2013 and CNPq process number 477462/2013-8.

Author Contributions

Guarantors of integrity of entire study, all authors.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Carugo, O., & Pongor, S. (2001). A normalized root- mean- spuare distance for comparing protein three- dimensional structures. Protein science, 10(7), 1470-1473.
2. Eyal, E., Najmanovich, R., Sobolev, V., & Edelman, M. (2001). MutaProt: a web interface for structural analysis of point mutations. Bioinformatics, 17(4), 381-382.
3. Feyfant, E., Sali, A., & Fiser, A. (2007). Modeling mutations in protein structures. Protein Science, 16(9), 2030-2041.
4. Khan, S., & Vihinen, M. (2010). Performance of protein stability predictors. Human mutation, 31(6), 675-684.
5. Kumar, M. S., Bava, K. A., Gromiha, M. M., Prabakaran, P., Kitajima, K., Uedaira, H., & Sarai, A. (2006). ProTherm and ProNIT: thermodynamic databases for proteins and protein–nucleic acid interactions. Nucleic Acids Research, 34(suppl 1), D204-D206.
6. Parthiban, V., Gromiha, M. M., & Schomburg, D. (2006). CUPSAT: prediction of protein stability upon point mutations. Nucleic acids research, 34(suppl 2), W239-W242.
7. Pires, D. E., Ascher, D. B., & Blundell, T. L. (2014). mCSM: predicting the effects of mutations in proteins using graph-based signatures. Bioinformatics, 30(3), 335-342.
8. Potapov, V., Cohen, M., & Schreiber, G. (2009). Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. Protein Engineering Design and Selection, 22(9), 553-560.
9. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., & Serrano, L. (2005). The FoldX web server: an online force field. Nucleic acids research, 33(suppl 2), W382-W388.
10. Tian, J., Wu, N., Chu, X., & Fan, Y. (2010). Predicting changes in protein thermostability brought about by single-or multi-site mutations. Bmc Bioinformatics, 11(1), 370.

SciForum
Mol2Net

An Insight to Segment Based Genetic Exchange in Influenza A virus: an *in silico* Study

Antara De [1,*], Ashesh Nandy [2]
[1]Centre for Interdisciplinary Research and Education, 404B Jodhpur Park, Kolkata 700068, India; E-Mail: antaradnet@gmail.com
[2]Centre for Interdisciplinary Research and Education, 404B Jodhpur Park, Kolkata 700068, India; Email-: anandy43@yahoo.com
***Author to whom correspondence should be addressed; Antara De E-Mail: antaradnet@gmail.com
Tel.: ++91-33-9038639153

Published: 4 December 2015

Abstract: Influenza virus is well recognized for high level of mutations that lead to development of new strains and subtypes, primarily through genetic shifts and drifts. Another mechanism of genetic change is through recombination, often observed in mammalian genes but controversial in viral genomes, which involves exchanges of short nucleotide sequences between two strains that coinfect the same cell. While evidence of such recombinations are rare to disputed in influenza genomes, we have observed that well-defined segments of influenza genes such as the hemagglutinin and neuraminidase have shown identical sequences between various strains that is best explained by segment exchange. Thus we had in our earlier study of the spread and proliferation of H5N1 bird flu observed that the neuraminidase with three segments – the transmembrane, stalk and body – shows evidence of exchange of one or other segments between different strains. Extending our work to the hemagglutinin of various subtypes, we noticed the same phenomena: Hemagglutinin has two segments, HA1 and HA2, where we found several instances where the segments seem to have been exchanged. Our analyses was based on RNA descriptors calculated in a 2D graphical representation scheme which have been proved to easily identify identical sequences. In this paper we discuss some of the details of this phenomenon in influenza genes which could be important in monitoring development of new highly pathogenic strains..
.

Keywords: Influenza; recombination; graphical representation

Mol2Net YouTube channel*: http://bit.do/mol2net-tube*

1. Introduction

Influenza A virus (IAV) has caused pandemic in human population since antiquity. It belongs to family Orthomyxiviridae and is a negative sense RNA virus with the genome divided into 8 segments which code for 11 proteins. The virus has been classified into different subtypes based on their cell-surface proteins hemagluttin (HA) and neuraminidase (NA). There are there are 18 HA and 11 NA subtypes identified [1]; however only few subtypes docking discrete combinations are found in nature [2]. These subtypes are further identified into different strains depending upon host type, geographical origin, strain number and year of isolation. Influenza exhibits remarkable degree of variability. Two well known mechanisms for the cause of variability are antigenic drift due to lack of proof-reading activity of RNA polymerase and antigenic shift by means of reassortment among the segmented genes of the genome [3]. There is also the question of whether recombination can cause variability, a phenomenon highly debated in negative sense RNA viruses [4]. Theoretically it can happen by non-homologous recombination between different genes or homologous recombination between same genes of different strains co-infecting the same cell. Our aim is to study the possibility of occurrence of homologous recombination in the HA and NA genes.

The choice of the two surface proteins, HA and NA, for our study is guided by the fact while HA is responsible for viral entry into the host cell, NA is responsible for exit of the progeny virions from the host cell. NA gene has three subunits consisting of transmembrane (TM), stalk (ST) and body (BD) (Fig 1); The HA gene consists of two subunits, HA1 and HA2 (Fig 2). Since recombination relates to exchange of whole sections of the sequence, we want to study whether homologous recombination can happen at the cleavage points of the segmented HA and NA genes. We hypothesize that during replication the RNA polymerase swaps its template from one strain to the other at the cleavage point of the subunits, thus producing unique recombinant having genetic segments of both strains undergoing recombination (Fig 3).

To determine whether recombination could have taken place by exchange of segments as we hypothesized, we analyzed the major human infecting influenza HA gene sequences, viz., the H1N1, H5N1, H3N2 and H7N9 subtypes over the period 2010 to 2014 in Asia, and the H5N1 bird flu NA from 1997-2009 reported previously [5]. We report here the results of our study that yielded several instances where sequence identities between segments of various strains could be interpreted as homologous recombination via segment exchange, albeit as a small fraction of subtypes tested, but implying enlarged possibility of evolution of new strains of such negative sense RNA viruses.

.

2. Results and Discussions

Our previous analyses of the neuraminidase sequences of the H5N1 bird flu epidemic of 1997-2009 had shown as an aside from the main thrust of the paper that there were worldwide evidences of duplications in part and whole of the sequences [5]. In particular, considering that the NA had three well-identified segments – transmembrane, stalk and body, there were evidences of sequences that had duplicates of one or two segments also. E.g., the transmembrane segment of A/treesparrow/Henan/4/2004(H5N1) was found to be identical to sequences from Hong Kong in 1999 and 2001

(A/Environment/HongKong/437-6/99(H5N1), A/Goose/HongKong/76.1/01 (H5N1)). The identity search among the rather large number of sequences was facilitated by computation of descriptors, g_R, of each segment whose equality implies sequence identity in the 2D graphical representation system as described in the Materials and Methods section.

This led to a number of sequence identities within each individual segment, which we had interpreted as being the result of RNA polymerase jump during replication in cases of co-infection in the same host cell, i.e. a homologous recombination. Although this was a preliminary study [5], the large number of such duplicated segmental sequences implied a new observation of homologous recombinations in negative sense viral genes.

These observations led us to replicate the analyses for possible such recombinations in hemagglutinin sequences of the Influenza genome. Our analysis of over a thousand hemagglutinin sequences using the 2D graphical representation descriptor, g_R, again revealed segmental duplication (more complete report in Ref [6]); though this time at a much lesser proportion than we had observed for the H5N1 neuraminidase. Of the 1200+ HA sequences analyzed, we found evidence for recombination in 73 instances of a daughter strain having its two segments from two parents under a strict protocol of considering only those daughter strains whose collection dates are marked later than the putative parents, and preferably from the same locality or country; Table 1 displays a short list of some selected sequences that display such connections indicated by the g_R values.. We expect such rigor if applied at the time would have reduced the number of recombination hits observed in the case of the neuraminidase counts.

However, some parent-daughter triplets have been identified with long time lag between their appearances, or the daughter sequence identified appears to have collection date before the parent. Based on previous studies that have shown that influenza strains can survive in the wild for years under suitable benign conditions depending on salinity, pH, temperature and other factors [7] it is possible that the apparent incongruent strain had taken part in the recombination in due course, but only discovered much later.

Segment exchange of the type we have hypothesized here do not fall into the “breakpoint” analysis favoured by researchers into these aspects. We note that such recombinations, if they are to happen, will take place during replication events when the RNA polymerase during its replication could jump from one template to another. For such jumps to take place accurately to create useful replicates, there has to be trigger points like the “breakpoints”. This we hypothesize to be the cleavage point between the HA1 and HA2 which is well conserved. Table 2 illustrates this motif by listing 10 bases, 5 on either side, around the cleavage junction. Our hypothesis is that since there are two segments on the HA, the RNA polymerase can jump from one template to another at the cleavage point and continue replicating till the end.

This should be slightly more difficult for the neuraminidase with three segments where replication with jumps from the transmembrane or the stalk will require the polymerase to jump back to the remainder of the first template. The fidelity of the transit points on the neuraminidase sequences are yet to be studied in detail.

We note too that not all segmental exchanges take place at the same frequency. The percentage of recombination observed amongst the total number of strains investigated in the neuraminidase genes varied significantly between segments. E.g., 2/3rds of all transmembranes showed high degree of similarity as shown in Table 4 of Ref [5]; taking a maximum of 1/3 of these being recombinants, that works out to 2/9ths of all H5N1 strains investigated, admittedly an upper bound. Similarly, it works out to <1/5th for the stalk region and <2% for the body segment. For the hemagglutinin, the recombinations varied by subtype: from 3.74% for H1 of H1N1 to 5.71% for H5 of H5N1, 7.35% of H3N2 and 7.87% of the H7N9 in our database.

It is noteworthy to watch the apparent disparity among four HA subtypes exhibiting different recombination frequency. However the phenomena can be explained by the fact that sometimes they have different evolutionary history as exemplified by H1 and H3 subtypes [8, 9] or a low pathogenic avian influenza virus (LPAIV) being suddenly mutated to a highly pathogenic avian influenza virus (HPAIV) as exemplified by H5 and H7 subtypes of H5N1 and H7N9 Influenza A virus which created bird flu and China Flu, 2013[10]. Often these HPAIV are adapted to human host; though possibility of human to human transmission remains debatable. More details can be found in our comprehensive report [Ref 6]. Interestingly, all valid recombinant possibilities noted above appears to have been restricted to parents and daughters from the same hemagglutinin subtype, i.e. there were no examples of mixed subtype marriages. It appears probable that the distinct antigenic sites in the different subtypes play a role to restrict the possible recombinations; the different base compositions, apart from the base distributions, thus may play a regulatory role.

Table 1. Few representative recombinants, identified on the basis of numerical characterization algorithm

Sub Type	Locus id	Description	g_R(HA1)	g_R(HA2)
H1N1	KM029055 (Parent 1)	A/swine/Hong Kong/NS4846/2011(H1N1)	66.18646	49.33498
	KM029103 (Parent2)	A/swine/Hong Kong/4902/2011(H1N1)	65.75885	49.06597
	KM029063 (Daughter)	A/swine/Hong Kong/NS4848/2011(H1N1)	66.18646	49.06597
H1N1	KM028335 (Parent 1)	A/swine/Guangxi/3614/2011(H1N1)	67.93622	49.80615
	KM028423 (Parent2)	A/swine/Guangxi/3880/2011(H1N1)	69.77442	48.80638
	KM028271 (Daughter)	A/swine/Guangxi/NS3248/2011(H1N1)	67.93622	48.80638
H3N2	KM069501 (Parent 1)	A/Singapore/C2010.307/2010(H3N2)	57.06303	36.58159
	JX437712 (Parent2)	A/Singapore/C2010.036V/2010(H3N2)	59.02475	37.64934
	JX437832 (Daughter)	A/Singapore/H2010.370C/2010(H3N2)	57.06303	37.64934

H3N2	JX437832 (Parent 1)	A/Singapore/H2010.370C/2010(H3N2)	57.06303	37.64934
	KM069503 (Parent2	A/Singapore/H2010.310/2010(H3N2)	56.53434	36.58159
	KM069501 (Daughter)	A/Singapore/C2010.307/2010(H3N2)	57.06303	36.58159
H5N1	KF369229 (Parent 1)	A/chicken/Cambodia/W0530391/2012(H5N1	61.16050	37.63436
	KF369214 (Parent2)	A/Cambodia/W0526301/2012(H5N1)	60.92035	37.15533
	KF369222 (Daughter)	A/chicken/Cambodia/W0530389/2012(H5N1)	61.16050	37.15533
H5N1	KF001474 (Parent 1)	A/civet cat/Cambodia/X0313306/2013(H5N1)	61.85977	36.95179
	KF001497 (Parent2)	A/duck/Cambodia/X0220302/2013(H5N1)	60.53024	39.06615
	KF001478 (Daughter)	A/civet cat/Cambodia/X0313307/2013(H5N1)	61.85977	39.06615
H7N9	CY147148 (Parent 1)	A/environment/Shanghai/S1438/2013(H7N9)	50.61318	44.19611
	CY147188 (Parent 2)	A/pigeon/Shanghai/S1423/2013(H7N9)	51.11281	43.71347
	CY147132 (Daughter)	A/environment/Shanghai/S1436/2013(H7N9)	50.61318	43.71347
H7N9	KF542876 (Parent 1)	A/chicken/Shanghai/017/2013(H7N9)	52.00389	43.08626
	KF667751 (Parent 2)	A/environment/Guangdong/30/2013(H7N9)	52.27365	43.67379
	KF667746 (Daughter)	A/environment/Guangdong/25/2013(H7N9)	52.00389	43.67379

Table 2: Conserved region at the cleavage point of HA1 and HA2 segments.

Locus ID	**Description**	**HA1 end**	**HA2 begin**
JF275925	A/swine/Nanchang/F9/2010(H1N1)	ct**aga**	**gg**cct
KJ955515	A/Ho Chi Minh/459.6/2010(H3N2)	ct**aga**	**gg**cat
KM276899	A/swine/Taiwan/NPUST0004/2013(H3N2)	at**aga**	**gg**cat
CY091837	A/Guangdong/322/2010(H3N2)	ct**aga**	**gg**cat
CY124183	A/Singapore/GP1147/2011(H3N2)	ct**aga**	**gg**cat
CY116636	A/Tbilisi/GNCDC0485/2012(H3N2)	ct**aga**	**gg**cat
CY124187	A/Singapore/GP1188/2011(H3N2)	ct**aga**	**gg**cat
AB569348	A/whooper swan/Mongolia/1/2010(H5N1)	aa**aga**	**gg**act

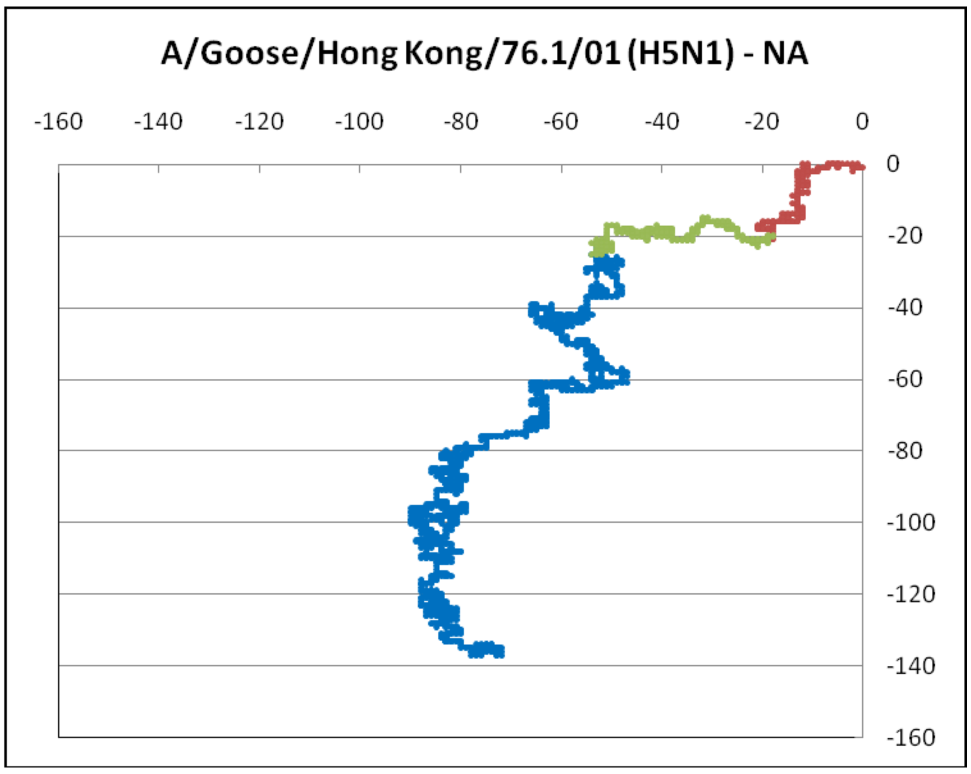


Figure 1: 2D graphical representation of a H5N1 strain neuraminidase sequence (Segments: Brown - Transmembrane; Green - Stalk; Blue - Body)

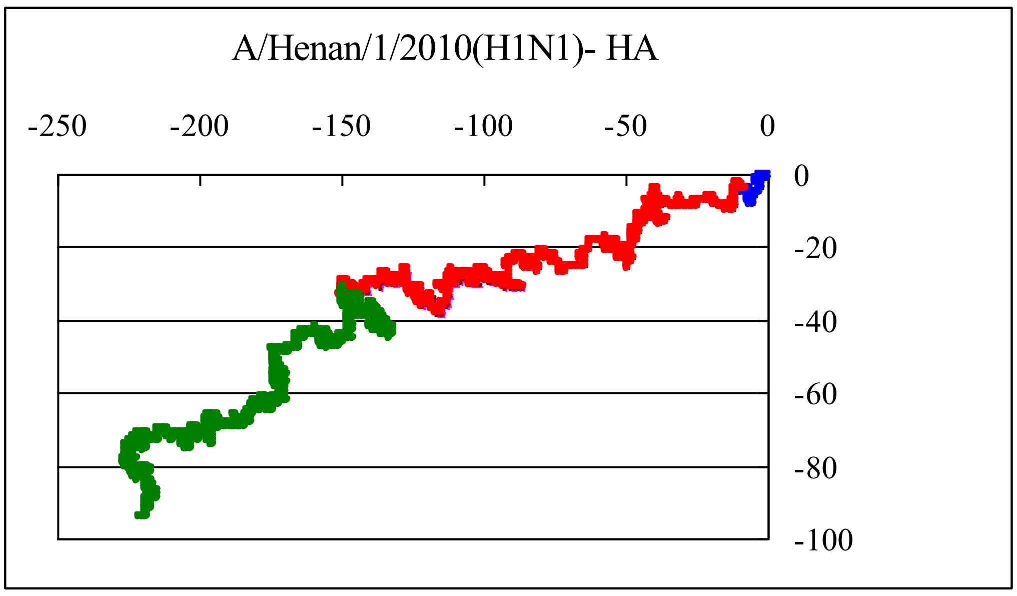


Fig 2: 2D graphical representation of a H1N1 strain hemagglutinin sequence. (Segments: Blue – Signal peptide; Orange- HA1; Green – HA2)

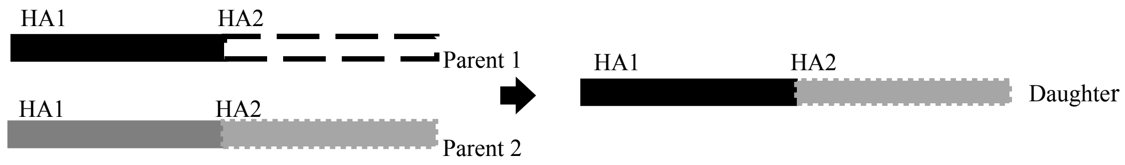


Figure 3. Recombination in HA gene of Influenza virus by genetic exchange at the cleavage point of HA1 and HA2 segments

3. Materials and Methods

Database: We downloaded 1274 full length HA sequences available in Genbank which were major human infecting subtypes i.e H1N1, H5N1, H3N2 and H7N9. All sequences were from Asia during time period of 2010 to 2014. Similarly we analyzed 682 NA of H5N1 subtype available in Genbank till March 2009.

Calculation of sequence similarity: We analyzed the sequences based on a 2D graphical representation system and determined sequence identity from the numerical characterization algorithm [11, 12]. In this method starting from the origin on Cartesian axes a 2D graphical plot is generated for HA and NA sequence of each strain by moving one step in the negative x-direction for an adenine, one step in the positive y-direction for a cytosine, one step in the positive x-direction for a guanine, and one step in the negative y-direction for a thymine. The 2D graphical plot so generated provides a visual representation of the base distribution pattern of the sequence. To numerically characterize the sequence we define a weighted centre of mass of the plot and a graph radius g_R as follows [12]
:

$$\mu_x = \sum_{i=1}^{N} x_i/_N \ , \ \mu_y = \sum_{i=1}^{N} y_i/_N$$

$$g_R = \sqrt{\mu_x^2 + \mu_y^2}$$

where x_i, y_i represents the coordinates of the i^{th} base and N is the total number of nucleotides in the sequence under consideration. The graph radius g_R is a base distribution index of the nucleotide sequences. g_R is found to be sensitive to any changes in the base distribution such that sequences having same value of g_R imply sequence identity [13].

.

4. Conclusions

Our recombination study is based on novel concept of segment exchange at the cleavage points of the well defined segments of a gene that codes for different subunits of the associated protein. Previous recombination studies on the influenza virus had been based on breakpoint analysis [14], where the polymerase recognizes a breakpoint sequence and creates a recombinant strain by switching back and forth between two parents, thus producing a daughter having copies of parts of sequences from both the parents. However, breakpoint recombination is a highly debated topic being postulated by some scientists while refuted by others [15, 16]. It is known that RNA polymerase remains loosely attached with few nucleotides of the template strand [17]. Our hypothesis is that such a loose attachment can facilitate a jump between two templates at consensus sequences, which, in the case of the HA and NA could be the cleavage points of their intrinsic segments. Our analyses revealed many examples where inter-segmental exchanges have led to new daughter strains being developed, thus demonstrating this kind of recombination as a valid evolutionary process. The fact that the exchanges we have observed all appear to occur within same subtypes implies more mechanisms at work than merely the consensus sequences, which though we are yet to clearly understand. However, this new evolutionary process, albeit restricted by segmental recombinations within same subtypes, opens up possibilities of development of many more subtype strains, some of which could turn out to be more pathogenic than the original subtype itself.

Acknowledgments

The authors would like to acknowledge with thanks the opportunity provided by the organizers of the MOL2NET-1 conference to present our findings to a general audience..

.

Author Contributions

AN suggested the problem, AD did the calculations and wrote the paper, which AN critically reviewed and edited.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Tong, S; Zhu, X; Li, Y; Shi, M, Zhang, J; Bourgeois, M; Yang, H; Chen, X; Recuenco, S; Gomez, J; Chen, L.M; Johnson, A; Tao, Y; Dreyfus, C; Yu, W; McBride, R; Carney, P. J; Gilbert, A. T; Chang, J; Guo, Z; Davis, C.T; Paulson, J.C; Stevens, J; Rupprecht, C.E; Holmes, E.C; Wilson, I.A; Donis, R.O. New World Bats Harbor Diverse Influenza A Viruses. PLoS Pathogens 2013, 9(10),e1003657

2. Nandy, A; Sarkar, T; Basak, S.C; Nandy, P; Das, S. Characteristics of influenza HA-N interdependence determined through a graphical technique. Curr Comput Aided Drug Des. 2014,10(4), 285-302.

3. Kageyama, T; Fujisaki, S; Takashita, E; Xu, H; Yamada, S; Uchida, Y; Neumann, G; Saito, T; Kawaoka, Y; Tashiro, M; Genetic analysis of novel avian A(H7N9) influenza viruses isolated from patients in China February to April 2013. Euro Surveill. 2013, 8(15), 20453

4. Worobey, M; Rambaut, A; Pybus, O.G; Robertson, D.L. Questioning the evidence for genetic recombination in the 1918 "Spanish flu" virus. Science 2002, 296(5566), 211

5. Ghosh, A; Nandy A; Nandy, P; Gute, B.D; Basak, S.C. Computational Study of Dispersion and Extent of Mutated and Duplicated Sequences of the H5N1 Influenza Neuraminidase over the Period 1997-2008. J. Chem. Inf. Model. 2009, 49(11), 2627–2638.

6. De, A; Sarkar, T; Nandy, A. Bioinformatics studies of Influenza A hemagglutinin sequence data indicate recombination-like events leading to segment exchanges. (Communicated)

7. Stallknecht, D.E.; Brown, J.D. Tenacity of avian influenza viruses. Rev. Sci. Tech. 2009, 28 (1), 59-67

8. Lindstrom, S.E.; Hiromoto, Y ; Nerome, R; Omoe, K; Sugita, S; Yamazaki, Y; Takahashi, T; Nerome, K. Phylogenetic Analysis of the Entire Genome of Influenza A (H3N2) Viruses from Japan: Evidence for Genetic Reassortment of the Six Internal Genes. J Virol. 1998 ,72, 8021-8031.

9. Zhong, J; Liang, L; Huang, P; Zhu, X; Zou, L; Yu, S; Zhang, X; Zhang, Y; Ni, H; Yan, J. Genetic mutations in influenza H3N2 viruses from a 2012 epidemic in Southern China. Virol J 2013, 10,345.

10. Sarkar, T; Das, S; De, A; Nandy, P; Chattopadhyay, C; Chawla-Sarkar M; Nandy, A. H7N9 influenza outbreak in China 2013: *In silico* analyses of conserved segments of the hemagglutinin as a basis for the selection of peptide vaccine targets. Computational Biology and Chemistry 2015, 59, 8 - 15.

11. Nandy, A. A new graphical representation and analysis of DNA sequence structure: I. Methodology and Application to Globin Genes. Current Sc 1994, 66(4), 309-314.

12. Raychaudhury, C.; Nandy, A. Indexing Scheme and Similarity Measures for Macromolecular Sequences. J Chem Infor Comput Sc. 1999, 39, 243-247.

13. Nandy, A.; Nandy, P. On the uniqueness of quantitative DNA difference descriptors in 2D graphical representation models. Chem Phys Letters 2003, 368,102-107.

14. Boni, M.F; Zhou, Y; Taubenberger, J.K; Holmes, E.C. Homologous recombination is very rare or absent in human influenza A virus. J Virol 2008, 82(10), 4807-4811.

15. Hao, W. Evidence of intra-segmental homologous recombination in influenza A virus Gene 2011, 481(2), 57-64.

16. Boni, M.F; Smith, G.J; Holmes, E.C; Vijaykrishna D. No evidence for intra-segment recombination of 2009 H1N1 influenza virus in swine. Gene. 2012, 494(2), 242-245.

17. Fleischmann, Jr., W.R. Medical Microbiology, 4th edition; Chapter 43 Viral Genetics. The University of Texas Medical Branch, Galveston, Texas, USA. 1996 (Online access at http://www.ncbi.nlm.nih.gov/books/NBK8439/

SciForum
Mol2Net

In SILICO Computer Simulation Risk Assessment of Triazole Fungicides on Human Cytochrome P450 Aromatase Enzyme: *CYP19A1* Inhibition by Triazoles Using Autodock Software

Tamar Chachibaia[1,2], **Joy Harris Hoskeri**[3]

1. Department of Analytical Chemistry, Food Science and Nutrition, Faculty of Pharmacy, University of Santiago de Compostela, Spain
2. Department of Public Health and Epidemiology, Faculty of Medicine, Iv.Javakhishvili Tbilisi State University, Tbilisi, Georgia
3. Institute of Experimental Pharmacology and Toxicology, Slovak Academy of Sciences, Bratislava, Slovakia

* Author to whom correspondence should be addressed; Tamar Chachibaia, (nanogeorgia@gmail.com)

Published: 4 December 2015

Abstract: Inhibitory effect of triazole fungicides were evaluated on the human aromatase enzyme and compared with the Letrozole (LTZ), the most potent inhibitor of aromatase, which is used as anti-estrogen for breast cancer treatment. For this study was used software AUTODOCK to calculate inhibition energy (IE) of triazoles on aromatase enzyme CYP19A1. Those compounds with minimal binding energy are safer in terms of toxicity and resistance of other prescription drugs like non-steroid AIs. In our study we found that four triazole fungicides compounds, Triticonazole, Tebuconazole, Metconazole and Fluquinconazole, exhibited minimal inhibition constant (IC).
.

Keywords: in silico, risk assessment, triazoles, pesticide, aromatase inhibitors, CYP19 enzyme

Mol2Net YouTube channel*: http://bit.do/mol2net-tube*

1. Introduction

Biosynthesis of estrogens from androgens is catalyzed by cytochrome P450 aromatase. Aromatase inhibition by the triazole compounds Letrozole (LTZ) and Anastrozole is a prevalent therapy for estrogen-dependent postmenopausal breast cancer.

Azoles are widely used as agricultural fungicides and antimycotic drugs that target 14α-demethylase. Some were previously shown to inhibit aromatase, thereby raising the possibility of endocrine disruptive effects. However, mechanistic analysis of their inhibition has never been undertaken.

We have evaluated the inhibitory effects of 15 common fungicides in human aromatase enzyme in comparison with the Letrozole (LTZ), the most potent inhibitor of aromatase used as anti-estrogen for breast cancer treatment using AUTODOCK software for calculation of inhibition energy on CYP19 aromatase enzyme. [i]

Triazole containing compounds as systemic fungicides are widely used in agriculture due to its high efficiency, broad spectrum, low toxicity and long effectiveness [ii]. Currently 16 triazole fungicides: bitertanol, cyproconazole, difenoconazole, epoxiconazole, fluquinconazole, flusilazole, flutriafol, hexaconazole, metconazole, myclobutanil, penconazole, propiconazole, tebuconazole, triadimefon, triadimenol and triticonazole, are approved by Swiss Federal Office of Public Health (Zürich, Switzerland). Switzerland no longer allows the use of many chemicals that are still sprayed on American fields (Rosensteil, 2015). By 2005 was set the goal to halve the pesticide pollution of water bodies [iii]. Although, by 2014 report was released that in the five rivers in Switzerland's found heavily polluted in spring and summer by a cocktail of different pesticides [iv].

Target enzymes of triazoles in steroidogenesis are the sterol *14-alfa-demethylase* (encoded by the CYP51 gene) and the *aromatase* (encoded by the CYP19 gene).

The human aromatase enzyme is a member of the cytochrome P450 family and is the product of the *CYP19A1* gene, located on chromosome 15 [v, vi]. Aromatase is the only known vertebrate enzyme that can aromatize a six-membered ring; aromatase is, therefore, the sole source of estrogen in the body [vii].

Nevertheless, since aromatase was first characterized, research has been impeded by the lack of its three dimensional structure. In 2009, Ghosh *et al.* successfully solved the crystallized structure of human aromatase enzyme and provides a structural basis for the specificity to androgen [viii, ix].

The catalytic site of aromatase is located at the juncture of the I and F helices, β-sheet 3, and as the B-C loop. Androstenedione binds into the steroid binding pocket such that its β-face orientates towards the heme group of aromatase, placing C19 within 4.0 Å of the Fe atom. This binding site is only possible if the I-helix backbone is moved 3.5 Å, creating a binding pocket that is approximately 400 $Å^3$. This important distortion is created by residue P308, without which N309, steric hindrance would prevent catalytic activity.

This crystal structure of aromatase will not only allow better structure-based drug design than previous models, but it has also allowed a direct analysis of why some currently available aromatase inhibitors function better than others [x].

As triazole moieties are widely used in fungicides, some studies reported that agricultural triazole pesticides are culprits for the development of resistance to other triazole containing drugs [xi] e.g. triazole aromatase inhibitor antiestrogens. Among them are Anastrozole and Letrozole, the third generation nonsteroidal aromatase inhibitors (AIs), which are now used as first-line therapy in the treatment of breast cancer in postmenopausal women (Scheme 2) [xii, xiii, xiv]. In recent years, some triazole residues have been found in agricultural products, including fruits, wheat, tea leaves and wine and water [xv,xvi,xvii,xviii,xix]. In one study was concluded that many azole compounds developed as inhibitors of fungal sterol 14-alfa-demethylase are inhibitors also of mammalian sterol 14-alfa-demethylase and mammalian aromatase with unknown potencies [xx].

To avoid the risk of possible development of resistance to other triazole drugs and to reduce toxicity of aromatase inhibitors in the treatment of breast cancer, there are used different methods in order to find out new preventive strategies.

Hypothesis of this study is that common agricultural triazole fungicides may express inhibitory effect on human aromatase enzyme Cyp19A1, in this way contributing to drug resistance and increased risk of cumulative toxicity of anti-estrogen medications.

In our study we performed virtual screening of 15 fungicides and AI reference drug Letrozole to measure inhibitory effect on human aromatase. In this way we aim to range which pesticides are most potent inhibitors of Cyp19A1 enzyme to predict and prevent possible summative cumulating effect of fungicide undesirably overlapping with the activity of anticancer drugs.

The publication of a high resolution X-ray structure of human aromatase has opened the way to a greater understanding of the structural basis for estrogen synthesis and substrate/inhibitor recognition [xxi]. Triazole aromatase inhibitors (AIs) bind to the active site of CYP19 by coordinating the heme iron atom of the enzyme through a heterocyclic nitrogen lone pair.

In our docking study we used together the X-ray structure of human cytochrome P450 aromatase Cyp19A1 (PDB code 3S79, resolution 2.75 Å) [xxii] associated with the metabolism of estrogens and carcinogens with breast cancer, with a collection of commercially available compounds, particularly, 15 triazole fungicides and anticancer drug Letrozole as reference standard (Table 1).

Molecular docking is established method for analysis of molecular associations, which is mostly used in the drug discovery field to study the binding of small molecules (ligands) to macromolecules (receptor) [xxiii].

Cytochrome P450 aromatase homology models were published and used to perform docking and molecular dynamics simulations on known AIs [xxiv, xxv]

.

. **Materials and methods:**

The availability of X-ray structure of human aromatase enables us to set up docking protocol by AutoDock software to identify iron - ligand interactions between heme protein and 16 different triazole ligands, as chemical scaffolds able to inhibit aromatase, thus testing interactions within the aromatase binding site.

Computational ligand docking methodology, AutoDock 4.0, based on Lamarckian genetic algorithm [xxvi] was employed for virtual screening of a compound library with 16 entries including reference compound as Letrozole, the 3rd generation aromatase inhibitors for the treatment of breast cancer, with the enzyme Cytochrome P450 aromatase(Cyp19A1), a potential drug target.

Autodock 4.0 uses GA as a global optimizer combined with energy minimization as a local search method [xxvii].

The macromolecule, Cytochrome P450 aromatase or Cyp19A1 (PDB code 3S79, resolution 2.75 Å) was retrieved by using AutoDock 4 (The Scripps Research Institute, Molecular Graphics Laboratory, 10550 North Torrey Pines Road, CA, 92037) running on operative system Windows 7 (Miscosoft corporation 2007)

PRODRG was used to draw the 2D structures of different ligands. All the structures were written in protein database (PDB) format. Input molecules files for an AutoDock experiments must confirm to the set of atom types supported by it. Therefore, PDBQT format was used to write ligands, recognized by AutoDock.

Torsional degree of freedom (TORSDOF) is used in calculating the change in the free energy caused by the loss of torsional degree of freedom upon binding. In the AutoDock 4.0 force field, the TORSDOF value for a ligand is the total number of rotatable bonds in the ligand.

The 3D crystal structure of Cytochrome P450-aromatase Cyp19A1 (Picture 1) PDB code 3S79, resolution 2.75 Å was downloaded from Brookhaven Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB; http://www.rcsb.org/pdb).

The nonbonded oxygen atoms of waters, present in the crystal structure were removed. After assigning the bond orders, missing hydrogen atoms were added, then the partial atomic charges was calculated using Gasteiger–Marsili method (Gasteiger). United atom charges were assigned, non-polar hydrogens were merged, and rotatable bonds were assigned, considering all the amide bonds as non-rotatable. The receptor file was converted to PDBQT format, which is PDB plus "q" charges and "t" AutoDock type. (To confirm the AutoDock types, polar hydrogens should be present, whereas non-polar hydrogens and lone pair should be merged, each atom should be assigned Gasteiger partial charges). Amino acids which form target pocket or inhibition cite of aromatase

.

Molecular Docking Study:

In the present work, we have studied the in silico binding affinities to the active pocket (Pic. 2.) of enzyme 3S79 (Pic. 1.) to the selected 15 triazole fungicides (Scheme 1.) and the standard anti-aromatase drug Letrozole.

Of the three different search algorithms offered by AutoDock 4.0, the Lamarckian Genetic algorithm (LGA) based on the optimization algorithm was used in favor to other two – simulated annealing and genetic algorithm.

For all dockings, 10 independent runs with step sizes of 0.2Å for translations and 5Å for orientations and torsions were used. AutoDock tools along with AutoDock 4.0 and AutoGrid 4.0 was used to generate both grid and docking parameter files (i.e., gpf and.dpf files) respectively.

A grid box size of 42 x 42 x 42 Å points with a grid spacing of 0.375 Å was generated using AutoGrid [xxviii]. The grid was centered at x,y,z coordinates of 85.51, 52.282, 48.114, which was reported as the binding site residues.

For each docking experiment, the lowest energy docked conformation was selected from 10 runs. The successful completion of docking experiment took from 1 to 4 hours, on a 2.0 GHz Intel (R) core 2 duo machine with 3.0 GB of RAM and Windows 7 operating system.

Prior to actual docking run, AutoGrid 4.0 was introduced to precalculate grid maps of interaction energies of various atom types.

The energy of interaction of this single atom with the protein is assigned to the grid point. An affinity grid is calculated for each type of atoms in the substrate, typically carbon, oxygen, nitrogen, and hydrogens as well as grid of electrostatic potential using a point charge of 1 as the probe. Autodock 4.0 uses these interaction maps to generate ensemble of low energy conformations. It uses a scoring function based on AMBER force field, and estimates the free energy of binding of a ligand to its target. For each ligand atom types, the interaction energy between the ligand atom and the receptor is calculated for the entire binding site which is discretized through a grid. This has the advantage that interaction energies do not have to be calculated at each step of the docking process but only looked up in the respective grid maps.

Since a grid map represents the interaction energy as a function of the coordinates, their visual inspection may reveal the potential unsaturated hydrogen acceptors or donors or unfavorable overlaps between the ligand and the receptor.

Results:

The binding affinity was evaluated by the binding energies, docking energy, inhibition constant, intermolecular energy, and RMSD values. It was demonstrated that the docking protocol could reliably reproduce the interaction of aromatase with its substrate with an RMSD of 0 Å.

The results of LGA docking experiments of the triazoles using AutoDock 4.0 and AutoGrid 4.0 are summarized in Table 1.

Binding energy for reference compound Letrozole (Fig.3) in our docking study is comply with other studies and is in agreement with them [xxix].

Triazole compounds (Fig 1a,b,c) 1, 3, 10 are chosen as possessing aromatase inhibitory potency based on obtained algorithmic parameters docking: highest binding energy, highest inhibition constant, and hydrogen bonds.

Compound 1 (Bitertanol) exhibits RMS equivalent to zero in 7th orientation to heme molecule of protein, and at this position binding energy is - 6,19; IC – 2,9X10-5; and two hydrogen bonds with amino acids of target pocket ASP371 and LEU372.

Compound 3 (Difenoconazole) demonstrated best compliance of inhibitory bound in 7th orientation to RMS=0 posing heme molecule, binding energy -7,36, IC – 4,03X10-6, and one hydrogen bond with target pocket amino acid THR310.

Compound 10 (Penconazole) exhibits its highest binding energy -7.71 at orientation 6th to heme molecule parallel alignment at RMS zero point, IC – 2.22X10-6 and one H-BOND with THR310.

In our study we found that four triazole fungicides compounds 15, 12, 8, 5 exhibited minimal inhibition constant (IC). Those are Triticonazole, Tebuconazole, Metconazole and Fluquinconazole. (Fig. 4 a,b,c,d).

Compound 15 exhibits its binding energy -21.65 at orientation 2nd to heme molecule parallel alignment at RMS zero point, IC – 1.35X10-016 and no H-BOND.

Compound 12 exhibits its binding energy -21.09 at orientation 7th to heme molecule parallel alignment at RMS zero point, IC – 3.5X 10-016 and no H-BOND.

Compound 8 exhibits its binding energy -19.69 at orientation 5th to heme molecule parallel alignment at RMS zero point, IC – 3.68X10-015 and no H-BOND.

Compund 5 exhibits its binding energy -17.25 at orientation 9th to heme molecule parallel alignment at RMS zero point, IC 2.29 X10-013 and no H-BOND.

Conclusions:

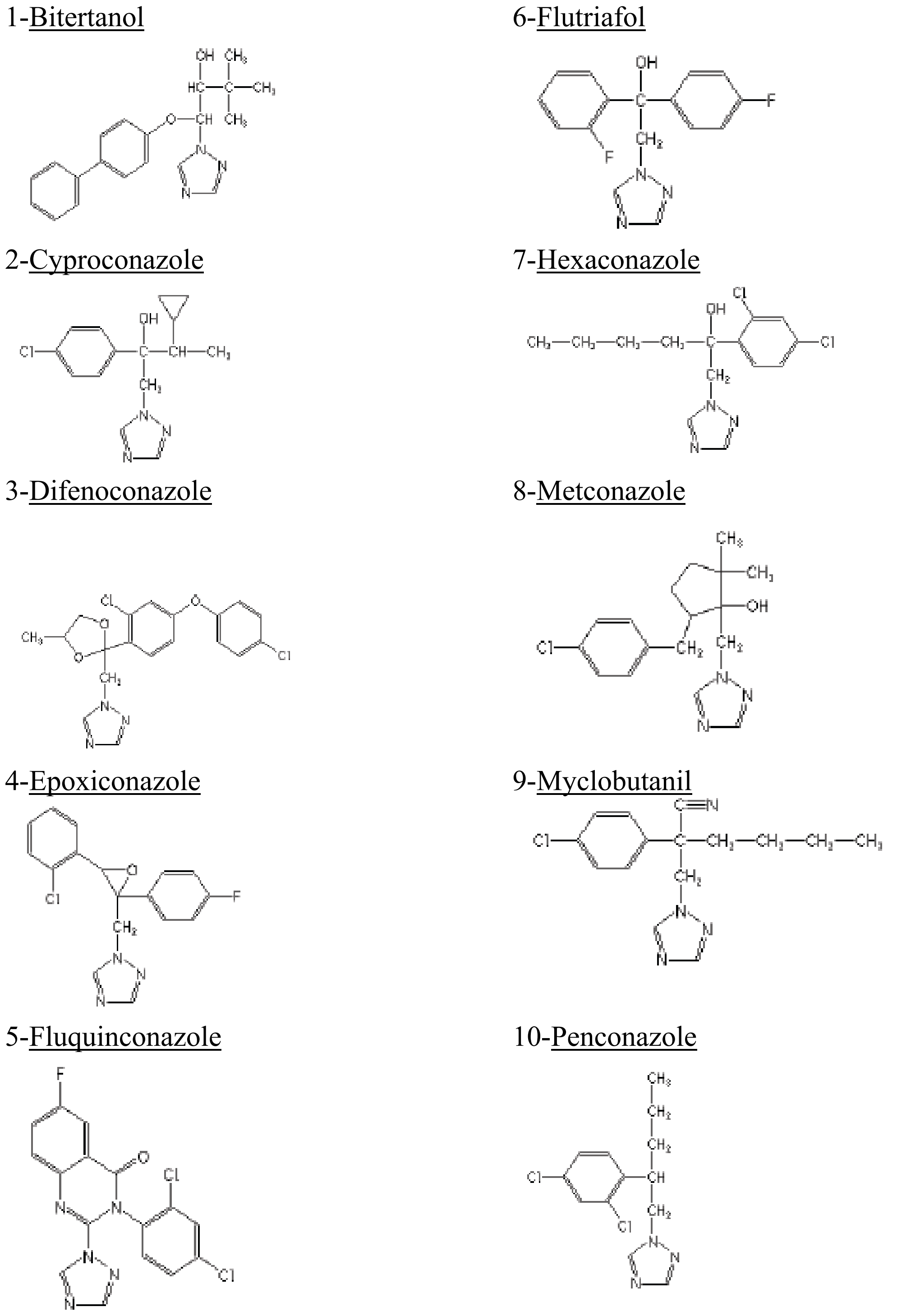
Those compounds with minimal binding energy are safer in terms of toxicity and resistance of other prescription drugs like non-steroid AIs. Those with higher binding energies may cause drug resistance or toxicity in cases of simultaneous administration and it should be taken cautiously during treatment with other triazole containing drugs like Letrozole.

.

Table 1. Results of docking of 15 fungicides and one non-steroid Aromatase inhibitor (NSAI) Letrosole.

SL. NO.	**MOLECULE**	**ORIENTATION**	**BINDING ENERGY** Kcal/mole	**DOCKING ENERGY** Kcal/mole	**INHIBITATION CONSTANT** (Ki) (nM)	**INTERMOL ENERGY**	**TORSIONAL ENERGY**	**INTERNAL ENERGY**	**RMSD**	**HYDROGEN BOND**
1.	Bitertanol	7th	-6.19	-5.09	2.9 e-005	-8.06	1.87	2.97	0.0	2 H-BONDS WITH **ASP371, LEU372**
2.	Cyproconazole	10th	-9.25	-8.73	1.66 e-007	-10.81	1.56	2.08	0.0	1 H-BOND WITH **MET374**
3.	Difenoconazole	7th	-7.26	-6.83	4.03 e-006	-8.92	1.56	2.09	0.0	1 H-BOND WITH **THR310**
4.	Epoxiconazole	8th	-10.83	-12.58	1.04 e-008	-12.14	1.25	-0.44	0.0	1 H-BOND WITH **MET374**
5.	**Fluquinconazole**	9th	-17.25	-17.83	**2.29 e-013**	-17.87	0.62	0.04	0.0	
6.	Flutriafol	2nd	-8.7	-7.94	4.18 e-007	-9.95	1.25	2.01	0.0	
7.	Hexaconazole	7th	-10.72	-11.52	1.38 e-008	-12.59	1.87	1.07	0.0	1 H-BOND WITH **MET374**
8.	**Metconazole**	5th	-19.69	-19.92	**3.68 e-015**	-20.95	1.25	1.01	0.0	
9.	Myclobutanil	1st	-8.5	-8.1	5.92 e-007	-10.36	1.87	2.27	0.0	1 H-BOND WITH THR310
10.	Penconazole	6th	-7.71	-9.07	2.22 e-006	-9.27	1.56	0.2	0.0	1 H-BOND WITH THR310
11.	Propiconazole	3rd	-8.64	-10.17	4.64 e-007	-10.2	1.56	0.03	0.0	
12.	**Tebuconazole**	7th	-21.09	-21.91	**3.5 e-016**	-22.95	1.87	1.04	0.0	
13.	Triadimefon	4th	-8.46	-9.97	6.29 e-007	-10.02	1.56	0.04	0.0	
14.	Triadimenol	9th	-10.59	-11.37	1.72 e-008	-12.15	1.56	0.78	0.0	
15.	**Triticonazole**	2nd	-21.65	-21.96	**1.35 e-016**	-22.89	1.25	0.94	0.0	
16	**Letrozole**	7th	-9.54	-8.77	**1.01 e-007**	-10.48	0.93	1.71	0.0	

Scheme 1. Downloaded from free the public domain http://www.alanwood.net/pesticides.



1-Bitertanol

6-Flutriafol

2-Cyproconazole

7-Hexaconazole

3-Difenoconazole

8-Metconazole

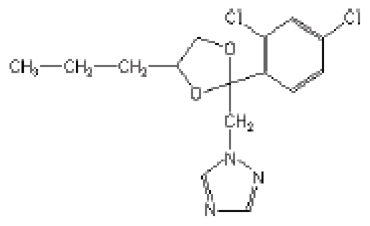
4-Epoxiconazole

9-Myclobutanil

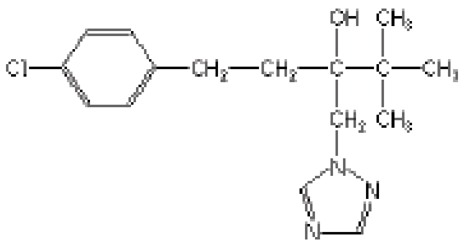
5-Fluquinconazole

10-Penconazole

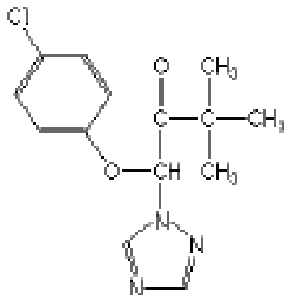
11-Propiconazole



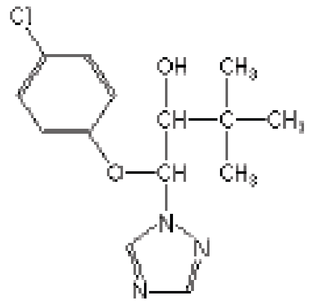
12-Tebuconazole



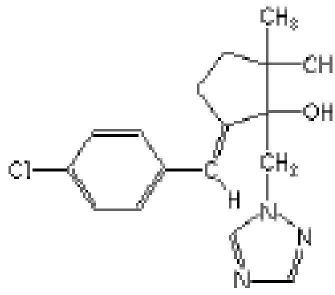
13-Triadimefon



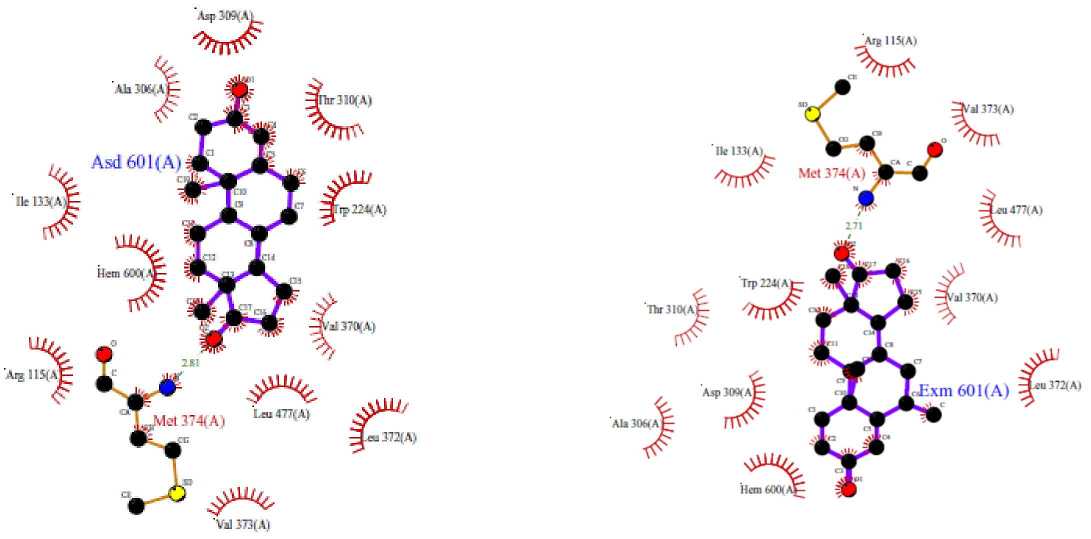
14-Triadimenol



15-Triticonazole



Pic. 1. Human placental aromatase cytochrome P450 aromatase (CYP19A1) refined at 2.75 angstrom 3S79 (ribbon model). [source: http://www.rcsb.org/pdb/explore/jmol.do?structureId=3S79].



Pic 2. Target pocket surrounding Heme-molecule of aromatase *CYP19A1*.

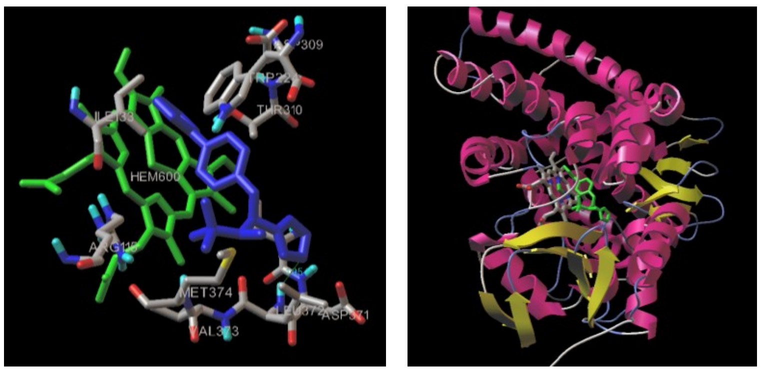


Fig. 2a: Compound 1 docked within the binding pocket of the enzyme 3S79. Predicted binding mode of compound 1 (). On the left, stick and ball model, and on the right, ribbon model of enzyme 3S79 in the binding pocket of which compound 1 is forming hydrogen bond with the amino acids **ASP371** and **LEU372**.

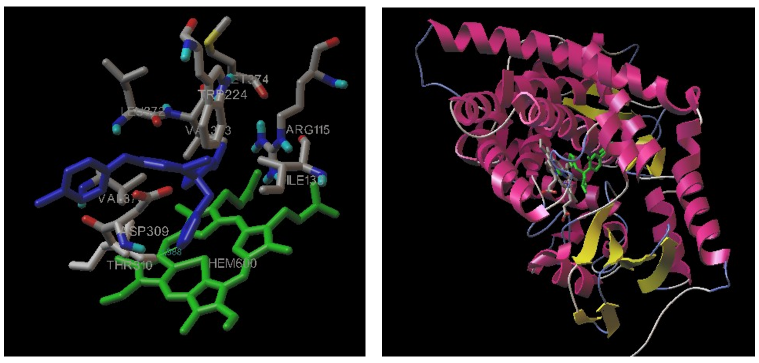


Fig. 2b: Predicted binding mode of compound 3. On the left, stick and ball model, and on the right, ribbon model of enzyme 3S79 in the binding pocket of which compound 3 is forming hydrogen bond with the amino acid **THR310**.

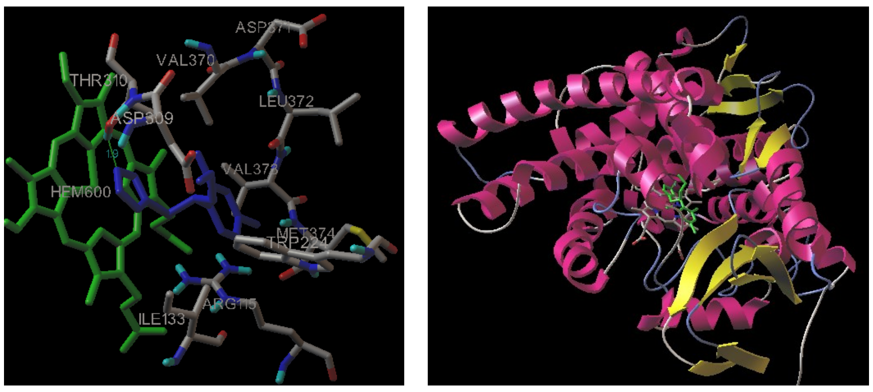


Fig. 2c: Predicted binding mode of compound 10. On the left, stick and ball model, and on the right, ribbon model of enzyme 3S79 in the binding pocket of which compound 3 is forming hydrogen bond with the amino acid **THR310**.

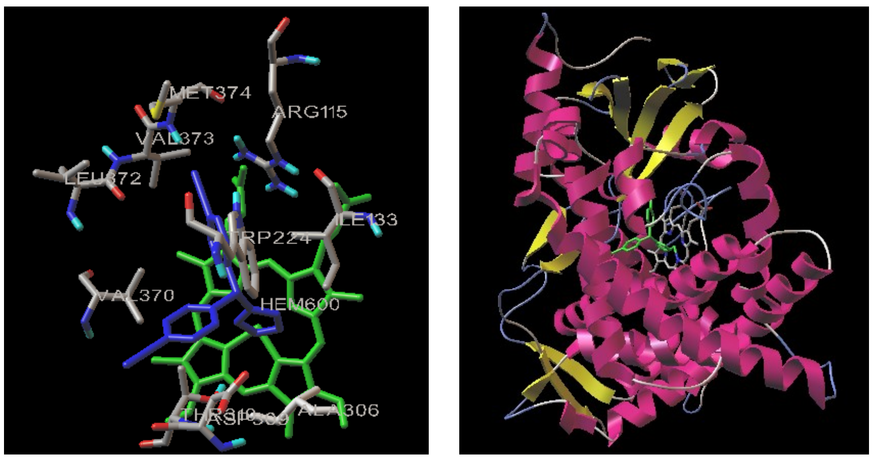
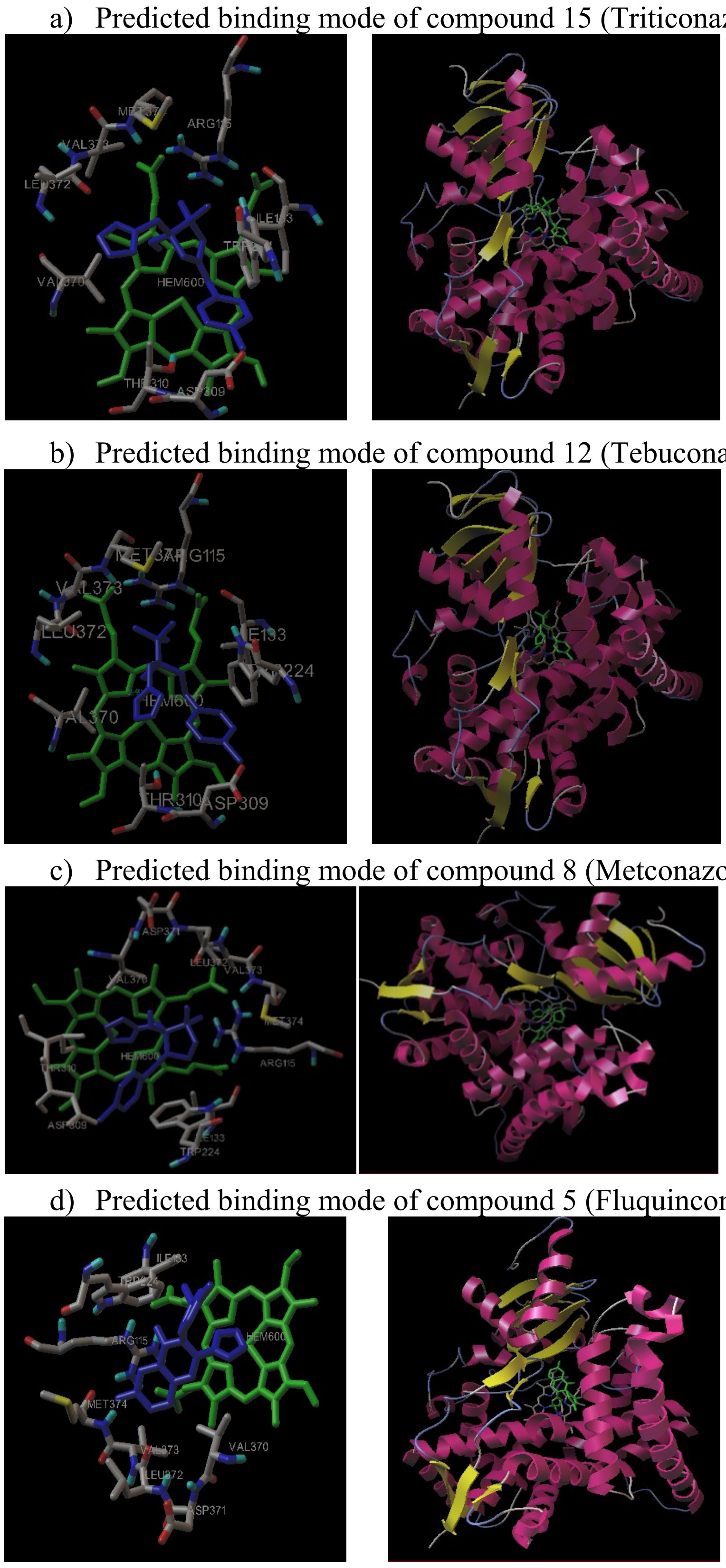


Fig. 3: Reference compound, Letrozole docked in the binding pocket of the enzyme 3S79. (Ball and stick model and ribbon model).



a) Predicted binding mode of compound 15 (Triticonazole).

b) Predicted binding mode of compound 12 (Tebuconazole)

c) Predicted binding mode of compound 8 (Metconazole)

d) Predicted binding mode of compound 5 (Fluquinconazole)

Fig. 4. Molecular surface view of compounds **15, 12, 8, 5** docked within the binding pocket of the enzyme 3S79 without H-bonds.

References and Notes

[i] C. Egbuta, J. Lo, D.Ghosh (2014) Mechanism of inhibition of estrogen biosynthesis by azole fungicides. Endocrinology. 155(12):4622-8. doi: 10.1210/en.2014-1561.

[ii] D. Feng, J. Guo , W. Song, W. Hu, Z. Li (2011). Comparative quantitative structure–activity relationship (QSAR) study on acute toxicity of triazole fungicides to zebrafish, Chemistry and Ecology, 27:4, 359-368, DOI: 10.1080/02757540.2011.585780

[iii] H. Singer. (2002), Pesticides in Water - Research Meets Politics. J. EAWAG News, 59. 16-19.

[iv] http://www.swissinfo.ch/eng/pollution_pesticide--cocktail--found-in-swiss-rivers/38096166

[v] E. Thompson Jr, P. Siiteri (1974) Utilization of oxygen and reduced nicotinamide adenine dinucleotide phosphate by human placental microsomes during aromatization of androstenedione. J Biol Chem. Sep 10; 249(17):5364-72.

[vi] S.Chen, M. Besman, R.Sparkes, S.Zollman, I. Klisak, et al. (1988) Human aromatase: cDNA cloning, Southern blot analysis, and assignment of the gene to chromosome 15. DNA, 7(1):27-38.

[vii] B. Amarneh, C. Corbin, J. Peterson, E. Simpson, S. Graham-Lorence (1993) Functional domains of human aromatase cytochrome P450 characterized by linear alignment and site-directed mutagenesis. Mol Endocrinol. 7(12):1617-24.

[viii] D. Ghosh, J. Griswold, M. Erman, W. Pangborn (2009). Structural basis for androgen specificity and oestrogen synthesis in human aromatase. Nature. 8; 457(7226):219-23.

[ix] D. Ghosh, J. Griswold, M. Erman, W. Pangborn (2010) X-ray structure of human aromatase reveals an androgen-specific active site. J Steroid Biochem Mol Biol. 118(4-5):197-202

[x] S. Chumsri, T.Howes, T. Bao, G. Sabnis, A. Brodie (2011). Aromatase, Aromatase Inhibitors, and Breast Cancer. The Journal of Steroid Biochemistry and Molecular Biology, 125(1-2), 13–22. doi:10.1016/j.jsbmb.2011.02.001

[xi] E. Snelders, S. Camps, A. Karawajczyk, G. Schaftenaar, G. Kema, et al (2012) Triazole Fungicides Can Induce Cross-Resistance to Medical Triazoles in Aspergillus fumigatus. J. PLOS ONE. DOI: 10.1371/journal.pone.0031801

[xii] A. Brodie. (2002) Aromatase inhibitors in breast cancer. Trends in Endocrinology & Metabolism, 13-2. 61-65.

[xiii] J. Geisler. (2011) Differences between the non-steroidal aromatase inhibitors anastrozole and letrozole - of clinical importance? British Journal of Cancer, 104(7), 1059–1066.

[xiv] R. Brueggemeier, J. Hackett, E. Diaz-Cruz (2005), Aromatase Inhibitors in the Treatment of Breast Cancer, Endocrine Reviews, 26:3, 331-345

[xv] V. Kumar, S. Ravindranath, A. Shanker, (2004) Fate of hexaconazole residues in tea and its behavior during brewing process, J. Chem. Health Saf. 11, 21–25.

[xvi] E. Trösken, N. Bittner, W. Völkel, (2005), Quantitation of 13 azole fungicides in wine samples by liquid chromatography–tandem mass spectrometry, J. Chromatogr. A 1083, 113–119.

[xvii] Q. Zhou, J. Xiao, Y. Ding, (2007) Sensitive determination of fungicides and prometryn in environmental water samples using multiwalled carbon nanotubes solid-phase extraction cartridge, Anal. Chim. Acta, 602, 223–228.

[xviii] R. Jeannot, H. Sabik, E. Sauvard, E. Genin, (2000) Application of liquid chromatography with mass spectrometry combined with photodiode array detection and tandem mass spectrometry for monitoring pesticides in surface waters, J. Chromatogr. A 879, 51–71.

[xix] L. Paraíba, (2007) Pesticide bioconcentration modelling for fruit trees, Chemosphere 66, 1468–1475.

[xx] J. Zarn, B.Brüschweiler, J. Schlatter (2003) Azole Fungicides Affect Mammalian Steroidogenesis by Inhibiting Sterol 14α-Demethylase and Aromatase Environmental Health Perspectives, 111(3):255–261.

[xxi] D.Schuster, C.Laggner, T. Steindl, A Palusczak, R. Hartmann, T. Langer, (2006), Pharmacophore modeling and in silico screening for new P450 19 (aromatase) inhibitors. J. Chem. Inf. Model., 46, 1301–1311.

[xxii] D. Ghosh, J. Lo, D. Morton, D. Valette, J. Xi, J. Griswold et al. Novel Aromatase Inhibitors by Structure-Guided Design (2012), J. Medicinal Chemistry 55 (19), 8464-8476. DOI: 10.1021/jm300930n

[xxiii] X. Barril, S.Morley (2005) Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. J. Med. Chem. 48, 4432–4443.

[xxiv] A. Favia, A. Cavalli, M.Masetti, A.Carotti, M. Recanatini (2006) Three-dimensional model of the human aromatase enzyme and density functional parameterization of the iron-containing protoporphyrin IX for a molecular dynamics study of heme-cysteinato cytochromes. Proteins, 62, 1074–1087.

[xxv] S.Karkola, H. Holtje, K. Wahala, (2007) A three-dimensional model of CYP19 aromatase for structure-based drug design. Steroid Biochem. Mol. Biol., 105, 63–70.

[xxvi] F.Solis, R.Wets, (1981) Minimization by random search techniques. Math. Oper. Res., 6, 19-30.

[xxvii] G. Morris, D. Goodsell, R. Halliday, R Huey, W.Hart, R.Belew, A.Olson (1998) Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. J. Comput. Chem. 19, 1639–1662.

[xxviii] G. Morris, D.Goodsell, R.Halliday, R. Huey, W.Hart, R.Belew, A.Olson (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comput. Chem., 19, 1639-1662.

[xxix] N. Suvannang, C. Nantasenamat, C. Isarankura-Na-Ayudhya, V. Prachayasittikul (2011) Molecular docking of Aromatase Inhibitors. Molecules, 16(5), 3597-3617; doi:10.3390/molecules16053597.

SciForum
Mol2Net

*MetAlgNet :*Metabolic Pathway Network Reconstruction from Algae Genome Annotation Data

Kirtan Dave [1]*, DarshanChoksi [1], Hetalkumar Panchal [1]

[1] G.H. Patel P.G. Dept. of Computer Science and Technology, Sardar Patel University, Vallabh Vidyanagar, Gujarat, India; E-Mail: kirtandave11@gmail.com

* E-Mail: kirtandave11@gmail.com; Tel: +91-02692-236829; Fax: +91-02692-236829;

Published: 4 December 2015

Abstract: Post-genomic molecular biology embodies high-throughput experimental techniques and hence it is a data-rich field. The goal of development of this tool is to utilize free available biological data of green algae in order to produce new metabolic pathway knowledge and to aid mining of newly generated data. The variety of biological sequence and functional information are stored in different online database, so getting annotation information of genome from different database is challenging task for reconstruction of pathways. Here we apply data integration approach to provide rich representation that enables pathway names based text mining of biological data in terms of integrated networks and conceptual spaces. The publicly available green algae genome annotated data can be used to aid mining of important biological enzymes in metabolic networks. We developed an integrative bioinformatics approach that utilizes publicly available knowledge of enzyme-metabolites interactions, network topological analysis like betweenness, closeness and degree for assigning node importance with quantitative values. The application of our software is revealed importance of role of potential enzymes in biological functions in view of network centrality values, which were calculated by various algorithms. The results provided in this work indicate that integration of heterogeneous biological data facilitates advanced mining of data to create metabolic pathway networks. The methods can be applied for gaining insight into functions of enzymes, metabolites and other molecules, as well as for offering interpretation of functional evolution of metabolites with help of topological analysis and reconstruction of phylogenetic tree from sequence data.

Keywords: metabolic networks, green algae, centrality, Python

1. Introduction

The advancement of new technology leads to production of large amount of biological data such as high-throughput sequencing data, metabolomics data, transcriptomics data and many more.The metabolic networks are complex due to their size and the presence of bimolecular reactions; so combined knowledge of biology, computer science and graph theory will help understand molecular network complexity[1]. Within the biological sciences, one of the primary challenges is to investigate how the collective behavior of cells, tissues, or organisms can be understood in terms of the properties of their molecular constituents from a metabolic network [2]. There is an essential role of metabolic networks in all biological processes of a living cell. Some are like biochemical pathways to protein interactions and gene regulation to cellular communication. Traditionally, genes and proteins involved in different functionalities have been studied in isolation or in small clusters. However, the complex nature of a cell cannot be fully understood by studying individual components in isolation. To investigate this intricate connectivity of cellular systems, the analysis of complex networks has become an important part of molecular biology [3]. Cellular system can be viewed as a combination of omics technologies, data integration, analysis, mining, and visualization often involving use of these techniques iteratively over hypothesis driven systematic experimental design to gain increased understanding of the structure and dynamics of the biological systems [4]. In Fig-1 an integrative bioinformatics starts with the integration of multiple datasets from one or more omics and also possibly from multiple organisms, and forms the basis for systems biology analysis.

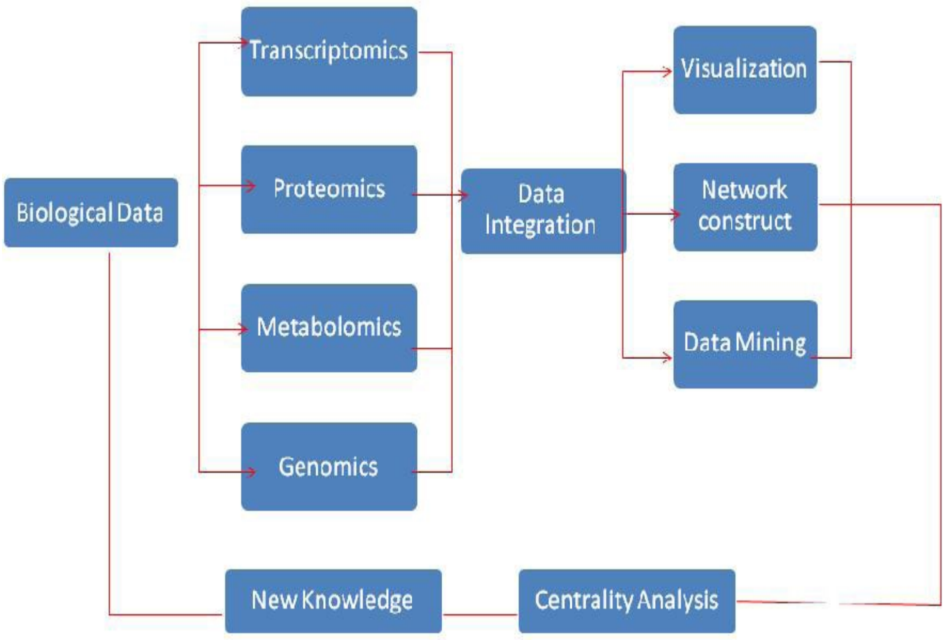


Figure 1. Basic schema for **MetAlgNet**

2. Results and Discussion

MegAlgNet data integration system facilitates mining of biological data and hence exploration ofsome useful patterns, novel relationships between different biological entities from the data, andmay provide novel insights into metabolic functions and context- specific biological functions.

The MetAlgNet creates a network from annotated data and the purpose for developing this network is to get knowledge of potential element from topological analysis with generated network. The software created random network from GMT file (see Fig. 3).We created 55 different pathway networks from the standard biological pathway as per KEGG database; the main purpose of creating this network is getting inference from it. However, the tool has ability to generate more number of networks with respective search term.

Along with centrality, we also reconstruct phylogenetic tree from respective annotated data of particular pathway. Here, we summarize result of pathways which generated from MetAlgNet data mining that were four main result in consideration 1) Network generated from particular pathway from specific search term 2) identification of potential node of respective network with help of node ranking algorithm 3) degree, closeness and betweenness centrality calculation and bar chart generation and 4) phylogenetic tree generation from sequence data.The interpretation lead to identification of major role of particular enzyme in network and chemical compound. The networks given below are generated using 1.*Chlamydomonas reinhardtii*, 2.*Ostreococcus lucimarinus*, 3.*Ostreococcus tauri*and 4.*Volvox carteri*. So, collecting all data from each of the organism database tool, creates a comprehensive GMT file.

The GMT file is further utilized for creating a network of enzymes and metabolites. The resulting networks showed surprisingly high level of connectivity across different stages of linear metabolic pathways via enzyme and metabolite interactions. The centrality analysis plays major role to identify a potential node in the network. If the network has a very high average closeness value, it leads to more organized functional units or modules. The degree could indicate a central role in a biological network. It may indicate relevance of a node as functionally capable of holding together other nodes in the network. Betweenness of a node effectively indicates the capability of a node to bring in distant nodes to perform communication in network (see Fig 4) .

3. Materials and Methods

Primary requirement for annotation is collecting genome data of desired organisms. However, if complete annotation data is not available, so we can annotate data with available genome annotationpipeline. The raw data collected from NCBI sequence read archive database or DDBJ or EBI-SRA database[5-7] for 1.Chlamydomonas reinhardtii, 2.Ostreococcus lucimarinus, 3.Ostreococcus tauri and 4.Volvox carteri. In context of four different algae, there is list of data available, which we have downloaded and used in annotation. The major data used in making database are listed below in table 1. However, there are lots of incomplete data, so we try to avoid use in study[8-10].

We took permanent draft and complete sequenced data for study. The study also considers other source annotated data of respected algae. There are many community based genome annotation projects going on like OrcAE (Online Resource for Community Annotation of Eukaryotes) and Phytozome (JGI annotation resource). Both online databases provide very much useful annotation data which contain Gene locus name, transcript name, protein name, PFAM, Panther ID, KOG, EC NO, KEGG Orthology and Gene ontology[11].

Table 1. The various data numbers representing each database, 1. CRE (Chlamydomonasreinhardtii) 2. OLU (Ostreococcuslucimarinus) 3. OTA (Ostreococcustauri) and 4.VCN (Volvox carteri).

	Org_db	**Org_anno_db**	**Org_cds_db**	**Org_gene_db**	**Org_protein_db**	**Org_rxn_db**
CRE	113	19526	19526	3433	19526	4635
OLU	109	7796	7796	3146	1196	5624
OTA	108	6912	6912	3309	6912	4067
VCU	114	15285	14971	3484	14971	3915

Data from various public data sources were collected into our local database systems The curation of a free available data from database involves several steps required in the curation process.

Multiple molecular biology databases provide descriptions of biological systems at different levels of abstraction. Some common biological information, along with names of primary databases providing information is indicated in figure-2.

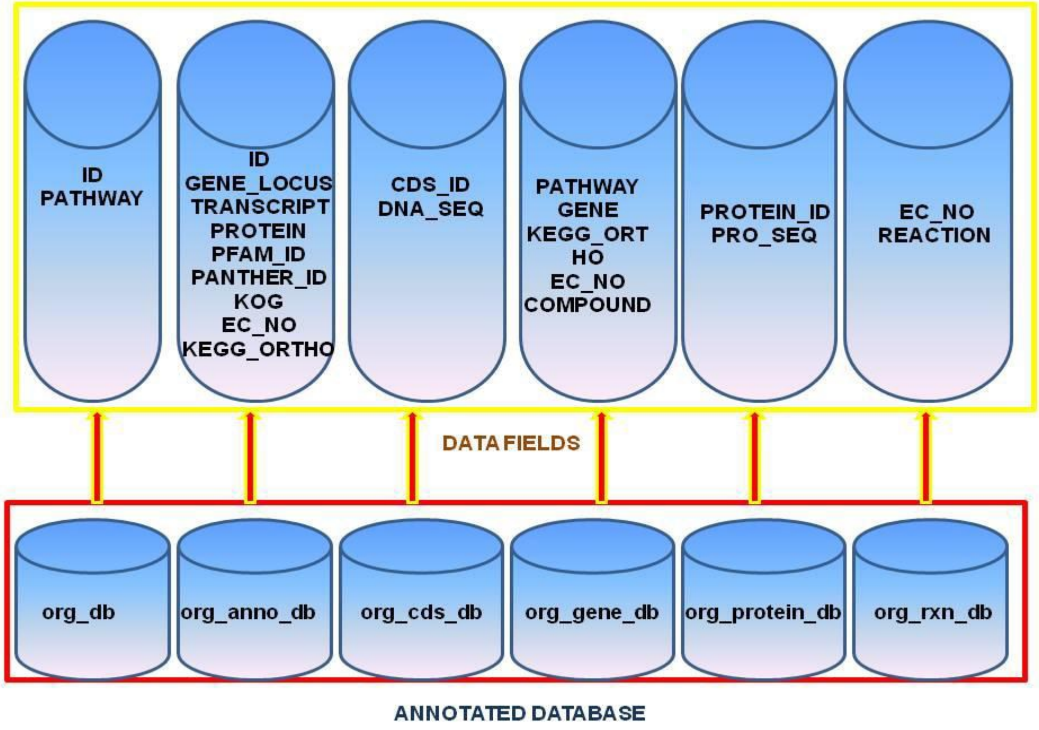
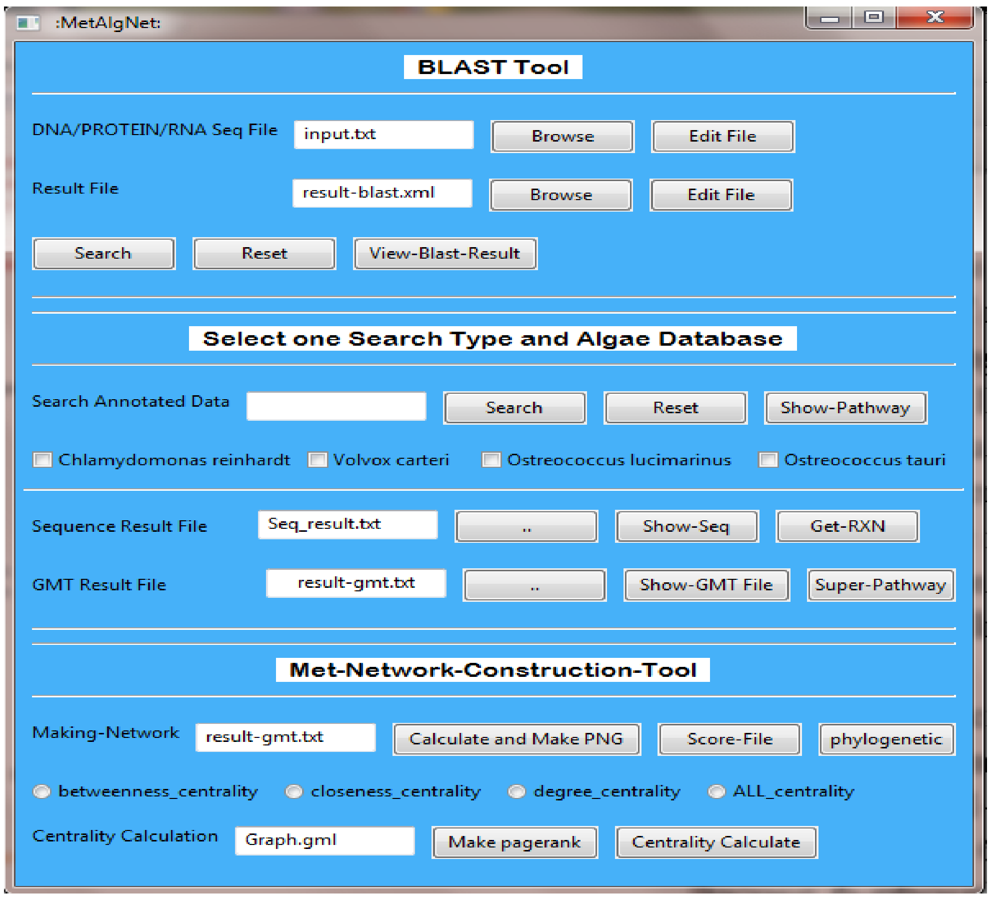


Figure 2. Overview of our in-house database architecture



DNA/PROTEIN/RNA Seq File: input.txt

Result File: result-blast.xml

Search Annotated Data:

□ Chlamydomonas reinhardt □ Volvox carteri □ Ostreococcus lucimarinus □ Ostreococcus tauri

Sequence Result File: Seq_result.txt

GMT Result File: result-gmt.txt

Making-Network: result-gmt.txt

○ betweenness_centrality ○ closeness_centrality ○ degree_centrality ○ ALL_centrality

Centrality Calculation: Graph.gml

Figure 3. The MetAlgNet tool

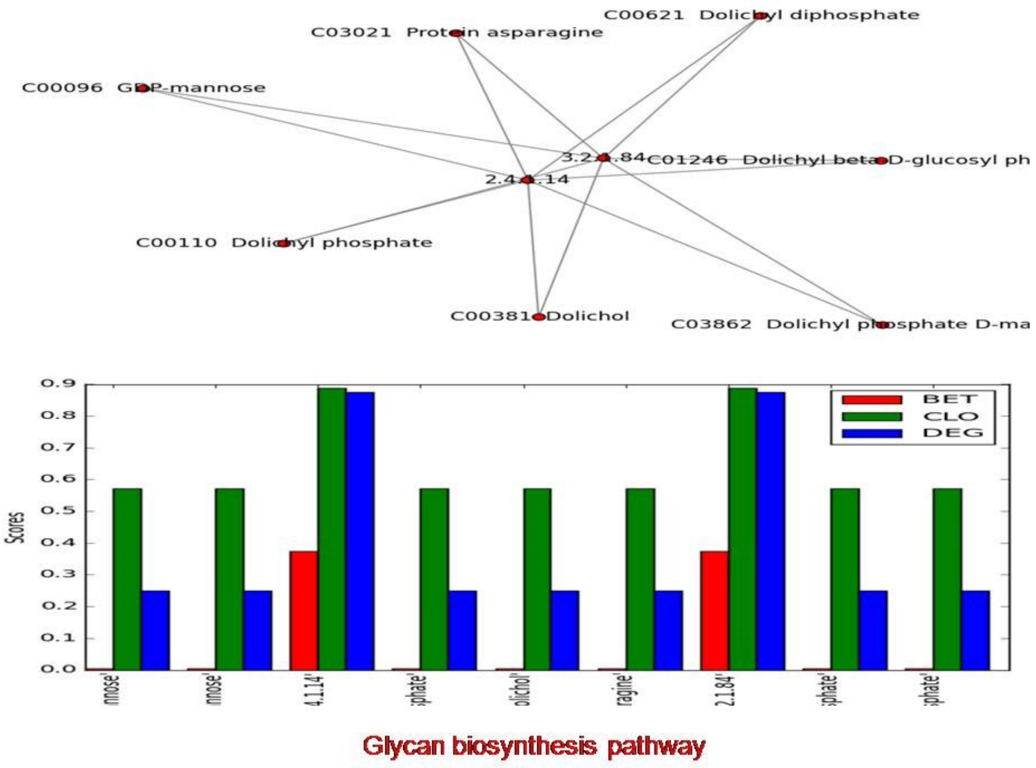


Figure 4. Sample result output

4. Conclusions

The creation of interactions network is through retrieval of data from multiple annotated databases, and the MetAlgNet software system allows visualization of the networks. Integrative text-based mining of the data from 24 various databases is facilitated by representing the annotated data as raw material for network construction, and visualizing the similarities using different python library.

The MetAlgNet-based data mining approach may facilitate discovery of novel or unexpected relationships among enzymes and metabolites, formulation of new hypotheses, data annotation, interpretation of new experimental data, and construction and validation of new network-based models of biological systems. Our approach takes advantage of connectivity of different annotated metabolic data of respective green algae in heterogeneous interactome network constructed by MetAlgNet, and shows that connectivity-based approach is superior to traditional pathway analysis. The findings from this study establish the applicability of our network analysis strategy, and support the hypothesis that modeling of local network topology dynamics can be used as an effective tool to study the activity of biological modules. Also, omics data are ever expanding and this poses challenges to updating and mining of data. The data warehousing approaches for data integration are really useful and effective from user point of view. It is not possible to completely avoid these problems, but by taking standards-based approach to data integration, we can minimize the problem of data integration. The integration approach is still found missing in online biological data available with different databases. It is better to develop databases which are interconnected with specific groups of organisms. The diversity of the data and the fact that not all data sources adapt the standards forces us to create our own schemas. We adapted a combination of multiple approaches in data integration. Although we imported all the databases to the local warehouse, the individual schemas were kept intact. We created an additional semantic mapping with the help of Python cursor and SQLite database to facilitate resolution of entities across databases, which often doesn’t need to change even when a new data source is added. The integration of data across databases and sophisticated queries are handled using Python programs. The technique of data integration is applicable more broadly to any organism for which we have large scale genome annotation data availability. As enzyme identifiers are the central entities to data integration in our method, data mining shows different interaction databases that use consistent identifiers.

Acknowledgments

We special thanks to Jignesh V. Smart for debug the **MetAlgNet** Program. This work is supported by the Councilof Scientific & Industrial Research (CSIR) - Senior Research Fellowship (SRF) grant no. 112511/2k11/1, India at G. H.Patel Post Graduate Department of Computer Science and Technology (GDCST), Sardar Patel University.

Conflicts of Interest

“The authors declare no conflict of interest.”

References

1. Chain, P.S.G.; Grafham, D.V.; Fulton, R.S.; FitzGerald, M.G.; Hostetler, J.; Muzny, D.; Ali, J.; Birren, B.; Bruce, D.C.; Buhay, C., *et al.* Genome project standards in a new era of sequencing. *Science (New York, N.Y.)* **2009**, *326*, 10.1126/science.1180614.

2. Hwang, D.; Rust, A.G.; Ramsey, S.; Smith, J.J.; Leslie, D.M.; Weston, A.D.; de Atauri, P.; Aitchison, J.D.; Hood, L.; Siegel, A.F., *et al.* A data integration methodology for systems biology. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 17296-17301.

3. Reeves, G.; Thornton, J.; Excellence, t.B.N.o. Integrating biological data through the genome. *Hum Mol Genet* **2006**, *15*, R81 - 87.

4. Christensen, C.; Thakar, J.; Albert, R. Systems-level insights into cellular regulation: Inferring, analysing, and modelling intracellular networks. *Systems Biology, IET* **2007**, *1*, 61-77.

5. NCBI Resource Coordinators. Database resources of the national center for biotechnology information. *Nucleic Acids Research* **2015**, *43*, D6-D17.

6. Kodama, Y.; Shumway, M.; Leinonen, R. The sequence read archive: Explosive growth of sequencing data. *Nucleic Acids Research* **2012**, *40*, D54-D56.

7. Pruitt, K.D.; Tatusova, T.; Maglott, D.R. Ncbi reference sequences (refseq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research* **2007**, *35*, D61-D65.

8. Palenik, B.; Grimwood, J.; Aerts, A.; Rouzé, P.; Salamov, A.; Putnam, N.; Dupont, C.; Jorgensen, R.; Derelle, E.; Rombauts, S., *et al.* The tiny eukaryote ostreococcus provides genomic insights into the paradox of plankton speciation. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 7705-7710.

9. Merchant, S.S.; Prochnik, S.E.; Vallon, O.; Harris, E.H.; Karpowicz, S.J.; Witman, G.B.; Terry, A.; Salamov, A.; Fritz-Laylin, L.K.; Maréchal-Drouard, L., *et al.* The chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* **2007**, *318*, 245-250.

10. Prochnik, S.E.; Umen, J.; Nedelcu, A.M.; Hallmann, A.; Miller, S.M.; Nishii, I.; Ferris, P.; Kuo, A.; Mitros, T.; Fritz-Laylin, L.K., *et al.* Genomic analysis of organismal complexity in the multicellular green alga volvox carteri. *Science* **2010**, *329*, 223-226.

11. Goodstein, D.M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N*., et al.* Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research* **2012**, *40*, D1178-D1186.

SciForum
Mol2Net

Green Processing of Nanoporous Biodegradable Carriers of Bioactive Agents for Pharmaceutical and Biomedical Applications

Jorge Santos [1], Pasquale del Gaudio [2], Mariana Landín [1] and Carlos A García-González [1,*]

[1] Departamento de Farmacia y Tecnología Farmacéutica, Facultad de Farmacia, Universidade de Santiago de Compostela, E-15782-Santiago de Compostela, Spain

[2] Department of Pharmacy (DIFARMA), University of Salerno, Via Giovanni Paolo II 132, Fisciano (SA) 84084, Italy

* Author to whom correspondence should be addressed; E-Mail: carlos.garcia@usc.es;
Tel.: +34-881-815-252; Fax: +34-981-547-148.

Published: 4 December 2015

Abstract: Pharmaceutical and biomedical industries demand simple, safe and reproducible processing methods thus urging the development of novel straightforward manufacturing approaches. The product manufacturing by the green processing of admixtures and end-product would avoid long and costly purification (downstream) steps. In this work, the supercritical fluid technology is used for the green processing of nanoporous carriers (aerogels) for bioactive agents [1,2]. Aerogels in the form of one micron-sized particles were processed and loaded with a model bioactive compound (ketoprofen). Results show that the carrier has excellent textural properties (specific surface area of 205 m^2/g) and a good loading capacity (8 wt.%) of the bioactive compound in the amorphous form. Release profile tests show the capacity of the carrier to modulate the drug release to the medium (pH 1.2 and 7.4). The resulting material can be potentially incorporated in the formulation of several pharmaceutical and biomedical products.

Keywords: supercritical fluid technology; green processing; aerogel; nanoporous material; bioactive agent carrier

1. Introduction

Nanotechnology research is facing the challenge of reaching mass markets in the near future by providing innovative and sustainable solutions to the demands of the society. Indeed, smart, reproducible and environmentally friendly solutions for nanotechnology are been sought in the life sciences field to be aligned with current and forthcoming environmental regulations. In this context, pharmaceutical and biomedical industries are focusing part of their research and innovation on the efficient use of raw materials and on the development of novel green processes with low environmental impact.

Supercritical fluid (SCF) technology emerges as a green processing method of nanomaterials. The possibility of modulation of the physicochemical properties of SCF (e.g., density, viscosity, diffusivity, dielectric constant) allows the processing with a fine control of the nanomaterial end properties. Moreover, solvent-free nanomaterials are obtained using this technology without need of downstream processes. Supercritical CO_2 is the most common choice of SCF attending to its low cost, health and safety properties (innocuous, non-flammable, GRAS substance) and the mild operating conditions needed to reach its critical point (31.1ºC, 73.8 bar). Finally, CO_2 used in supercritical processes can be potentially reused in a closed loop without contribution to greenhouse gas emission.

Aerogels are a special class of nanostructured materials obtained using SCF technology [1]. Aerogels have high porosity (>90%) in the mesoporous range and with outstanding textural properties (specific surface areas higher than $150 m^2/g$). These nanomaterials are obtained by the supercritical extraction of the solvent contained in gels [2], the unique technique able to preserve the porous gel network in the dry form. Among the aerogel types, polysaccharide aerogels are especially attractive for life sciences purposes since it combines the above-mentioned intrinsic properties of aerogels with the availability, renewability, low toxicity and usual biodegradability and biocompatibility of polysaccharides [3, 4]. Finally, the preparation of aerogels in the form of fine powder would facilitate the incorporation of the material as an admixture in pharmaceutical and biomedical formulations [5, 6].

In this work, starch aerogels in the form of one-micron particles are processed by a combination of emulsion-gelation and supercritical drying techniques. Morphological and textural properties of aerogels are evaluated. Finally, aerogel particles are loaded with an anti-inflammatory drug (ketoprofen) by supercritical impregnation and the drug release from the aerogel is evaluated using simulated gastric (pH 1.2) and blood (pH 7.4) conditions to assess its behavior for different administration routes.

.

2. Results and Discussion

Starch aerogel particles were obtained as a fine white powder with a particle size distribution in the 200 nm to 1.5 μm range and a mean particle size of 1.110±0.149 μm. Aerogel powder was free-flowing with an angle of repose of 34º and a compressibility of 35%.

SEM pictures highlight the presence of spherical microparticles (Fig. 1,top). Primary particles were loosely joined together forming agglomerates likely due to the presence of emulsifier (PGPR) remnants.

The nanoporous structure of the aerogel particles processed by supercritical drying can be observed at the highest magnification (Fig. 1,bottom). Excellent textural properties of the aerogel powder were obtained with specific surface areas of 205 m^2/g (BET-method), specific pore volume of 0.98 cm^3/g and mean pore size of 17 nm (BJH-method). In contrast, specific surface areas below 5 m^2/g was obtained for the same gel particles dried under ambient conditions, highlighting the high performance of the supercritical drying method.

The poor solubility of ketoprofen in water and its good solubility in supercritical CO_2 suggested the use of supercritical impregnation of aerogels as the best choice for the drug loading. The use of this solvent-free method led to a high ketoprofen loading (*ca.* 8 wt.%) in the aerogel matrix. Supercritical impregnation of drugs in aerogels usually leads to the adsorption of the drug in the amorphous form [6]. The ketoprofen amorphization and its loading in the aerogel matrix significantly increased the drug release rate with respect to the raw ketoprofen in both simulated physiological media (pH 7.4 and pH 1.2) (Fig. 2). Ketoprofen loaded in the aerogels was released faster at pH 7.4 than at pH 1.2 as also observed with the raw crystalline drug. This behaviour was related to the pK_a value of ketoprofen (4.4) in-between the two pH media studied thus leading to different ionic forms of the drug in the two simulated physiologicl conditions studied. After an initial burst during the first 15 min, the release profile of ketoprofen from the aerogel was almost linear at pH 1.2 during the release period studied (Fig. 2a). At pH 7.4, a two-stage release profile is observed for ketoprofen loaded in the aerogels (Fig. 2b): a fast release profile of *ca.* 60% of the ketoprofen payload during the first 2 h is followed by a slower release in the following hours.

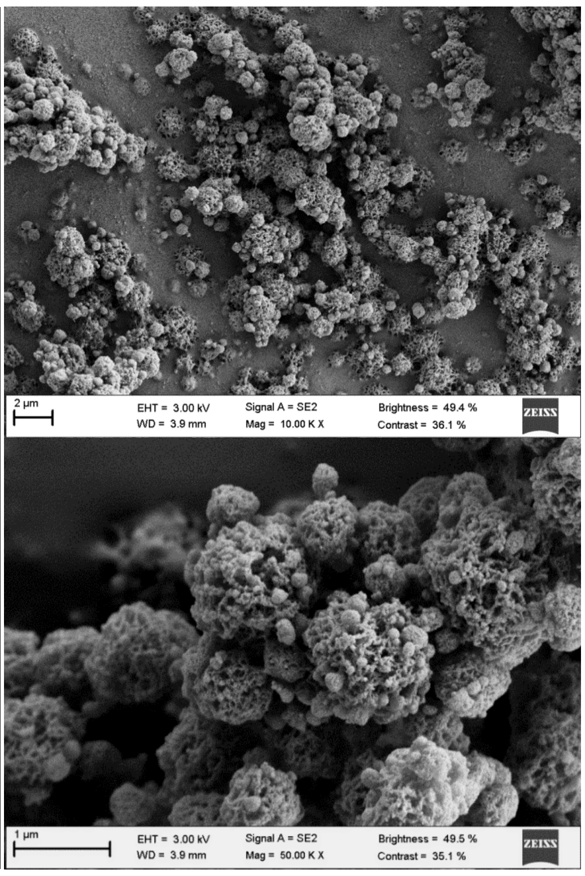


Figure 1. SEM images of starch aerogel microspheres (top). Higher magnifications show the nanoporous structure of the particles (bottom).

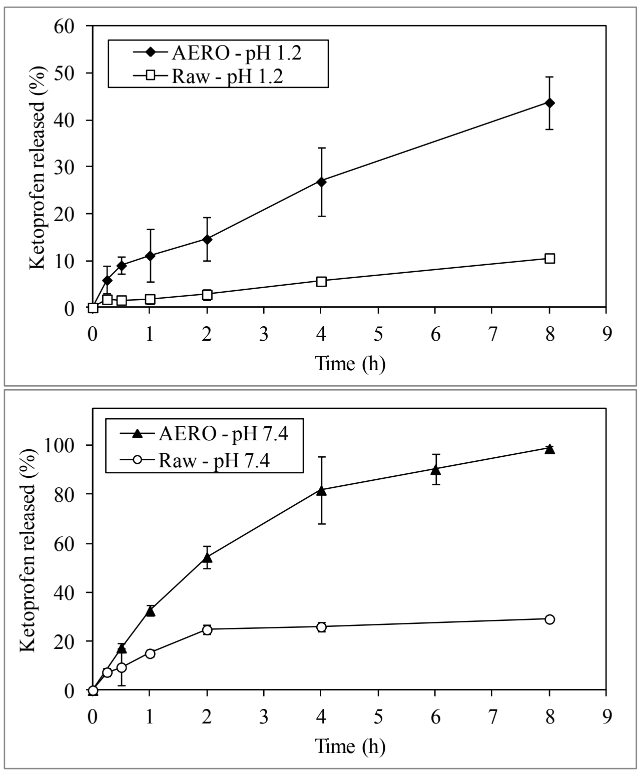


Figure 2. *In vitro* release profiles of ketoprofen from starch aerogel microspheres (dark markers) at two different pH media: pH 1.2 (top) and pH 7.4 (bottom). Release profiles of the raw ketoprofen (blank markers) are also plotted for the sake of comparison

3. Materials and Methods

3.1. Reagents and chemicals

Native corn starch (Starch Amylo N-460; 52.6% amylose content) was from Roquette (France). Ketoprofen (Raw, 99.7% purity) was provided by Acofarma (Spain). Ethanol (99.8% purity) was from Omnilab (Germany). Polyglycerol polyricinoleate (PGPR) was from Palsgaard (Denmark). Paraffin oil was provided by Panreac (Spain). CO_2 (99.8%) was obtained from Praxair (USA).

3.2. Methods

3.2.1. Aerogel preparation

Starch gel microspheres were prepared by the method of emulsification-gelation [4]. A water-in-oil emulsion (W/O) was prepared from a mixture of paraffin oil (continuous phase)/aqueous starch solution 15% (w/w) (dispersed phase) in the 3:1 weight ratio and 3% (w/w) of emulsifier. The mixture obtained was autoclaved (Trade Raypa Steam Sterilizer, USA) at 121°C and 1.1 bar for 20 min. After sterilization and partial cooling (95 °C), the mixture was homogenized using an ultrasound probe (Branson Digital Sonifier; Mexico) and kept for 48 h at 4 °C for starch retrogradation. The resulting starch hydrogel dispersion in paraffin was centrifuged to separate the hydrogel from the oil phase. Starch hydrogel microspheres were transferred to a fresh solution of ethanol for solvent exchange, which was replaced daily for two days until complete removal of water to get an alcogel.

Aerogel microspheres (AERO) were obtained by the $scCO_2$-assisted drying of the starch gels with a CO_2 flow of 5-7 g/min during 3.5 h in a 100-mL autoclave (TharSFC, USA). The autoclave was heated to 40°C and the pressure increased progressively until 120 bar.

Particle size of the aerogel particles were measured by the dynamic light scattering method (Malvern Zetasizer Nano ZS, UK). Textural properties were obtained by a low-temperature N_2 adsorption–desorption analysis (ASAP 2000 Micromeritics Inc., USA).

3.2.2. Aerogel impregnation and release studies

Aerogels were impregnated with ketoprofen by the supercritical CO_2-assisted impregnation process at 40ºC and 150 bar for 11 h in the batch mode.

Release of ketoprofen was performed using dialysis membranes (Float-A-Lyzer, MWCO 8-10 KD, V 1ml; USA) in 100 mL of PBS solution (pH 7.4) and 0.1N HCl aqueous solution (pH 1.2) as release media. Release studies were carried out at 37ºC and at the stirring rate of 100 rpm for 8 h using 10 mg of sample. 2 ml-aliquots are collected at selected times (0.5, 1, 2, 4 and 8 h) and refilled with fresh medium. Ketoprofen concentrations were spectrophotometrically determined by UV/Vis at the wavelength of λ=258 nm.

4. Conclusions

Green processing approaches have been used for the preparation of nanoporous starch (aerogel) microparticles. One-micron aerogel microspheres with excellent textural properties were obtained. Supercritical impregnation technique allowed the loading of the aerogels with drugs (8 wt.% for ketoprofen). Ketoprofen release results highlighted the significant effect of the drug incorporation into starch aerogels with respect to the raw crystalline drug in enhancing the ketoprofen dissolution properties at the two experimental pH conditions tested (1.2 and 7.4). The drug-loaded aerogels can be potentially incorporated in the formulations of several pharmaceutical and biomedical products to improve the drug release profile.

Acknowledgments

Work funded by MICINN (SAF2012-39878-CO2-01) and MINECO (EUIN2015-62748). C.A. García-González acknowledges MICINN for the financial support through the Juan de la Cierva Fellowship Programme (JCI-2012-12705).

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. García-González, C.A., M. Alnaief, and I. Smirnova, *Polysaccharide-based aerogels - Promising biodegradable carriers for drug delivery systems.* Carbohydrate Polymers, 2011. 86(4): p. 1425-1438.
2. García-González, C.A., et al., *Supercritical drying of aerogels using CO 2: Effect of extraction time on the end material textural properties.* Journal of Supercritical Fluids, 2012. 66: p. 297-306.
3. Del Gaudio, P., et al., *Design of alginate-based aerogel for nonsteroidal anti-inflammatory drugs controlled delivery systems using prilling and supercritical-assisted drying.* Journal of Pharmaceutical Sciences, 2013. 102(1): p. 185-194.
4. García-González, C.A., et al., *Preparation of tailor-made starch-based aerogel microspheres by the emulsion-gelation method.* Carbohydrate Polymers, 2012. 88(4): p. 1378-1386.
5. García-González, C.A., A. Concheiro, and C. Alvarez-Lorenzo, *Processing of materials for regenerative medicine using supercritical fluid technology.* Bioconjugate Chemistry, 2015. 26(7): p. 1159-1171.
6. García-González, C.A., et al., *Polysaccharide-based aerogel microspheres for oral drug delivery.* Carbohydrate Polymers, 2015. 117: p. 797-806.

SciForum
Mol2Net

Influence of Synthetic Conditions in the Hydrothermal Preparation of TiO_2 Nanotubes

Izaskun Gil de Muro [1,2] *, Amaia Ereño [1], M. Insausti [1,2].

[1] Inorganic Chemistry Department, FCT-ZTF, UPV/EHU, Bº Sarriena, 48940 Leioa, Spain; E-Mail: izaskun.gildemuro@ehu.eus; aereno002@ikasle.ehu.eus, maite.insausti@ehu.eus.

[1] BCMaterials. Parque Científico y Tecnológico de Bizkaia, 48160 Elexalde Derio, Spain.

* Author to whom correspondence should be addressed; E-Mail: izaskun.gildemuro@ehu.eus;
Tel.: +34 946015990; Fax: +34 946013500.

Published: 4 December 2015

Abstract: TiO_2 nanotubes have been synthesized using a 2-step strategy that involves the hydrothermal preparation of intermediate titanates followed by a subsequent thermal/hydrothermal treatment of these species to produce the oxide. In the first step the influence of parameters such as temperature, pH or reaction time on the composition, structure and morphology of the intermediate species has been studied. Then we have examined how temperature and the presence of surfactants may affect the composition, structure, morphology and particle size of the titanium dioxide obtained in the second thermal/hydrothermal treatment. In particular, we have studied the effects that the presence of CTAB has upon the morphology of the final product. Both intermediate and final species have been studied by means of X-ray diffraction, transmission electron microscopy, IR spectroscopy and thermogravimetric analysis. This way we have been able to identify $NaTi_3O_6(OH).2H_2O$ and $(TiO_2)_x(H_2O)_y$ as the intermediate titanate species and rutile and anatase as the final TiO_2 polymorphs. Finally, it is worth mentioning the preparation of spindle- or oval-shaped anatase, obtained via hydrothermal synthesis in the presence of CTAB.

Keywords: TiO_2; hydrothermal synthesis; nanotube.

1. Introduction

Titanium dioxide exists as three polymorphs: anatase, rutile and brookite, and although rutile is the thermodinamically stable phase, in the nanoscale anatase becomes very stable too [1]. As a wide-band gap semiconductor, TiO_2 has been widely used in sunscreens, paints and tooth-pastes and, today, especially in photocatalysis [2]. For many of these applications, it is of great importance to maximize the specific surface to achieve a maximum efficiency. In particular, it has been found that in photocatalytic applications, nanotubes and nanorods -that is, 1D nanostructures- may allow for a much higher control of the chemical and physical behaviour [3]. Indeed, by diminishing dimensions to the nanoscale, not only the specific surface area increases significantly but also the electronic properties may change considerably [4].

1D titanium dioxide nanostructures may be obtained by various routes such as sol–gel, template-assisted methods, hydro/solvothermal synthesis or by electrochemical processes [5, 6]. Among them, hydrothermal synthesis has become a promising chemical strategy. This is usually a two-step route that involves, first, the generation of alkali titanate nanofibers by hydrothermal reaction in alkali solution followed by exchange of alkali ions with protons [7], and the subsequent thermal dehydration reactions at high temperature in air. Less frequently, hydrothermal reactions may also be performed to produce the final TiO_2 nanotubes or nanorods [8].1D morphology is already achieved in the first hydrothermal process when alkali titanates are formed, and it is based on the exfoliation of TiO_2 crystal planes in the alkali environment and then stabilization due to the insertion of Na^+ cations between the exfoliated planes. The formed nanolayer sheets are then rolled into tubes during cooling or in the acid treatment [9]. Finally, thermal treatments at 350-450 ºC tend to lead anatase, which seems to favour a faster electronic transport than the rutile phase [10].

Here, we present synthesis of TiO_2 1D nanostructures *via* a 2-step route. Initially, the influence of parameters such as temperature or reaction time on the composition, structure and morphology of the $NaTi_3O_6(OH).2H_2O$ and $(TiO_2)_x(H_2O)_y$ intermediate titanates has been studied. Then, we have examined how temperature can affect the final TiO_2 rutile/anatasa composition in the dehydration process. Finally, we have performed an alternatively hydrothermal route starting from Na-titanate phase in the presence of the CTAB surfactant at different pH values. This way, a spindle- or oval-shaped anatase product was obtained.

2. Results and Discussion

Hydrothermal syntheses were carried out in order to study the influence of temperature and reaction time. Thus, synthesis at 200 ºC for 6, 24 and 48 h were performed. Nanotubes start to appear after 6 h of hydrothermal heating, but the reaction was not complete and crystallinity of the product was low. The products obtained after 24 and 48 h were very similar, both in composition and degree of crystallinity. Syntheses at 170ºC were performed during 24 h but low crystallinity was obtained again. For this reason, the syntheses were carried out at 200 ºC for 48 h (6 sample).

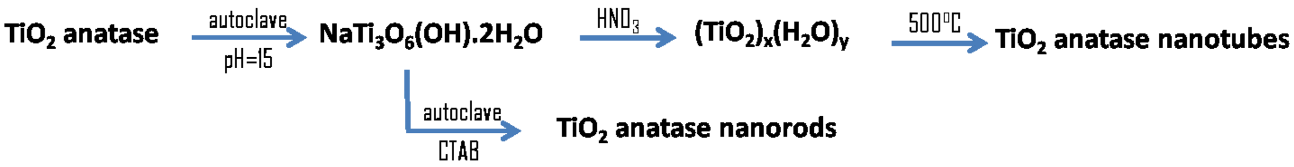
XRD patterns of the most representative as synthesized titanates shown in Fig. 1 indicate the presence of $NaTi_3O_6(OH).2(H_2O)$ [11] and $(TiO_2)_x(H_2O)_y$ [12] phases before and after HNO_3 treatment, respectively. Subsequent thermal treatment in air at 400-500°C leads to the formation of anatase and rutile TiO_2 phases (JCPDF 21-1272 and JCPDF 21-1276) together with some $NaTi_8O_{13}$ (JCPDF 48-0523) impurities. At the same time, after the second hydrothermal treatment anatase phase was also obtained.

Thermogravimetric analysis of titanate samples shows mass losses of about 20 and 15%

in good agreement with the presence of H_2O and some Na_2CO_3 (formed in the basic conditions due to absorption of atmospheric CO_2 [13]) in Na-titanate samples and only H_2O in H-titanate ones. This was confirmed by IR spectroscopy (Figure 2).

TEM images of some representative samples are shown in Figure 3. Nanotubes were already formed in the first hydrothermal step and neither the acidic nor the thermal nor the hydrothermal processes alter significantly the morphology of the samples. Both Na- and H-titanate samples are very similar, 40-80 nm wide, although they widen to 70-170 nm after thermal or hydrothermal treatments when deintercalation and dehydration take place and lead to TiO_2. HTEM images of titanates shown in Figure 3 reveal lattice fringes with plane spacings of about 11 Å and 10 Å corresponding to the interlayer distances in the 2D crystal structures of $NaTi_3O_6(OH).2(H_2O)$ and $(TiO_2)_x(H_2O)_y$. Both of them exhibit similar crystal structures with packed planes formed by titanium oxide octahedra and Na^+, OH^- H^+ and H_2O as inserted species between them [11, 12].

Some authors suggest that the mechanism of formation of the tube shape is based on a topochemical process of exfoliation of TiO_2 crystal planes in the alkali medium [14], after which nanolayer sheets start rolling into tubes during cooling. Na^+ ions are then exchanged by H^+ ions by washing products in acid solutions to form layered hydrogen titanate, from $Na_{2-x}H_xTi_3O_7.2H_2O$ to $H_2Ti_3O_7.yH_2O$ [15, 16, 17]. Finally, the H-titanate transforms into anatase through a dehydration reaction. This is accompanied by an *in situ* rearrangement of the structural units to give 1D TiO_2 nanostructures [8]. In this work, not only traditional heat treatments of H-titanates to obtain anatase TiO_2 nanotubes, but also hydrothermal treatments of Na-titanates in presence of CTAB as surfactant at different pH values were performed. Heating at 400ºC and 500ºC in air led to the formation of anatase and anatase/rutile nanotubes, respectively, about 70-170 nm wide and without significant morphological changes. However, some impurities of $NaTi_8O_{13}$ oxide were also found in XRD, indicating that the exchange of Na^+ ions was not completely achieved. At the same time, hydrothermal processes led to the formation of anatase as the main phase but the morphology of the products was found to be slightly altered depending on pH. Thus, the synthesis at pH=2 led to obtain anatase porous nanotubes, very similar in thickness and morphology to their precursors, but a considerable amount of intermediate titanate remained unaltered. The non-acidified synthesis (pH=8.5) led to a purer and more crystalline product. It seems that CTAB in a slight alkali environtment favoured the extraction of Na^+ ions previous to the stacking of the TiO_2 layers, as thicker and denser nanostructures of 70-170 nm were found. It is possible that the CTAB surfactant made the exfoliation and later rearrangement of oxide layers in Na-titanates easier than in H-titanates, as H^+ could be more strongly bounded to the TiO_6 octahedra. The morphology also changed. Oval- or spindle-shaped anatase was obtained suggesting a non-uniform stacking mechanism. Both hydrothermally synthesized samples did, however show similar microstructures formed by aggregated nanofibres, although these were more densely stacked in the oval-shaped oxides**.**



Scheme 1. Scheme of the syntheses of TiO2 nanotubes

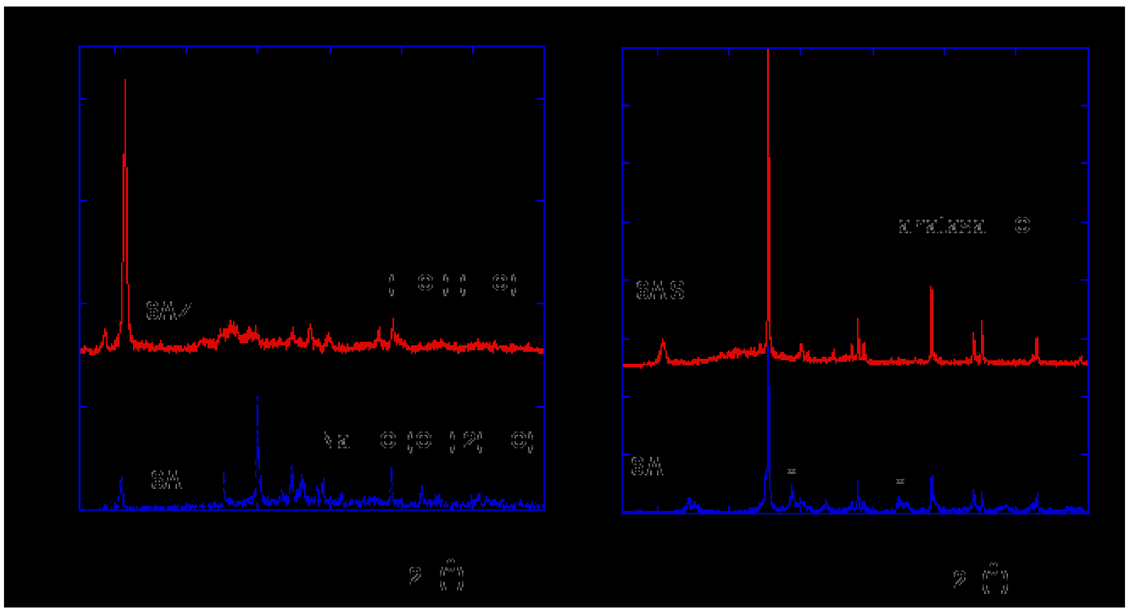


Figure 1. XRD patterns of 6A (after first hydrothermal treatment), 6AZ (acidified sample), 6AZT (after thermal treatment) and 6AS (after second hydrothermal treatment with CTAB).

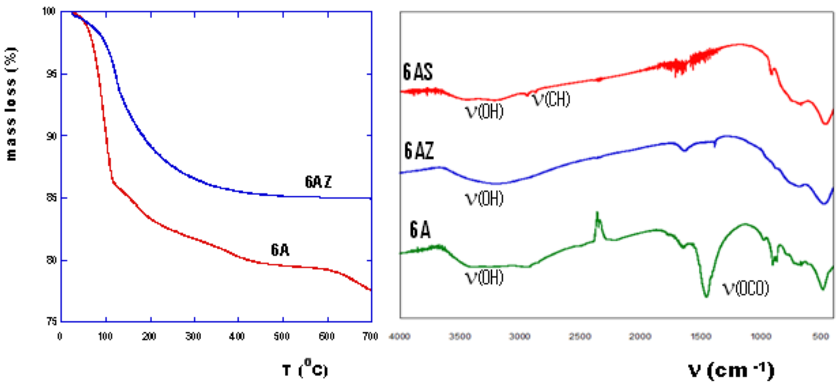


Figure 2. TG curves and IR spectra of synthesized samples.

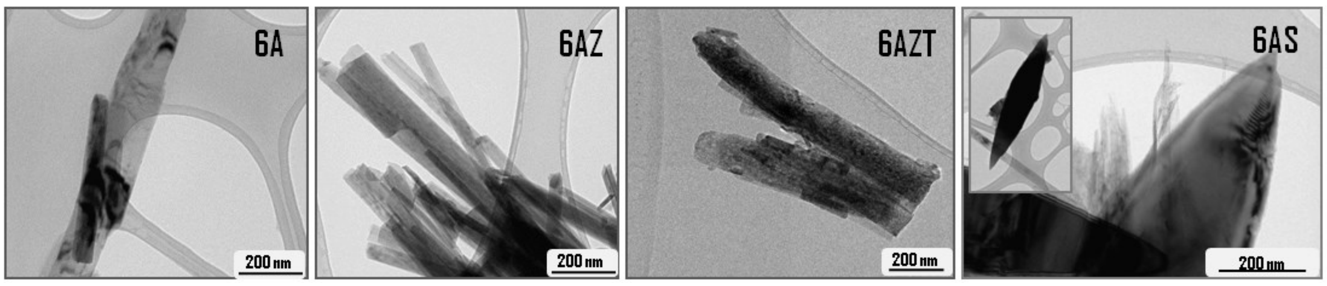


Figure 3. TEM images of 6A (after first hydrothermal treatment), 6AZ (acidified sample), 6AZT (after thermal treatment) and 6AS (after second hydrothermal treatment with CTAB).

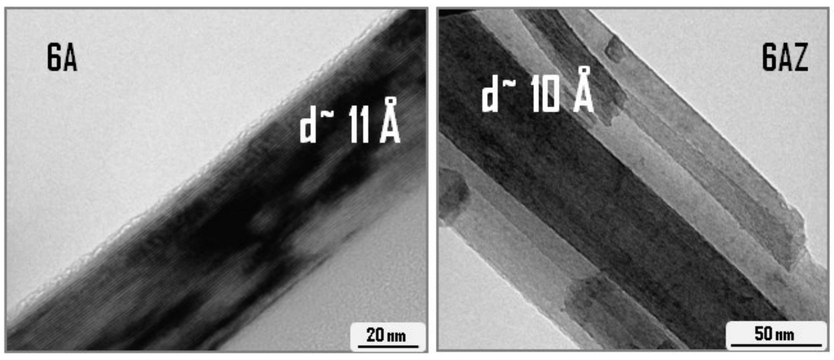


Figure 4. HRTEM images of 6A and 6AZ, Na- and H-titanate samples.

3. Materials and Methods

Sodium hidroxide (97%), titanium (IV) oxide (99,8%), cetyltrimethylammonium bromide (CTAB) (99%), nitric acid (65%) and acetone were purchased from Sigma-Aldrich and used as received without further purification. Hydrochloric acid (37%) was purchased from Panreac.

3.1. Synthesis of sodium and hydrogen titanates nanorods. 0,4 g commercial TiO_2 was dispersed in 20 mL of 10M NaOH with magnetic stirring for 30 minutes. The white suspension was then transferred into a Teflon-lined autoclave of 25 mL capacity. After being heated at 170-200 °C for 6-48 h, the autoclave was naturally cooled to room temperature. The resulting products were filtered and washed several times with diluted HCl [18, 19] and dried at 80°C for 2 h.

3.2. Synthesis of sodium and hydrogen titanates nanorods. For the ion-exchange, the sodium titanate samples were immersed into a 0.2 M HNO_3 solution for 6 h and 1-2 hour in an ultrasonic bath [20]. After that, they were washed with deionized water for several times, filtered and dried at 80°C for 2 h.

Finally, different hydrothermal [21] and thermal treatments were carried out. 0,3 g of previously synthesised Na-titanate were mixed with 0,5 g CTAB in 20 mL of distilled water and transferred again into a Teflon-lined autoclave and heated at 200°C-tan for 24 h. In one case, before heating, the pH was acidified by adding HCl 0,1M. In the other hand, some H-titanates were heated at 400-500 °C for 4-5 h.

3.4. Characterization. IR spectra (400-4000 cm^{-1}) were recorded on a MATTSON FTIR 1000 spectrophotometer with samples prepared as KBr pellets. The powder X-ray diffraction (XRD) pattern was taken using a Philips PW1710 diffractometer. Thermogravimetric measurements were performed in a Netzsch STA 449C thermogravimetric analyzer. Crucibles containing 9 mg of sample were heated at $10ºC.min^{-1}$ under dry argon. For microstructure analysis, powders were dispersed in acetone, dropped-cast onto copper grid and examined using a CM200 transmission electron microscope.

4. Conclusions

1D anatase samples with different shapes had been obtained by a 2-step method that implied an initial hydrothermal treatment in alkali environment followed by acid exchange and final thermal/hydrothermal heating. 1D nanostructures were already achieved in the first stage during the formation of sodium titanates and remained without significant morphological alteration, except a slight broadening, until the formation of anatase TiO_2 nanotubes. Exceptionally, a second hydrothermal heating in the presence of CTAB of the alkali intermediates led to an oval-shaped anatase sample where TiO_2 fibres seemed to be more densely stacked.

Acknowledgments

This work has been supported by institutional funding from MINECO and Basque Government under Projects MAT2013-41128-12 and IT-570-13, respectively.

Author Contributions

This text is based on "**Final Degree Project"** made by E. Ereño.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Hanaor, D.A.H.; Sorrell, C.C. Review of the anatase to rutile phase transformation. *J. Mater. Sci*. **2011**, *46*, 855–874
2. Abdullah, N; Kamarudin, S.K. Titanium dioxide in cell fuel technology: an overview. *J. Power Sources* **2015**, *278*, 109-118.
3. Roy, P.;Beger, S.; Schmuki, P. TiO_2 nanotubes; synthesis and applications. *Angew. Chem. Int. Ed.* **2011**, *50*, 2904-2939.
4. Henglein, A., Small-particle research: physicochemical properties of extremely small colloidal metal and semiconductor particles. *Chem. Rev*. **1989**, *89*, 1861-1873.
5. Kasuga, T.; Hiramatsu, M.; Hoson, A.; Sekino, T.; Niihara K., Formation of Titanium Oxide Nanotube. *Langmuir* **1998**, *14*, 3160;
6. Kasuga, T.; Hiramatsu, M.; Hoson, A.; Sekino, T.; Niihara K. Titania Nanotubes Prepared by Chemical Processing. *Adv. Mater*. **1999**, *11*, 1307-1311.
7. Yang, D.; Liu, H.; Zheng, Z.; Sarina, S.; Zhu H. Titanate-based adsorbents for radioactive ions entrapment from water. *Nanoscale*, **2013**, *5*, 2232-2242.
8. Sui, X.L.; Wang, Z.B.; Li, C.H.; Zhang, J.J.; Zhao, L. Effect of pH value on $H_2Ti_2O_5/TiO_2$ composite nanotubes as Pt catalyst support for methanol oxidation. *J. Power Sources* **2014**, *271*, 196-202.
9 . Suzuki, Y.; Yoshikawa, S. Synthesis and Thermal Analyses of TiO_2-Derived Nanotubes Prepared by the Hydrothermal Method. *J. Mater. Res.* **2004**, *19(4),* 982-98
10. Aradi, B. ; Deak, P.; Huy, H.A.; Rosenauer, A., Frauenheim, T. Role of Symmetry in the Stability and Electronic Structure of Titanium Dioxide Nanowires, *J. Phys. Chem. C* **2011**, *115*, 18494–18499.
11. Andrusenko, I.; Mugnaioli, E.; Gorelik, T.E.; Koll, D.; Panthöfer, M.; Tremel,W.; Kolb, U. Structure analysis of titanate nanorods by automated electron diffraction tomography. *Acta Crystallogr B* **2011**, *67(3),* 218-25.
12. Yang, J.; Jin, Z.; Wang, X.; Li, W.; Zhang, J.; Zhang, S.; Guo, X.; Zhang, Z. Study on composition, structure and formation process of nanotube $Na_2Ti_2O_4(OH)_2$. *Dalton Trans.* **2003**, 3898–3901.
13. Andrusenko, I.; Mugnaioli, E.; Gorelik, T.E.; Koll, D.; Panthöfer, M.; Tremel,W.; Kolb, U. Structure analysis of titanate nanorods by automated electron diffraction tomography. *Acta Crystallogr B* **2011**, *67(3),* 218-25
.14 Lee, L.; Park, H.; Paik, U.; Han, T.H. Exfoliation of titanium dioxide powder into nanosheets using hydrothermal reaction and its reassembly into flexible paper for thin-film capacitors. *J. Solid State Chem*. **2015**, *224*, 76-81.
15. Du, *G*. H. ; Chen, Q.; Che, R. C.; Yuan, Z. Y.; Peng, L. M. Preparation and structure analysis of titanium oxide nanotubes. *Appl. Phys. Lett.* **2001**, *79, 3702 – 3704.*

16. Suzuki, Y.; Yoshikawa, S. Synthesis and Thermal Analyses of TiO_2-Derived Nanotubes Prepared by the Hydrothermal method. *J. Mater. Res.* **2004***, 19,* 982.
17. Peng, C.W.; Richard-Plouet, M.; Ke, T.-Y.; Lee, C.-Y; Chiu, H.-T.; Marhic, H.; Puzenat, E.; Lemoigno, F.; Brohan, L. Chimie douce Route to sodium hydroxo titanate nanowires with modulated structure and conversion to highly photoactive titanium dioxides. *Chem. Mater*. **2008**, *20*, 7228-7236.
18. Yuxiang, L.I.; Mei, Z.; Min, G.; Xidong, W. Hydrothermal growth of well-aligned TiO2 nanorod arrays: Dependence of morphology upon hydrothermal reaction conditions. *Rare Metals* **2010**, 29(3), 286.
19. Yu, J.; Wang, G.; Cheng, B.; Zhou, M. Effect of hydrothermal temperature and time on the photocatalytic activity and microstructures of bimodal mesoporous TiO_2 powder. *Applied Catalysis B: Environmental* **2007**, *69*, 171-180
20. Zhu, K.; Gao, H.; Hu, G.; Shi, Z. A rapid transformation of titanate nanotubes into single-crystalline anatase TiO_2 nanocrystals in supercritical water. *J. of Supercritical Fluids* **2013**, *83*, 28-34.
21. Dorian, A.H.; Hanaor, C.; Sorrel, C. Review of the anatase to rutile phase transformation. *J. Mater. Sci*. **2011,** *46*, 855-874.

SciForum
Mol2Net

Bio-AIMS Chemoinformatics Web Tools for Proteins

Cristian R. Munteanu [1], Humberto González-Díaz [2,3], Carlos Fernandez-Lozano [1], José Antonio Seoane Fernández [4], José M. Vázquez-Naya [1,*], Mabel Loza [5] and Alejandro Pazos [1,6]

[1] RNASA-IMEDIR Group, Faculty of Computer Science, University of A Coruna, 15071 A Coruña, Spain

[2] Department of Organic Chemistry II, Faculty of Science and Technology, University of the Basque Country UPV/EHU, 48940 Leioa, Vizcaya, Spain

[3] IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Vizcaya, Spain

[4] Stanford Cancer Institute, Stanford University, C.J. Huang Building, 780 Welch Road, Palo Alto, CA 94304, USA

[5] Grupo BioFarma-USEF, Departamento de Farmacología, Facultad de Farmacia, Campus Universitario Sur s/n, 15782 Santiago de Compostela, Spain

[6] Instituto de Investigación Biomédica de A Coruña (INIBIC), Complexo Hospitalario Universitario de A Coruña (CHUAC), 15006 A Coruña, Spain

E-Mails: crm.publish@gmail.com, gonzalezdiazh@yahoo.es, carlos.fernandez@udc.es, seoane@stanford.edu, jmvazquez@udc.es, mabel.loza@usc.es, apazos@udc.es

* Correspondent author: José M. Vázquez-Naya, Information and Communication Technologies Department, Faculty of Computer Science, University of A Coruna, 15071 A Coruña, Spain; E-Mail: jmvazquez@udc.es; Tel.: +34-981-167-000; Fax: +34-981-167-160.

Published: 4 December 2015

Abstract: The peptide biological screening represents a difficult task due to the complexity of the amino-acid sequences. One solution is the encoding of the molecular information using complex networks or graphs of the peptides into QSAR-like models in Web tools. Bio-AIMS contains free Web tools on an Artificial Intelligence Model Server in Biosciences: http://bio-aims.udc.es/TargetPred.php. These in silico peptide screening tools are implementing models to predict different protein activities, drug – protein and protein – protein interactions. The inputs are using 3D protein structures or 1D peptide amino acid sequences and the SMILES formulas for drugs, and the classification models are based on Machine Learning techniques. The Web tools are implemented using Python, PHP and XHTML programming languages.

Keywords: molecular information, Machine Learning, protein graphs, Python scripts, QSAR models, Web tools

1. Introduction

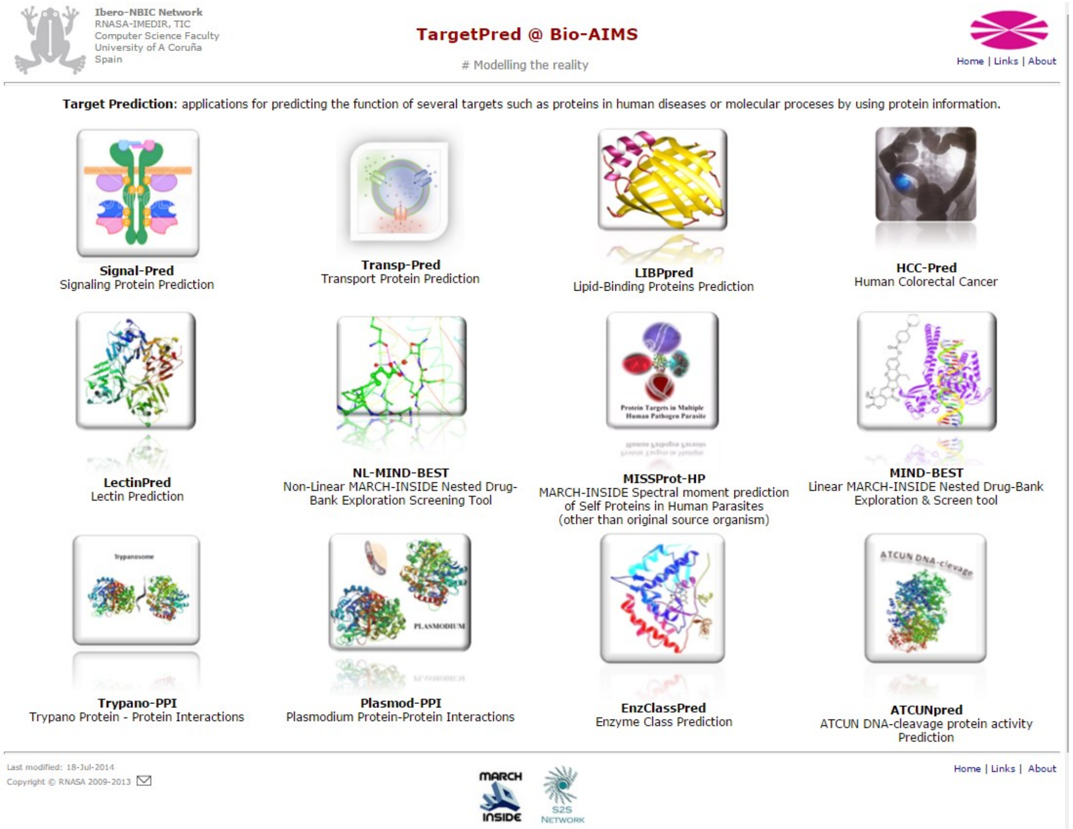
The *in silico* screening methods are very important in Drug Development or proteomics. The theoretical screening is a fast and low cost option to filter the large number of molecules or macromolecules for a specific biological action or chemical property.

These methods are proposing prediction models such as Qualitative Structure-Activity/Property Relationships (QSAR/QPDR), relations between the molecular structure and its activity [1,2]. Extended publications are using small molecule QSAR models. The current collection of QSAR-like models implemented into Web tools are extended the QSAR methodology to macromolecules [3].

2. Results and Discussion

The collection of free Web tools of Target Prediction section of Bio-AIMS server are implementing 12 classifiers (http://bio-aims.udc.es/TargetPred.php, see Figure 1):

- **Signal-Pred**: Signaling Protein Prediction [4]
- **Transp-Pred**: Transport Protein Prediction [5]
- **LIBPpred**: Lipid-Binding Proteins Prediction [6]
- **HCC-Pred**: Human Colorectal Cancer Protein Prediction [7]
- **LectinPred**: Lectin Prediction [8]
- **NL-MIND-BEST**: Non-Linear MARCH-INSIDE Nested Drug-Bank Exploration Screening Tool [9]
- **MISSProt-HP**: MARCH-INSIDE Spectral moment prediction of Self Proteins in Human Parasites (other than original source organism) [10]
- **MIND-BEST**: Linear MARCH-INSIDE Nested Drug-Bank Exploration & Screen tool [11]
- **Trypano-PPI**: Trypano Protein - Protein Interactions [12]
- **Plasmod-PPI**: Plasmodium Protein-Protein Interactions [13]
- **EnzClassPred**: Enzyme Class Prediction [14]
- **ATCUNpred**: ATCUN DNA-cleavage protein activity Prediction [15].


Ibero-NBIC Network
RNASA-IMEDIR, TIC
Computer Science Faculty
University of A Coruña
Spain
TargetPred @ Bio-AIMS
# Modelling the reality
Home | Links | About
Signal-Pred
Signaling Protein Prediction
Transp-Pred
Transport Protein Prediction
LIBPpred
Lipid-Binding Proteins Prediction
HCC-Pred
Human Colorectal Cancer
LectinPred
Lectin Prediction
NL-MIND-BEST
Non-Linear MARCH-INSIDE Nested Drug-Bank Exploration Screening Tool
MISSProt-HP
MARCH-INSIDE Spectral moment prediction of Self Proteins in Human Parasites (other than original source organism)
MIND-BEST
Linear MARCH-INSIDE Nested Drug-Bank Exploration & Screen tool
Trypano-PPI
Trypano Protein - Protein Interactions
Plasmod-PPI
Plasmodium Protein-Protein Interactions
EnzClassPred
Enzyme Class Prediction
ATCUNpred
ATCUN DNA-cleavage protein activity Prediction
Last modified: 18-Jul-2014
Copyright © RNASA 2009-2013
Home | Links | About
MARCH INSIDE
S2S Network

Target Prediction: applications for predicting the function of several targets such as proteins in human diseases or molecular proceses by using protein information.

Figure 1. Bio-AIMS Target Prediction with 12 free Web tools for proteins

3. Materials and Methods

The molecular information was encoded into graph/network molecular descriptors [16]: in the case of proteins/peptides, the nodes are the amino acids and the edges are the peptide bonds and graph specific properties [17-19].

The set of molecular descriptors with the specific protein activities or properties have been used as input for the Machine Learning techniques to obtain the best classifier predictors.

The best protein predictors are implemented into 12 free Web tools as Target Prediction section of Bio-AIMS server: http://bio-aims.udc.es/TargetPred.php.

The inputs of these tools are protein PDB name [20,21], SMILE chemical formulas for drugs or peptide sequences. The tools to calculate the descriptors are MARCH-INSIDE (Python version) [22] and S2SNet – Sequence to Star Network [23,24] (programmed in Python/Biopython [25] The Machine Learning methods [26] have been used from STATISTICA [27], Weka [28] and R [29]. The Web tools were programmed in XHTML [30], PHP [31], Python [25], and R [29].

4. Conclusions

This short communication is presenting a collection of free Web tools for protein prediction at Bio-AIMS. These tools are based on protein descriptors obtained with molecular graphs, Machine Learning methods to search for the best classifier and Python/PHP/XHTML/R programming languages.

This collection is an important contribution to the open science and demonstrate the power of encoding of the molecular information into molecular graph descriptors for proteins/peptides.

Acknowledgments

The authors acknowledge the support provided by the Galician Network of Drugs R+D REGID (Xunta de Galicia R2014/025) and by the "Collaborative Project on Medical Informatics (CIMED)" PI13/00280 funded by the Carlos III Health Institute from the Spanish National plan for Scientific and Technical Research and Innovation 2013-2016 and the European Regional Development Fund / GAIN (FEDER - CONECTAPEME - INTERCONECTA). This work was partially supported by the Galician Network for Colorectal Cancer Research (Red Gallega de Cáncer Colorrectal - REGICC, Ref.: 2014/039), Institute for Biomedical Informatics of A Coruña (INIBIC), and Center for Research of Information and Communication Technologies (CITIC).

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Archer, S. Qsar: A critical appraisal. *NIDA Res. Monogr.* **1978**, 86-102.
2. Rehn, D.; Zerling, W. A critical comment on the use of the plate diffusion test for qsar considerations. *Methods Find. Exp. Clin. Pharmacol.* **1983**, *5*, 457-460.
3. Munteanu, C.R.; Gonzalez-Diaz, H.; Garcia, R.; Loza, M.; Pazos, A. Bio-aims collection of chemoinformatics web tools based on molecular graph information and artificial intelligence models. *Comb. Chem. High Throughput Screen.* **2015**, *18*, 735-750.
4. Fernandez-Lozano, C.; Cuinas, R.F.; Seoane, J.A.; Fernandez-Blanco, E.; Dorado, J.; Munteanu, C.R. Classification of signaling proteins based on molecular star graph descriptors using machine learning models. *J. Theor. Biol.* **2015**, *384*, 50-58.
5. Fernandez-Lozano, C.; Gestal, M.; Pedreira-Souto, N.; Postelnicu, L.; Dorado, J.; Munteanu, C.R. Kernel-based feature selection techniques for transport proteins based on star graph topological indices. *Curr. Top. Med. Chem.* **2013**, *13*, 1681-1691.
6. Gonzalez-Diaz, H.; Munteanu, C.R.; Postelnicu, L.; Prado-Prado, F.; Gestal, M.; Pazos, A. Libp-pred: Web server for lipid binding proteins using structural network parameters; pdb mining of human cancer biomarkers and drug targets in parasites and bacteria. *Mol. BioSyst.* **2012**, *8*, 851-862.
7. Munteanu, C.R.; Magalhaes, A.L.; Uriarte, E.; Gonzalez-Diaz, H. Multi-target qpdr classification model for human breast and colon cancer-related proteins using star graph topological indices. *J. Theor. Biol.* **2009**, *257*, 303-311.
8. Munteanu, C.R.; Pedreira-Souto, N.; Dorado, J.; Pazos, A.; Pérez-Montoto, L.G.; Ubeira, F.M.; González-Díaz, H. Lectinpred: Web server that uses complex networks of protein structure for prediction of lectins with potential use as cancer biomarkers or in parasite vaccine design. *Molecular Informatics* **2014**, *33*, 276-285.
9. Gonzalez-Diaz, H.; Prado-Prado, F.; Sobarzo-Sanchez, E.; Haddad, M.; Maurel Chevalley, S.; Valentin, A.; Quetin-Leclercq, J.; Dea-Ayuela, M.A.; Teresa Gomez-Munos, M.; Munteanu, C.R., *et al.* Nl mind-best: A web server for ligands and proteins discovery-theoretic-

experimental study of proteins of giardia lamblia and new compounds active against plasmodium falciparum. *J. Theor. Biol.* **2011**, *276*, 229-249.

10. Gonzalez-Diaz, H.; Muino, L.; Anadon, A.M.; Romaris, F.; Prado-Prado, F.J.; Munteanu, C.R.; Dorado, J.; Sierra, A.P.; Mezo, M.; Gonzalez-Warleta, M.*, et al.* Miss-prot: Web server for self/non-self discrimination of protein residue networks in parasites; theory and experiments in fasciola peptides and anisakis allergens. *Mol. BioSyst.* **2011**, *7*, 1938-1955.
11. Gonzalez-Diaz, H.; Prado-Prado, F.; Garcia-Mera, X.; Alonso, N.; Abeijon, P.; Caamano, O.; Yanez, M.; Munteanu, C.R.; Pazos, A.; Dea-Ayuela, M.A.*, et al.* Mind-best: Web server for drugs and target discovery; design, synthesis, and assay of mao-b inhibitors and theoretical-experimental study of g3pdh protein from trichomonas gallinae. *J. Proteome Res.* **2011**, *10*, 1698-1718.
12. Rodriguez-Soca, Y.; Munteanu, C.R.; Dorado, J.; Pazos, A.; Prado-Prado, F.J.; Gonzalez-Diaz, H. Trypano-ppi: A web server for prediction of unique targets in trypanosome proteome by using electrostatic parameters of protein-protein interactions. *J. Proteome Res.* **2010**, *9*, 1182-1190.
13. Rodriguez-Soca, Y.; Munteanu, C.R.; Dorado, J.; Rabuñal, J.; Pazos, A.; González-Díaz, H. Plasmod-ppi: A web-server predicting complex biopolymer targets in plasmodium with entropy measures of protein-protein interactions. *Polymer* **2010**, *51*, 264-273.
14. Concu, R.; Dea-Ayuela, M.A.; Perez-Montoto, L.G.; Prado-Prado, F.J.; Uriarte, E.; Bolas-Fernandez, F.; Podda, G.; Pazos, A.; Munteanu, C.R.; Ubeira, F.M.*, et al.* 3d entropy and moments prediction of enzyme classes and experimental-theoretic study of peptide fingerprints in leishmania parasites. *Biochim. Biophys. Acta* **2009**, *1794*, 1784-1794.
15. Munteanu, C.R.; Vazquez, J.M.; Dorado, J.; Sierra, A.P.; Sanchez-Gonzalez, A.; Prado-Prado, F.J.; Gonzalez-Diaz, H. Complex network spectral moments for atcun motif DNA cleavage: First predictive study on proteins of human pathogen parasites. *J. Proteome Res.* **2009**, *8*, 5219-5228.
16. Harary, F. *Graph theory*. Westview Press: MA, 1969.
17. Balaban, A.T. Chemical graphs. Xxxiv. Five new topological indices for the branching of tree-like graphs. *Theor. Chim. Acta* **1979**, *53*, 355-375.
18. Balaban, A.T. Topological indices based on topological distances in molecular graphs. *Pure Appl. Chem.* **1983**, *55*, 199-206.
19. Randic, M. On graphical and numerical characterization of proteomics maps. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1330-1338.
20. Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.; Meyer, E.F., Jr.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535-542.
21. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank *Nucleic Acids Res.* **2000**, *28*, 235-242.
22. González-Díaz, H.; Torres-Gomez, L.A.; Guevara, Y.; Almeida, M.S.; Molina, R.; Castanedo, N.; Santana, L.; Uriarte, E. Markovian chemicals "in silico" design (march-inside), a promising

approach for computer-aided molecular design iii: 2.5d indices for the discovery of antibacterials. *J Mol Model* **2005**, *11*, 116-123.

23. Munteanu, C.R.; Gonzáles-Díaz, H. *S2snet - sequence to star network, reg. No. 03 / 2008 / 1338, santiago de compostela, spain*, Santiago de Compostela, Spain, 2008.
24. Munteanu, C.R.; Magalhaes, A.L.; Duardo-Sanchez, A.; Pazos, A.; Gonzalez-Diaz, H. S2snet: A tool for transforming characters and numeric sequences into star network topological indices in chemoinformatics, bioinformatics, biomedical, and social-legal sciences. *Curr. Bioinf.* **2013**, *8*, 429-437.
25. Rossum, G.v. Python reference manual. http://docs.python.org/ref/ref.html
26. Mitchell, T. *Machine learning*. 1997.
27. StatSoft.Inc. *Statistica (data analysis software system), version 6.0, www.Statsoft.Com.Statsoft, inc.*, 6.0; 2002.
28. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.A. The weka data mining software: An update. *SIGKDD Explorations* **2009**, *11*.
29. Team, R.D.C. *R: A language and environment for statistical computing. R foundation for statistical computing, vienna, austria*. R Foundation for Statistical Computing: Vienna, Austria, 2008.
30. Pemberton, S.; Altheim, M.; AskJeeves, A.D.; Boumphrey, F.; Mitre, G.B.; Donoho, A.W.; Dooley, S.; Hofrichter, K.; Hoschka, P.; Ishikawa, M., *et al.* Xhtml™ 1.0: The extensible hypertext markup language. W3c recommendation. http://www.w3.org/TR/2000/REC-xhtml1-20000126/
31. Lerdorf, R. Dynamic web pages with php3. Webtechniques. http://www.php.net

SciForum
Mol2Net

CORAL: The Dispersion of SWNTs in Different Organic Solvents

Alla P. Toropova, Andrey A. Toropov*
IRCCS, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, 20156, Via La Masa 19, Milano, Italy
*To whom correspondence should be addressed: E-mail: andrey.toropov@marionegri.it
Tel: +39 02 3901 4595 Fax: +3902 3901 4735 (AAT)。

Published: 4 December 2015

Abstract: Single-walled carbon nanotubes (SWNTs) are group of new substances with specific cylindrical architecture of their molecules. The dispersion of SWNTs in different organic solvents is parameter that can be valuable information for development of nanomaterials. The CORAL software is a tool to build up model for different endpoints using the Monte Carlo technique. In this work, the ability of the CORAL software to be a tool to predict dispersion of SWCTs in different organic solvents demonstrated.

Keywords: QSPR; Monte Carlo method; SMILES; Validation; Domain of applicability; CORAL software

1. Introduction

The development of nanotechnology indicates that use of carbon nanotubes (CNTs), in general, and single-walled nanotubes (SWNTs), in particular, gives attractive possibilities for chemical technology [1], biochemistry [2], and medicine [3]. The dispersibility of SWNTs in various solvents is important physicochemical characteristics [4] from point of view of technology [5, 6].

The theoretical approaches to predict of the endpoint for different solvents developed and described in the literature [5, 6]. Apparently, however, similar studies based on the quantitative structure – property / activity relationships (QSPRs/QSARs) [7-10] be continued.

In particular, this work dedicated to search for a new alternative approaches to predict the dispersibility of SWNTs in organic solvents using the Monte Carlo method [11, 12].

2. Method

2.1. Data

The dispersibility of SWNTs in a series of 29 different organic solvents taken in the literature [5, 6]. The endpoint is decimal logarithm of dispersibility C_{max} expressed in mg/mL. Three random splits into the visible training set (in fact this is structured into two sets: the training and calibration sets) and the invisible validation set are examined in order to check up the actual ability of the approach.

2.2. Optimal descriptors

The optimal descriptor used in this work calculated as the following:

$$DCW(T^{*}, N^{*}) = \Sigma CW(V_k)$$

(1)

In Eq. 1: The T* is the coefficient to classify vertex degree into two categories rare and not rare. The parameter has influence upon the results of the Monte Carlo optimization

The Vk is vertex in the hydrogen-suppressed molecular graph [13-15]. Table 1 contains example of the hydrogen suppressed graph together with (0, 1) adjacency matrix and Vk values, which are calculated using the elements of the matrix; the $CW(V_k)$ is correlation weight of the V_k. The T* is threshold or a coefficient for the classification of vertices into two classes: (i) rare (the number of V_k in the training set is less than T^*) and (ii) active (the number of V_k in the training set is larger than T^*). The rare vertices are not involved building up model: their correlation weights fixed equal to zero. The N^* is the number of epochs of the Monte Carlo optimization. In fact, one can use arbitrary T and N, but the T* and N^* are values of these parameters which give preferable statistical quality of the model for the calibration set, hoping that the model is avoided of the overtraining (i.e. the situation where the excellent quality for the training set accompanied by poor quality for the calibration set).

[Table 1, around here]

Having the numerical data on the correlation weights, one can calculated the DCW(T*,N*) for all compounds of the training, calibration, and test sets. Using the data on the training set, one should to calculate the model

$$Endpoint = C_0 + C_1 * DCW(T^*,N^*)$$

(2)

The predictive potential of the model calculated with Eq. 2 should be checked with data on the calibration and validation sets.

2.3. Mechanistic interpretation

The CORAL models give the possibility to interpret the role of different molecular features as the promoters of increase or decrease of an endpoint. For instance, if in several runs of the Monte Carlo optimization the correlation weight of the V_k is larger than zero, then this feature is promoter of the endpoint increase, whereas if the correlation weights of the V_k are less than zero in several runs of the optimization then the V_k should be interpreted as promoter of the endpoint decrease.

2.4. Domain of applicability

The domain of applicability for the CORAL model defined according to prevalence of different molecular features in the training and the calibration sets: each molecular feature has the statistical defect. The defect is equal to difference between probabilities of the molecular feature in the training set and in the calibration set.

Ideal situation if the difference is zero, however in praxis, this value is not zero. Apparently, the preferable distribution should be characterized by the minimal sum of these parameters for all active molecular features. Thus, the approach gives possibility not only to define the domain of applicability, but, also, to compare different distributions into the training and calibration sets.

3. Results and Discussion

3.1. Models

The models for dispersibility of SWNTs in different organic solvents for three different random splits into the training, calibration, and validation sets are the following:

Split 1: $\log_{10}C_{max}$ = -2.9944 (± 0.1266) + 0.2076 (± 0.0183) * DCW(3,24) (3)

Split 2: $\log_{10}C_{max}$ = -3.2119 (± 0.1310) + 0.2380 (± 0.0200) * DCW(1,25) (4)

Split 3: $\log_{10}C_{max}$ = -3.1077 (± 0.1187) + 0.2165 (± 0.0175) * DCW(2,25) (5)

Table 2 contains numerical data on the correlation weights used to calculate the DCW(T*,N*) for calculation with Eqs. 3-5. Table 3 contains the statistical characteristics of models calculated with Eqs. 3-5.

[Table 2 and 3, around here]

3.2. Domain of applicability

The estimation of the domain has been done by scheme described in the literature [16]: the solvent with sum of defects for the SMILES less than average value of this parameter (for the training set) multiplied by 2:

$$\sum Defect \leq 2 \times \overline{\sum defect}$$

(6)

[Table 4, around here]

One can see (Table 4) the distribution into the training, calibration, and validation sets has influence upon the domain of applicability, but this situation gives possibility to select preferable from the statistical point of view the distribution (minimum of the above-mentioned defect).

3.3. Mechanistic interpretation

Three runs of the Monte Carlo optimization with selected T* and N* give correlation weights collected in Table 5. One can hypothesizes about the role of molecular features represented by the V_k in the behavior of a solvent: if all runs give positive value of correlation weight for a V_k then the molecular feature can be classified as promoter of an endpoint increase, if all runs gives negative value of correlation weight then the molecular feature represented by the V_k can be classified as promoter of endpoint decrease [16].

3.4. Selection of molecular features for increase (decrease) of dispersibility of SWNT

The analysis of data collected in Table 5 lead to hypothesis that presence (in hydrogen suppressed molecular graph which is representation of a solvent) of carbon and nitrogen atoms with vertex degree 3, oxygen with vertex degree 1, and carbon atom with vertex degree 2 are promoter of dispersibility increase. The presence in molecular graph represented a solvent carbon vertex with vertex degree 1 is promoter of the endpoint decrease.

3.5. Comparison with QSAR models from the literature

The statistical characteristics of model of $\log_{10}C_{max}$ (for validation set, the same 29 solvents) suggested in work [5] are n=6, r^2=0.932; $\overline{r_m^2}$ =0.844, Δr_m^2 = 0.066; the statistical quality of model (for the same 29 solvents) suggested in work [6] are n=7, r^2=0.807; $\overline{r_m^2}$ =0.744, Δr_m^2 = 0.125. The above-mentioned models related to fixed splits into the training and validation sets, whereas models suggested in this work are checked up with three different splits. It is to be noted, different splits into the training and validation sets used in work [5] and in work [6].

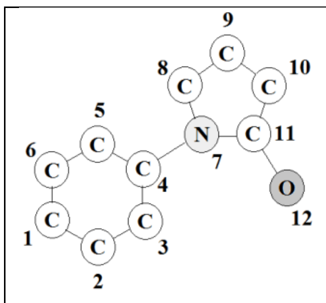
4. Conclusions

The described version of the Monte Carlo method gives satisfactory prediction for the disprsibility of SWNT in different solvents. The distribution into the visible training set (together with calibration set) and the invisible validation set has influence on the predictive potential models. The approach gives quite convenient measure of quality of distribution into the training and the validation sets together with convenient criterion of the domain of applicability.

Acknowledgements

Authors thank the EU project PROSIL (project LIFE12 ENV/ IT/ 000154) for financial support.

Table 1
Example of the hydrogen suppresed graph together with the adjacecncy matrix and vertex degree values (V_k).



	C_1	C_2	C_3	C_4	C_5	C_6	N_7	C_8	C_9	C_{10}	C_{11}	O_{12}	V_k
C_1	0	1	0	0	0	1	0	0	0	0	0	0	2
C_2	1	0	1	0	0	0	0	0	0	0	0	0	2
C_3	0	1	0	1	0	0	0	0	0	0	0	0	2
C_4	0	0	1	0	1	0	1	0	0	0	0	0	3
C_5	0	0	0	1	0	1	0	0	0	0	0	0	2
C_6	1	0	0	0	1	0	0	0	0	0	0	0	2
N_7	0	0	0	1	0	0	0	1	0	0	1	0	3
C_8	0	0	0	0	0	0	1	0	1	0	0	0	2
C_9	0	0	0	0	0	0	0	1	0	1	0	0	2
C_{10}	0	0	0	0	0	0	0	0	1	0	1	0	2
C_{11}	0	0	0	0	0	0	1	0	0	1	0	1	3
O_{12}	0	0	0	0	0	0	0	0	0	0	1	0	1

Table 2

Correlation weights of different vertices (chemical element together with the vertex degree) calculated by the Monte Carlo method for split 1, 2, and 3

V_k	$CW(V_k)$	Prevalence in training set	Prevalence in training set	Defect
Split 1, Eq.3				
C...1...	-0.15309	6	4	0.0071
C...2...	0.09204	13	6	0.0094
C...3...	1.92499	11	6	0.0021
Cl..1...	0.0	1	0	0.0000
N...1...	0.0	2	1	0.0000
N...3...	5.52455	7	3	0.0125
O...1...	0.57923	13	8	0.0034
O...2...	0.0	2	1	0.0000
S...2...	0.0	0	1	0.0000
Split 2, Eq. 4				
C...1...	-0.30200	6	5	0.0179
C...2...	0.12657	14	5	0.0197
C...3...	1.96087	11	6	0.0021
Cl..1...	0.86431	1	0	1.0000
N...1...	1.90949	2	2	0.0268
N...3...	5.40373	6	4	0.0071
O...1...	0.12862	13	7	0.0027
O...2...	0.62649	3	0	1.0000
Split 3, Eq. 5				
C...1...	-0.29507	6	6	0.0268
C...2...	0.11067	14	5	0.0197
C...3...	2.10422	11	8	0.0113
Cl..1...	0.0	0	1	0.0000
N...1...	1.13484	2	0	1.0000
N...3...	5.39626	6	4	0.0071
O...1...	0.58729	13	8	0.0034
O...2...	0.12783	3	0	1.0000

Table 3. The statistical characteristics of models for dispersibility of SWNTs in the organic solvents

	Training set (n=14)			**Calibration set (n=8)**			**Validation set (n=7)**		
Split	r^2	Q^2	F	r^2	$\overline{r_m^2}$	Δr_m^2	r^2	$\overline{r_m^2}$	Δr_m^2
1	0.605	0.420	18	0.885	0.83	0.04	0.900	0.81	0.09
2	0.611	0.436	19	0.888	0.67	0.15	0.953	0.88	0.05
3	0.607	0.440	19	0.931	0.90	0.00	0.912	0.59	0.19

Table 3. The experimental and calculated

ID*	SMILES	$Log_{10}C_{max}$, experiment	$Log_{10}C_{max}$, calculated	∑Defect	Domain of Applicability
	Split 1 $2\times\overline{\sum defect}$ =0.1320				
M02	O=C1N(C)CCCN1C	-0.1870	-0.1872	0.0730	YES
M05	CN1CCCC1=O	-0.9360	-1.3023	0.0533	YES
M09	N1(C(CCC1)=O)C=C	-1.0760	-1.2832	0.0627	YES
M14	O=CN(C)C	-1.6380	-1.7718	0.0396	YES
M16	CCC#N	-1.8240	-2.9880	0.0259	YES
M17	C=CC(=O)O	-1.8600	-2.3670	0.0254	YES
M20	C1CCC(=O)C1	-1.8890	-2.3982	0.0431	YES
	Split 2 $2\times\overline{\sum defect}$ =0.7097				
M02	O=C1N(C)CCCN1C	-0.1870	-0.1957	0.1140	YES
M05	CN1CCCC1=O	-0.9360	-1.4100	0.0890	YES
M09	N1(C(CCC1)=O)C=C	-1.0760	-1.3798	0.1087	YES
M14	O=CN(C)C	-1.6380	-2.0088	0.0653	YES
M17	C=CC(=O)O	-1.8600	-2.7257	0.0451	YES
M18	OCCSCCO	-1.8670	-3.0302	0.0843	YES
M20	C1CCC(=O)C1	-1.8890	-2.5941	0.0837	YES
	Split 3 $2\times\overline{\sum defect}$ =0.8251				
M06	O=C1CCCN1CCC#N	-0.9390	-0.9675	1.1402	No
M09	N1(C(CCC1)=O)C=C	-1.0760	-1.3250	0.1276	YES
M12	O=CN1CCCCC1	-1.4090	-1.6687	0.1290	YES
M14	O=CN(C)C	-1.6380	-1.9162	0.0839	YES
M16	CCC#N	-1.8240	-2.8780	1.0663	No
M17	C=CC(=O)O	-1.8600	-2.4378	0.0646	YES
M18	OCCSCCO	-1.8670	-2.7576	0.0858	YES

[*)] ID taken in Ref. [5]

Table 5.Correlation weights of different kinds of the vertex degrees obtained in three runs of the Monte Carlo calculations.

V_k	Run 1	Run 2	Run 3	Effect	Prevalence in Training set	Prevalence in Calibration set	Defect
Split 1							
C...2...	0.08953	0.09126	0.08627	increase	13	6	0.0094
O...1...	0.48932	0.53427	0.55085	increase	13	8	0.0034
C...3...	1.84634	1.79047	1.83472	increase	11	6	0.0021
N...3...	5.40255	5.42669	5.35392	increase	7	3	0.0125
C...1...	-0.20080	-0.20290	-0.19975	decrease	6	4	0.0071
N...1...	0.0	0.0	0.0	N/A*	2	1	0.0000
O...2...	0.0	0.0	0.0	N/A	2	1	0.0000
Cl..1...	0.0	0.0	0.0	N/A	1	0	0.0000
S...2...	0.0	0.0	0.0	N/A	0	1	0.0000
Split 2							
C...2...	0.12572	0.12594	0.14700	increase	14	5	0.0197
O...1...	0.16092	0.05264	0.73299	increase	13	7	0.0027
C...3...	1.90844	1.96168	2.37844	increase	11	6	0.0021
N...3...	5.27071	5.42916	6.00261	increase	6	4	0.0071
O...2...	0.61590	0.72085	0.42493	increase	3	0	1.0000
N...1...	1.77083	1.97330	2.00166	increase	2	2	0.0268
Cl..1...	0.92213	0.91521	0.78688	increase	1	0	1.0000
C...1...	-0.30019	-0.30069	-0.00482	decrease	6	5	0.0179
Split 2							
C...2...	0.11570	0.10125	0.12910	increase	14	5	0.0197
O...1...	1.21165	0.65059	1.10164	increase	13	8	0.0034
C...3...	2.62358	2.03992	2.52190	increase	11	8	0.0113
N...3...	5.99585	5.34519	5.99873	increase	6	4	0.0071
N...1...	1.24679	1.06117	1.38648	increase	2	0	1.0000
C...1...	-0.00109	-0.30296	0.00496	N/A	6	6	0.0268
O...2...	-0.12445	0.17736	-0.06459	N/A	3	0	1.0000
Cl..1...	0.0	0.0	0.0	N/A	0	1	0.0000

[*)] N/A = classification is not available

Author Contributions

A.P.T. had prepared the group of the random splits of available organic solvents into the training, calibration, and validation sets; had taken part in the carrying out the Monte Carlo experiments; and the discussion of the final text of the manuscript. A.A.T. had prepared the preliminary strategy of selection of group of versions for the Monte Carlo method and had prepared the preliminary version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Backes, C.; Hauke, F.; Schmidt, C.D.; Hirsch, A. Fractioning HiPco and CoMoCAT SWCNTs via density gradient ultracentrifugation by the aid of a novel perylene bisimide derivative surfactant *Chem. Commun.* **2009**, *19*, 2643–2645.
2. Choi, W.; Ohtani, S.; Oyaizu, K.; Nishide, H.; Geckeler, K.E. Radical polymer-wrapped SWNTs at a molecular level: High-rate redox mediation through a percolation network for a transparent charge-storage material. *Adv. Mater.* **2011**, *23*, 4440-4443.
3. Cheung, W.; Pontoriero, F.; Taratula, O.; Chen, A.M.; He, H. DNA and carbon nanotubes as medicine *Adv. Drug Deliv. Rev.* **2010**, *62*, 633-649.
4. Bergin, S.D.; Sun, Z.; Streich, P.; Hamilton, J.; Coleman, J.N. New solvents for nanotubes: Approaching the dispersibility of surfactants. *J. Phys. Chem.* C **2010**, *114*, 231-237.
5. Rofouei, M.K.; Salahinejad, M.; Ghasemi, J.B. An alignment independent 3D-QSAR modeling of dispersibility of single-walled carbon nanotubes in different organic solvents. *Fuller. Nanotube. Car. N.* **2014**, *22*, 605-617.
6. Yilmaz, H.; Rasulev, B.; Leszczynski, J. Modeling the dispersibility of single walled carbon nanotubes in organic solvents by quantitative structure-activity relationship approach. *Nanomaterials* **2015**, *5*, 778-791.
7. Gonzáles-Díaz, H.; Gia, O.; Uriarte, E.; Hernádez, I.; Ramos, R.; Chaviano, M.; Seijo, S.; Castillo, J.A.; Morales, L.; Santana, L.; Akpaloo, D.; Molina, E.; Cruz, M.; Torres, L.A.; Cabrera, M.A. Markovian chemicals "in silico" design (MARCH-INSIDE), a promising approach for computer-aided molecular design I: Discovery of anticancer compounds. *J. Mol. Model.* **2003**, *9*, 395-407.
8. Speck-Planche, A.; Scotti, M.T.; de Paulo-Emerenciano, V. Current pharmaceutical design of antituberculosis drugs: Future perspectives. *Curr. Pharm. Design* **2010**, *16*, 2656-2665.
9. Toropov, A.A.; Toropova, A.P.; Veselinović, A.M.; Veselinović, J.B.; Nesmerak, K.; Raska, I.; Duchowicz, P.R.; Castro, E.A.; Kudyshkin, V.O.; Leszczynska, D.; Leszczynski, J. The Monte Carlo method based on eclectic data as an efficient tool for predictions of endpoints for nanomaterials-two examples of application. *Comb. Chem. High. T. Scr.* **2015**, *18*, 376-386.
10. Worachartcheewan, A.; Prachayasittikul, V.; Toropova, A.P.; Toropov, A.A.; Nantasenamat, C. Large-scale structure-activity relationship study of hepatitis C virus NS5B polymerase inhibition using SMILES-based descriptors. *Mol. Divers.* **2015**, *19*, 955-964.
11. Toropova, A.P.; Toropov, A.A. CORAL software: Prediction of carcinogenicity of drugs by means of the Monte Carlo method. *Eur. J. Pharm. Sci.* **2014**, *52*, 21-25.

12. Toropova, M.A.; Toropov, A.A.; Raška, I.; Rašková, M. Searching therapeutic agents for treatment of Alzheimer disease using the Monte Carlo method. *Comput. Biol. Med*. **2015**, *64*, 148-154.
13. Toropov, A.A.; Toropova, A.P.; Gutman, I. Comparison of QSPR models based on hydrogen-filled graphs and on graphs of atomic orbitals. *Croat. Chem. Acta* **2005**, *78*, 503-509.
14. Gutman, I.; Toropov, A.A.; Toropova, A.P. The graph of atomic orbitals and its basic properties. 1. Wiener index. *MATCH Commun. Math. Comput. Chem*. **2005**, *53*, 215-224.
15. Gutman, I.; Furtula, B.; Toropov, A.A.; Toropova, A.P. The graph of atomic orbitals and its basic properties. 2. Zagreb indices. *MATCH Commun. Math. Comput. Chem*. **2005**, *53*, 225-230.
16. Toropov, A.A.; Toropova, A.P. Quasi-QSAR for mutagenic potential of multi-walled carbon-nanotubes. *Chemosphere* **2015**, *124*, 40-46.

SciForum
Mol2Net

Phylogenetic and Genetic Analysis of Envelope Gene of the Prevalent Dengue Serotypes in India in Recent Years

Sumanta Das*, Centre for Interdisciplinary Research and Education, Jodhpur

Park, Kolkata 700068, India, Email: sumantadey13@gmail.com; anandy43@yahoo.com

Ashesh Nandy, Centre for Interdisciplinary Research and Education, Jodhpur

Published: 4 December 2015

Abstract: A fresh wave of dengue infection, particularly dengue serotypes 1 and 3, have been observed all across India in recent times and has led to several fatalities. Since the surface situated envelope protein of the dengue virion is responsible for virus entry into the host cell, we have laid special emphasis on its characterization and analyses of the envelope gene with an aim to eventually develop inhibitors of the dengue virus. There are four serotypes of the dengue virus of which types 1 and 3 are the most widely prevalent in India. 2D graphical representations of the envelope gene from various countries show that the gene from an Indian dengue type 1 virus bears a strong resemblance to the genes from Asia, whereas in the case of dengue type 3, the Indian strain representation shows simlarity to strains from North America. Phylogenetic trees using alignment procedures also bear this out, implying an inherent cross-national spread of the dengue virus. Moreover, hydropathy analysis shows that amino acid compositional changes are tending to increase hydrophobic residues in the dengue type 3 viruses leading to morphological changes that may explain, in part, the higher pathogenicity of the dengue virus in India in recent times. These exercises serve to show the urgency of comprehensive genetic surveillance of the dengue virus to anticipate further damaging changes in the viral sequence.

Keywords: Dengue envelope gene, 2D graphical representation, Phylogenetic analysis, transition/transversion ratio, hydropathy plot, amino acid composition changes, viral pathogenicity.

Introduction

The dengue virus (DENV) is estimated to infect over 50-100 million people across 60 countries annually [1], with fatalities of around 50 to 100 thousand per year [2, 3]. This virus, genus *Flavivirus*, family Falviviridae, consists of four antigenically distinct serotypes (DENV 1 to 4) [4, 5]. The genome comprises a single strand of RNA encoding three structural proteins - the capsid (C), premembrane/membrane (PrM/M) and envelope (E), and seven non-structural (NS) proteins - NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5 [6], arranged sequentially as shown in a 2D graphical representation in Fig.1. Of these, the surface situated envelope glycoprotein is usually targeted by vaccines and drugs for inhibition [1, 7]. Moreover it is found that dengue virus infectivity depends on envelope protein binding to target cell [8, 9, 10]; it is also interesting to note that even a single nucleotide change in the envelope protein gene of dengue DENV2 can affect its neurovirulence in mice [11], and could be important in the case of human patients also.

Because of the wide prevalence of dengue infections in India, we have done a bioinformatics study of the dengue focusing on the outer capsid envelope gene which is responsible for viral endocytosis. We report here results of the study for a selection of dengue viruses of DENV3 and DENV1 types from various countries of the world along with the same sera prevalent in India over a period of several years. In India these two serotypes are mostly observed [12], but in recent times cases of DENV3 appear to be much more prevalent. In this report we characterize and draw phylogeny of envelope gene sequences of DENV3 and DENV1 from India and other countries for comparative study along with 2D graphical representation of the gene for visualization of the base distributions and structure. We also report here the transition/transversion ratio, amino acid composition as well as hydropathy index of the envelope protein to get an indication of the protein morphological differences between the serotypes DENV3 and DENV1, with emphasis on the changes in DENV3 envelope protein in relation to DENV1.

Result and Discussions

Figs. 2 and 3 show the phylogenetic relationship between the envelope gene sequences of DENV3 and DENV1 respectively from various countries in recent times. The results show that DENV3 envelope gene from India is closely related to American strains, whereas for DENV1, the Indian gene is more closely related to Chinese and other Asian strains, which we had also reported in our previous paper [14].

Figs. 4 and 5 show the comparative analysis of the envelope gene by 2D graphical representation. These representations clearly show that the envelope gene of DENV 3 strains from India (e.g., Locus ID: JQ686083) are closely related to American strains (e.g., EU596494) whereas in case of DENV 1 strains from India (e.g., JF967939) are related to Asian strains (e.g., China KC006933).

Table 1 shows the transition/transversion ratio matrix of the dengue envelope genes of DENV3 and DENV1 from India, influenza A surface genes (Hamaglutinin and Neuraminidase) and mammalian beta globin genes. Numbers in bold are the transition frequencies (%), others are transversion frequencies (%). The data above show that while the rate of transition to transversion mutations is about 55: 45 in mammalian and other viral genes, in the case of dengue genes this ratio at 88:12 is significantly different.

Figs. 6 and 7 show the hydropathy index plot of the envelope gene sequences of both the serotypes 3 and 1, respectively, from India after or from the year 2000. The results indicate that hydrophobicity of the envelope protein of the Indian serotype DENV3 shows a slight tendency to increase with time but in case of the Indian serotype DENV1 it shows tendency towards decreasing with time implying morphological changes in the protein structure.

Fig 8 shows the differences in the amino acid composition of DENV1 and DENV3 for each amino acid of Indian strains in the period under study: positive values imply higher frequencies in DENV3. The figure shows changes vary mainly in amino acids like Asparagine, Isoleucine, Tyrosine, Aspartic acids, Arginine, Serine, Threonine, Valine and Glutamic acids while changes are at a minimum in amino acids like Cysteine, Lysine, Methionine, Glutamine and Tryptophan between the two serotypes, DENV3 and DENV1.

We infer these data have the following implications:

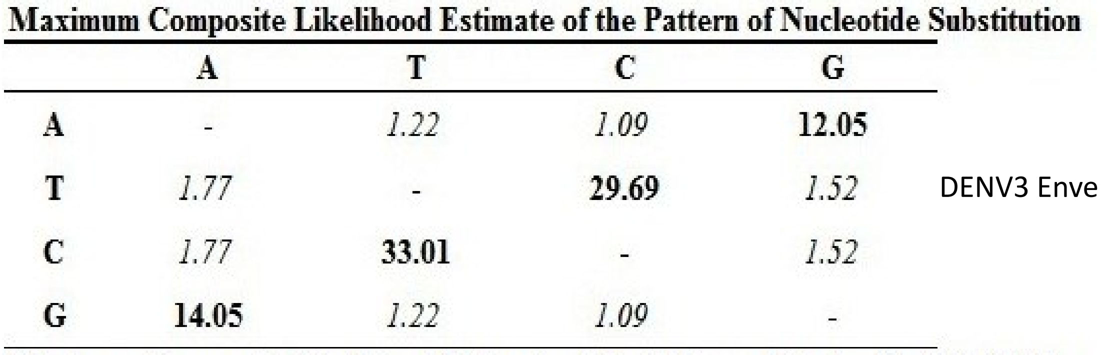
> Since, Isoleucine and Valine have large aliphatic hydrophobic side chains, their molecules are rigid and their mutual hydrophobic interactions are important for correct folding of proteins. So changes in the composition of these amino acids, e.g. higher Isoleucine for DENV3, can affect the 3D structure of the envelope protein for both the strains.

> Tyrosine contains a large rigid aromatic group on the side chain and is also one of the biggest amino acids. Moreover like Isoleucine and Valine, Tyrosines are hydrophobic and trend to orient towards the interior of the folded protein molecule. Excess Tyrosine in DENV3 could be making it more hydrophobic.

> Arginine contains a large flexible side chain with a positively-charged end. The flexibility of the chain makes Arginine suitable for binding to molecules with many negative charges on their surfaces. The strong charge makes the amino acid prone to be located on the outer hydrophilic surfaces of the proteins. Since the envelope protein is surfaced exposed, change in the Arginine composition of DENV3 might reduce the binding property of the protein.

> Serine and Threonine with a hydroxyl group are very hydrophilic. Their reduced frequency in DENV3 leads to higher hydrophobicity and consequent morphological change in the DENV3 envelope protein

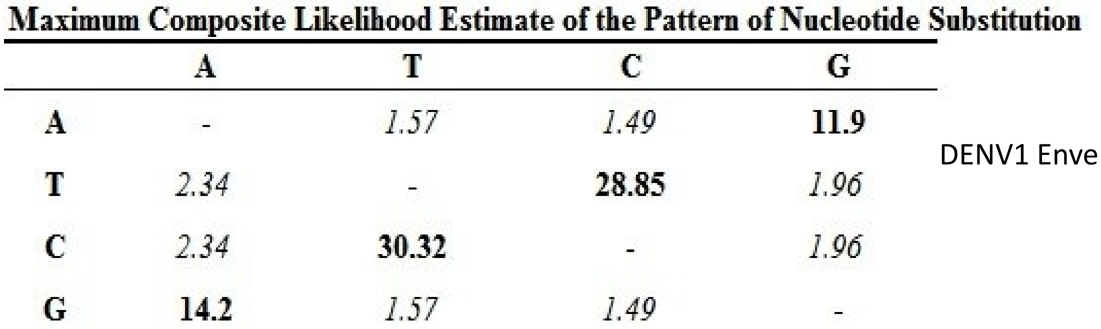
Table 1 showing transition/transversion matrix



Maximum Composite Likelihood Estimate of the Pattern of Nucleotide Substitution

	A	T	C	G
A	-	*1.22*	*1.09*	**12.05**
T	*1.77*	-	**29.69**	*1.52*
C	*1.77*	**33.01**	-	*1.52*
G	**14.05**	*1.22*	*1.09*	-

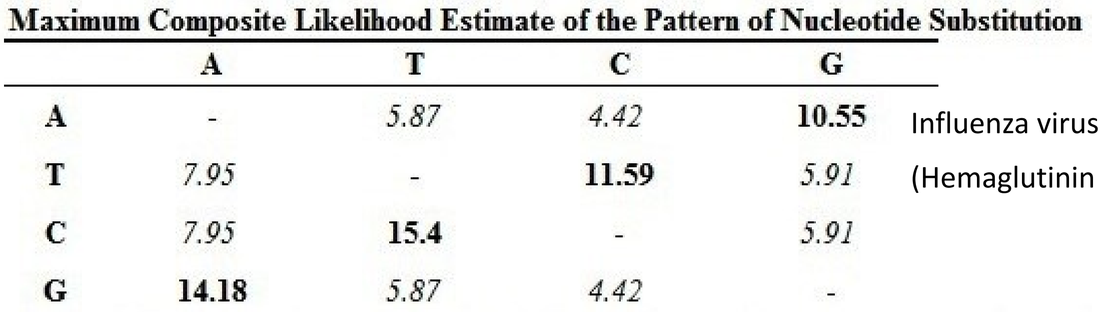
DENV3 Envelope genes



Maximum Composite Likelihood Estimate of the Pattern of Nucleotide Substitution

	A	T	C	G
A	-	*1.57*	*1.49*	**11.9**
T	*2.34*	-	**28.85**	*1.96*
C	*2.34*	**30.32**	-	*1.96*
G	**14.2**	*1.57*	*1.49*	-

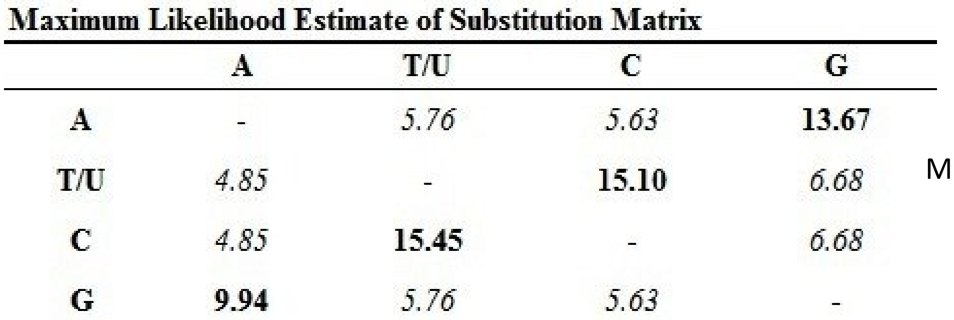
DENV1 Envelope genes



Maximum Composite Likelihood Estimate of the Pattern of Nucleotide Substitution

	A	T	C	G
A	-	*5.87*	*4.42*	**10.55**
T	*7.95*	-	**11.59**	*5.91*
C	*7.95*	**15.4**	-	*5.91*
G	**14.18**	*5.87*	*4.42*	-

Influenza virus A surface genes
(Hemaglutinin and Neuraminidase)



Maximum Likelihood Estimate of Substitution Matrix

	A	T/U	C	G
A	-	*5.76*	*5.63*	**13.67**
T/U	*4.85*	-	**15.10**	*6.68*
C	*4.85*	**15.45**	-	*6.68*
G	**9.94**	*5.76*	*5.63*	-

Mammalian Beta globin genes

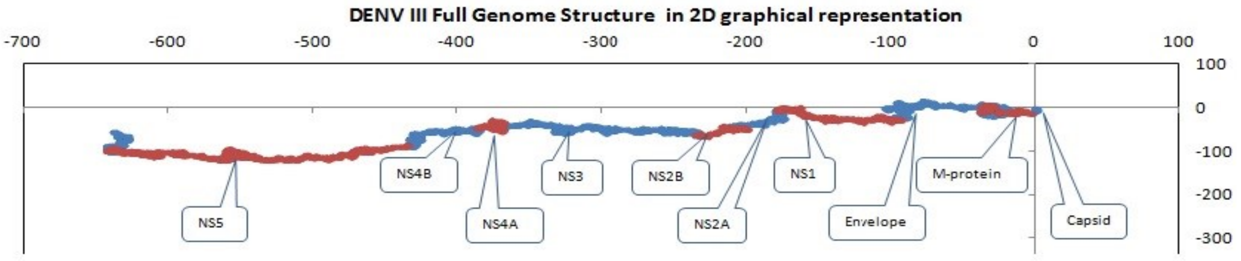


Figure 1: Genome structure of Indian DENV3 in 2D graphical representation

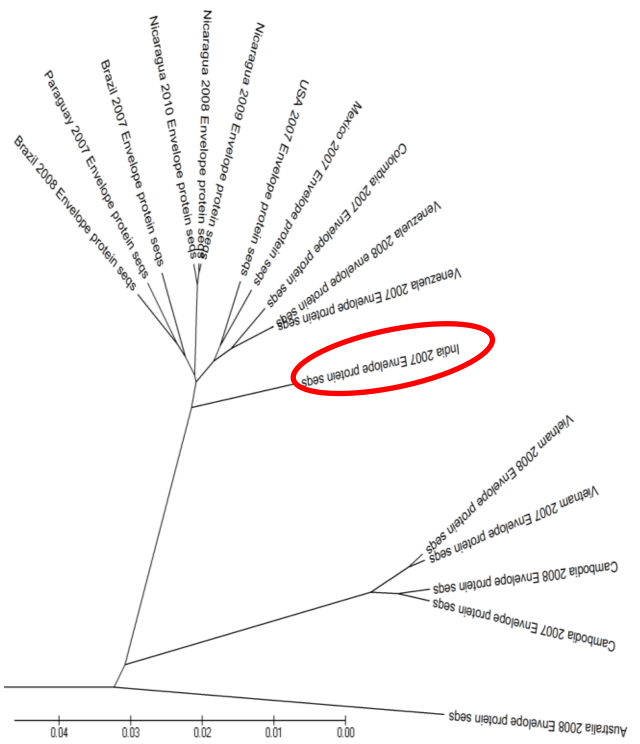

Brazil 2008 Envelope protein seqs
Paraguay 2007 Envelope protein seqs
Brazil 2007 Envelope protein seqs
Nicaragua 2010 Envelope protein seqs
Nicaragua 2008 Envelope protein seqs
Nicaragua 2009 Envelope protein seqs
USA 2007 Envelope protein seqs
Mexico 2007 Envelope protein seqs
Colombia 2007 Envelope protein seqs
Venezuela 2008 envelope protein seqs
Venezuela 2007 Envelope protein seqs
India 2007 Envelope protein seqs
Vietnam 2008 Envelope protein seqs
Vietnam 2007 Envelope protein seqs
Cambodia 2008 Envelope protein seqs
Cambodia 2007 Envelope protein seqs
Australia 2008 Envelope protein seqs
0.04
0.03
0.02
0.01
0.00

Figure 2: Phylogenetic relationship between 18 envelope gene sequences of DENV 3 from various

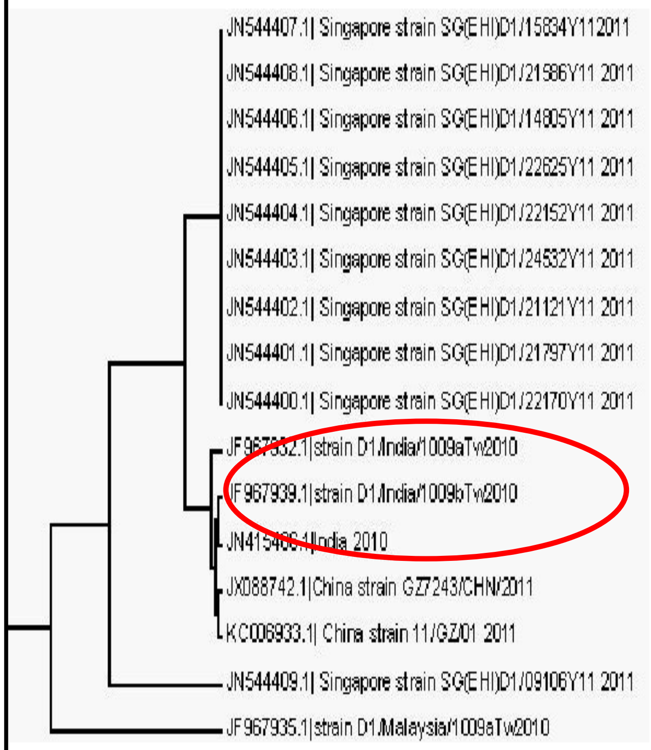

JN544407.1| Singapore strain SG(EHI)D1/15834Y112011
JN544408.1| Singapore strain SG(EHI)D1/21586Y11 2011
JN544406.1| Singapore strain SG(EHI)D1/14805Y11 2011
JN544405.1| Singapore strain SG(EHI)D1/22625Y11 2011
JN544404.1| Singapore strain SG(EHI)D1/22152Y11 2011
JN544403.1| Singapore strain SG(EHI)D1/24532Y11 2011
JN544402.1| Singapore strain SG(EHI)D1/21121Y11 2011
JN544401.1| Singapore strain SG(EHI)D1/21797Y11 2011
JN544400.1| Singapore strain SG(EHI)D1/22170Y11 2011
JF967932.1|strain D1/India/1009aTw2010
JF967939.1|strain D1/India/1009bTw2010
JN415406.1|India 2010
JX088742.1|China strain GZ7243/CHN/2011
KC006933.1| China strain 11/GZ/01 2011
JN544409.1| Singapore strain SG(EHI)D1/09106Y11 2011
JF967935.1|strain D1/Malaysia/1009aTw2010

Figure 3: Phylogenetic relationship between envelope gene sequences of DENV 1 from Asian countries in recent

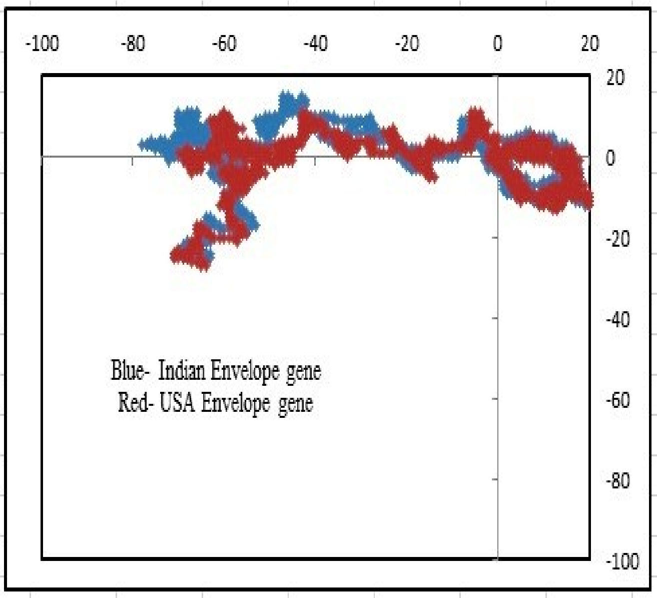


Figure 4:2D graphical representations of DENV3 envelope gene sequences from India (Locus ID: JQ686083) and USA (EU596494)

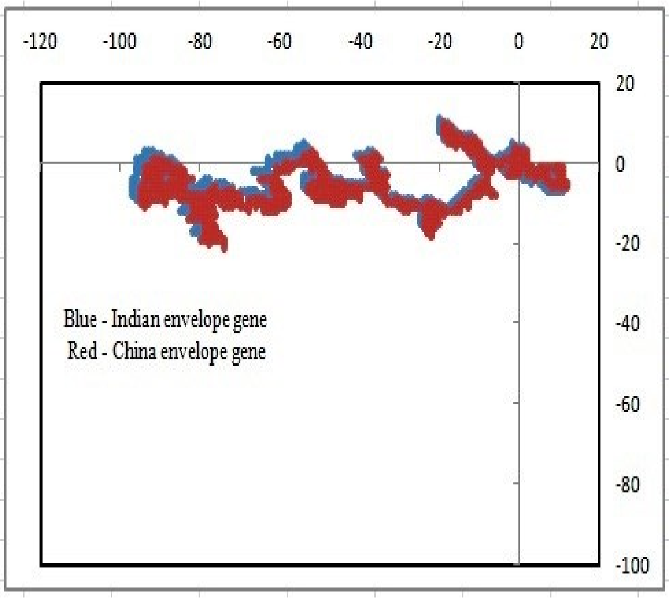


Figure 5:2D graphical representations of DENV1 envelope gene sequences from India (JF967939) and China (KC006933).

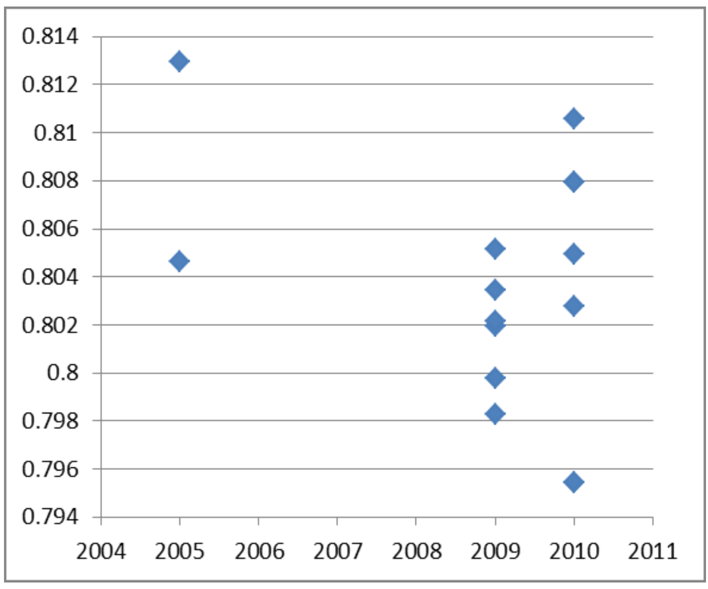


Figure 6: Hydropathy plot of envelope gene sequences of DENV 3 strains from India

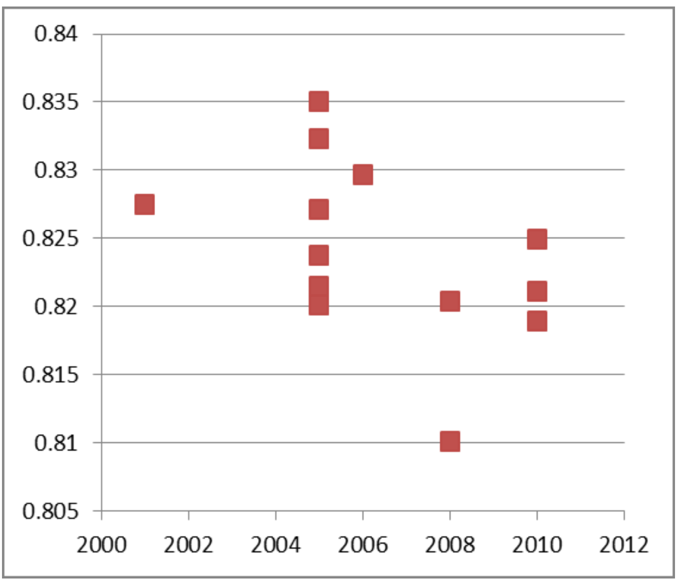


Figure 7: Hydropathy plot of envelope gene sequences of DENV 1 strains from India

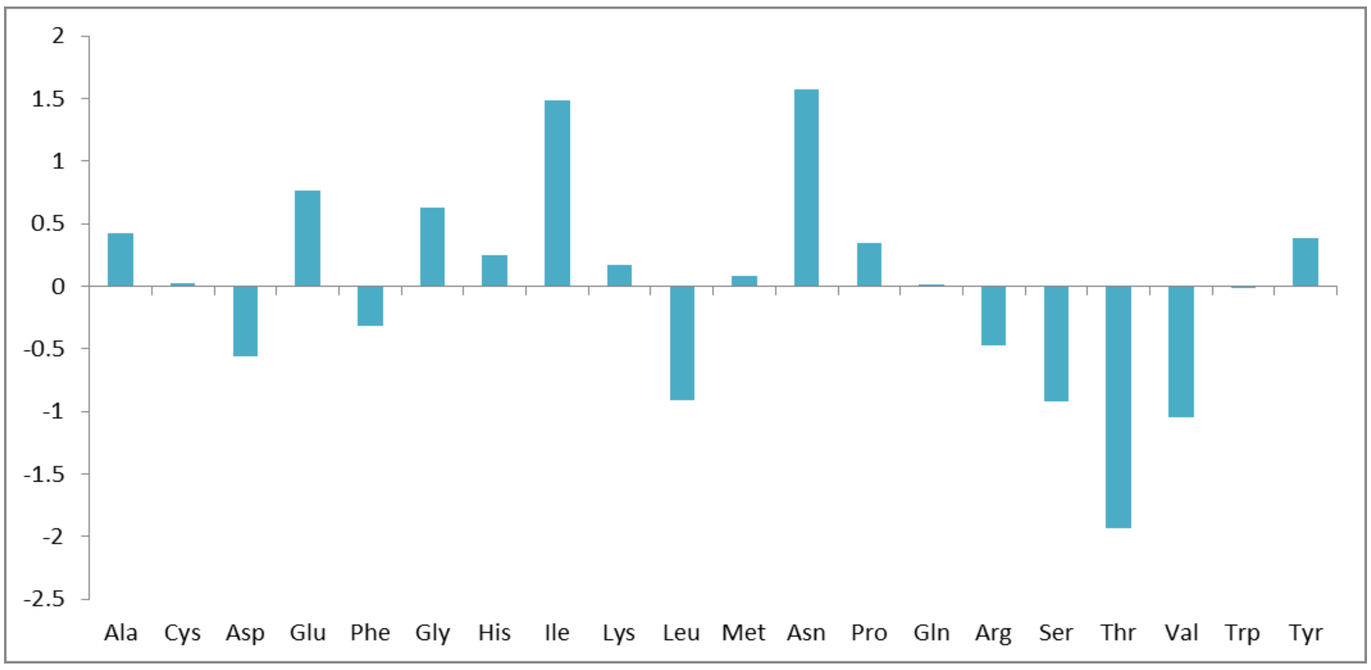


Figure 8: Chart of the differences in amino acid composition of DENV1 and DENV3 for each amino acid. Positive values imply higher frequencies in DENV3.

Materials and Methods

Sequence data of the envelope genes from across the world including India of the dengue virus serotype 3 and 1 (DENV3 and DENV1) were downloaded from the NCBI website (last accessed Sep15, 2015); the number of dengue gene and genomic sequences from India were comparatively less than from other countries like the USA, China, Australia, etc., but we downloaded as many complete gene sequences as were available. Plots of the gene sequences were made in a 2D graphical representation scheme with ACGT axis system for visualization of the sequence structure and analyzed for numerical characterisation. For the graphical representation the sequences are plotted as a walk in a 2D grid taking one step in the negative x-direction for an adenine, in the positive y-direction for a cytosine, the positive x-direction for a guanine and in the negative y-direction for a thymine/uracil. This yields (x,y) values for each base in a sequence and successive plots of the points generate a curve in the 2D graph [13]. We also used an alignment based procedure, the software MEGA 5.2, to draw phylogenetic trees showing the evolutionary relationship between the sequences. The transition/transversion ratio along with the amino acid composition of the envelope gene sequences were determined using the same software. The hydropathy index was determined using ExPasy server.

Conclusion

From phylogenetic as well as through 2D graphical representation point of view it is evident that Indian DENV3 strains are closely related to American strains, whereas Indian DENV1 strains are similar to Asian strains.

From the genetic point of view, we hypothesize from the hydropathy index and amino acid differences that morphological changes are occurring in the envelope gene structure in recent times. Such changes could be leading to enhanced viral pathogenecity and might explain part of the high incidence of dengue cases being observed now.

The lack of adequate representations of sequences of the dengue genes and genomes from India makes it difficult to define any trends with good statistics. However, the analyses we have done so far does indicate the propensity of morphological changes in the dengue envelope gene which could lead to higher pathogenicity. These exercises serve to show the urgency of comprehensive genetic surveillance of the dengue virus to anticipate further damaging changes in the viral sequence.

Acknowledgments
The authors would like to acknowledge with thanks the opportunity provided by the organizers of the MOL2NET-1 conference to present our findings to a general audience.

Author Contributions
S Dey worked on the problem and wrote the paper, which A Nandy critically reviewed and edited.

Conflicts of Interest
The authors declare no conflict of interest.

References

1. Whitehead S S et al, Prospects for a dengue virus vaccine, Nature Rev Microbiol 5 (2007) 518-528.
2. Guzman M G et al, Do escape mutants explain rapid increases in dengue case – fatality rates within epidemics?, Lancet 355 (2000) 1902-1903.
3. Fatima Z et al, Serotype and genotype analysis of dengue virus by sequencing followed by phylogenetic analysis using samples from three mini-outbreaks – 2007-2009 in Pakistan, BMC Microbiology (2011) 11:200.
4. Mukhopadhyay S, Kuhn RJ, Rossmann MG (2005) A structural perspective of the Flavivirus life cycle. Nat Rev Microbiol 3: 13-22
5. Heinz FX, Stiasny K (2012) Flaviviruses and falvivirus vaccines. Vaccine 30: 4301-4306.
6. .Rice, C.M., Lenches, E.M., Eddy, S.R., Shin, S.J., Sheets, R.L., Strauss, J.H., Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. Science 229 (1985), 726– 733.
7. Gubler DJ (1998) Dengue and Dengue Hemorrhagic Fever, Clin Microbiol Rev 11(3): 480-496.
8. Chen Y, Maguire T, Hileman RE, Fromm JR., Esko JD, Linhardt RJ, Marks RM (1997) Dengue virus infectivity depends on envelope protein binding to target cell heparin sulfate. Nat Med 3: 866-871.

9. Leitmeyer KC, Vaughin DV, Watts DM, Salas R, Chacon IVD, Ramos C, Rico-Hesse R (1999) Dengue Virus Structural Differences That Correlate with Pathogenesis. J Viral 73: 4738-4747.
10. Chen Y, Maguire T, Marks RM (1996) Demonstration of binding of dengue virus envelope protein to target cells. J Virol 70: 12, 8765-8772.
11. Sanehez IJ, Ruiz BH (1996) A single nucleotide change in the E protein gene of dengue virus 2 Mexican strain affects neurovirolence in mice. J Gen Virol 77: 2541-2545.
12. Anoop M, Mathew A J. Jayakumar B, Issac A, Nair S, Abraham R, Anupriya MG, Sreekumar E (2012) Complete genome sequencing and evolutionary analysis of dengue virus serotype 1 isolates from an outbreak in Kerala, South India. Virus genes 45: 1-13.
13. Nandy, A. A new graphical representation and analysis of DNA sequence structure: Methodology and Application to Globin Genes, Current Sci. 66(4), 309-314 (1994).
14. Dey, S.; Nandy, A.; Nandy, P.; Das, S. Diversity and evolution of the envelope gene of dengue virus type 1 circulating in India in recent times. Int. J Bioinfor Res and Appl. (in press)

SciForum
Mol2Net

Phosphorylated Sites on the Disordered Interface Signatures the Interacting Behavior of Proteins—A Comparative Mapping of Phosphorylation Propensities on Disordered Interfaces of Interactome and Negatome

Srinivas Bandaru[1], Deepika Ponnala[1], Chandana Lakkaraju[1], Chaitanya Kumar Bhukya[1], Uzma Shaheen[1] and Anuraj Nayarisseri[2,*]

[1.] Institute of Genetics and Hospital for Genetic Diseases, Osmania University, Hyderabad – 500 016, India.

[2.] In silico Research Laboratory, Eminent Biosciences, Indore – 452 010, India.

* Author to whom correspondence should be addressed; E-Mail:anuraj@eminentbio.com; Tel: +91 9752295342

Published: 4 December 2015

Abstract: Hub proteins in interaction networks involved in signaling pathways are known to have more disordered residues than non-hubs. Since the signaling mechanisms involving PPI are regulated by phosphorylation, disordered interfaces could be thought to be extremely phosphorylated. In the present study we sought to map the phosphorylated sites onto disordered regions on interacting proteins-Interactomes and non-interacting proteins-Negatomes. Dataset of non-interacting protein included 784 proteins retrieved from Negatome database 2.0. 2252 interacting proteins were retrieved from "GeneMania". Intrinsically disordered regions were predicted with "Disopred" program. The binding interfaces were defined by "PDBePISA" server, while, phosphorylation sites were derived from "NetPhos" program. All phosphorylation sites were mapped onto protein structures using alignments calculated by the MUSCLE program. As anticipated, the extent of phosphorylation in interactomes were significantly higher in disordered regions to its ordered counter parts (p=0.04). Insights revealed that the disordered regions in negatome were sparse in comparison to those in interactomes (p<0.0024). Declined phosphorylated sites were observed in negatomes. The widespread non-flexible and ordered regions in the negatomes confer the non interacting nature of the protein in turn makes it poor participant in signal transduction that involves phosphorylation. Our study sheds light on the importance of phosphorylated sites on disordered regions as a mark to decide whether protein would possibly interact or not.

1. Introduction

Protein–Protein interaction (PPI) forms the core of interactomics system and unsurprisingly, aberrant PPIs are the basis of diseases, such as Alzheimer's and cancers [1]. Growing body of evidence suggest that hub proteins in interaction networks in signaling pathways have more disordered residues than non-hubs [2, 3]. Intrinsically disordered proteins lack single well-defined structure and are characterized by specific amino acid composition, a propensity for post-translational modifications and the ability to bind to many different partners [4].The importance of disorder in protein–protein interactions is apparent from analysis of protein-protein interaction networks. Studies have shown that hub proteins in interaction networks have more disordered residues than non-hubs and that there may be a weak correlation between the disorder of a protein and the number of its partners [5 - 8]. The functional diversity of disordered proteins and their multi-binding properties, allow them to play a unique role in signaling networks [9]. Phosphorylations form the most important dynamic covalent modification involved in the signal transduction systems [10]. Therefore disordered interfaces which are involved in PPI could be thought to be extremely phosphorylated. Since the signaling mechanisms involving PPI are regulated by phosphorylation, disordered interfaces could be thought to be extremely phosphorylated. In the present study we sought to map the phosphorylation propensities of Ser,Tyr and Thr onto disordered regions on interacting proteins-"Interactomes" and non-interacting proteins-"Negatomes".

. METHODOLOGY

Dataset Compilation

For interactomes, we compiled a data set consisting of 2252 human protein complexes of known 3D structures from Protein Data bank. The interacting proteins were selected based on their interaction hubs; as provided from Genemania. 784 non-interacting proteins were selected from Negatome database, which provides the list of non-interacting proteins available in PDB. Further, redundant proteins were removed using BLAST, using a p-value threshold of 10e-07.

Identification of Phosphorylation Sites

Phosphorylation sites were derived from PhosphoSitePlus, Phospho.ELM, and PHOSIDA 26 web servers. All the phosphorylation sites were mapped onto protein structures in the PDB using alignments calculated by the MUSCLE program.

Prediction of Intrinsically Disordered Regions

Disordered regions were predicted using the support vector machine enabled Disopred programs. Amino acid propensities for Disorder was calculated using TopIDP program.

Statistical Analysis

The propensity of disorderness and phosphorylation in ordered and disordered regions was calculated by statistical functions like ANOVA, stepwise logistic regression analysis, Odds ratios and linear regression analysis using SPSS v17.0 suite

RESULTS AND DISCUSSION

Table 1 shows the distribution of disordered and ordered regions in Negatomes and

Interactomes. From the statistical analysis it is quite evident that the disordered regions in interactomes are 1.3 folds higher in comparison to negatomes (p value=0.033) (Figure 1). Given that disordered regions involves in PPI, the analysis reflects the non interacting nature of negatomes as evaluated through disordered prediction.

Table 1. Distribution of disordered and ordered regions in interactomes and negatomes.

	Ds %	Or %	χ2	OR (95% CI)	p Value	Ratio	Diff
Interactomes	62.3	37.7	3.367	1.67 (1.2, 2.9)	0.033	1.3	13.90%
Negatomes	48.4	51.6					
t-Test (95% CI)	F stats	df	p value				
Interactomes *vs.*Negatomes	5.6758	24917	<0.021				

Ds = Disordered regions , Or= Ordered Regions, OR= Odds Ratio, Diff= percentage difference

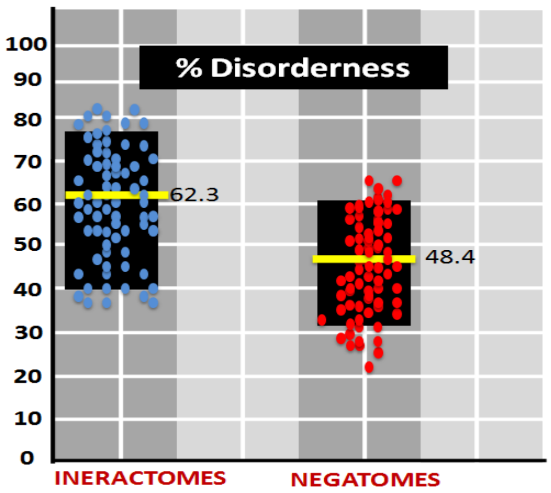


Figure 1. The plot shows the percentage distribution of disorderness in interactomes and negatomes. Each dot represents the disordered percentage of an individual protein falling in interactomes (blue dots) or negatomes (red dot). The bar represents the mean disorderness

We further analyzed the total phosphorylation propensities of Ser, thr and Tyr residues onto Interactomes and negatomes. We observed that percentage distribution of phosphorylation was 1.36 folds higher in the disordered region in both interactomes and negatomes combined (p value =0.0041) Table 2 and Figure 2.

Table 2. The distribution phosphorylation sites in ordered and disordered regions in interactomes and negatomes. The disordered regions have higher share of phosphorylated sites than ordered counterparts.

	Ds (I+N)%	Or (I+N)%	χ2	OR		p Value	Ratio	Diff
P	67	36						
NP	133	134	6.947	1.872 3.0)	(1.13,	0.0041	1.36	15.24 %
t-Test (95% CI)	F stats	df	p value					
P *vs.*NP	8.6758	2251,22	0.04					

P=Phophorylation percentage; NP= Non- Phophorylation percentage

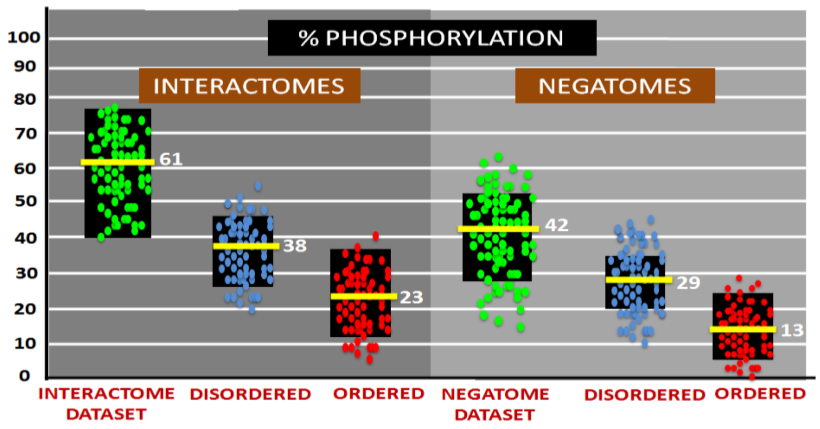


Figure 2.The plot shows the phosphorylation intensities in disordered and ordered regions in interactomes and negatomes. The phosphorylation dots being prominently intense in interactomes.

In the further perusal, we mapped for phosphorylation propensities on to ordered and disordered regions individually in interactomes and negatomes. From stepwise logistic regression analysis we found that disordered regions where 1.4 folds phosphorylated than ordered counter parts in interactomes while in case of negatomes phosphorylated regions were 24.1 folds higher disordered regions than ordered regions testifying the disordered regions are more prone to phosphorylation

TABLE 3. Table shows phosphorylated site with individualistic approach in interactomes and negatomes, In either cases (IN or N) disordered region has higher proportion of phosphorylation.

	Ds %	**Or %**	**χ2**	**OR**	**p Value**	**Ratio**	**Diff**
P (IN)	38	23	5.307				
NP (IN)	62	77		2.044 (1.1, 3.8)	0.01	1.4	19%
P(N)	29	13	7.715				
NP(N)	71	87		2.720 (1.3,5.7)	0.002	1.15	24.11%

P (IN) = Phosphorylated regions in Interactomes;
NP (IN) = Phosphorylated regions in Negatomes;
Ds = Disordered regions; Or= Ordered Regions; OR= Odds Ratio; Diff= percentage difference

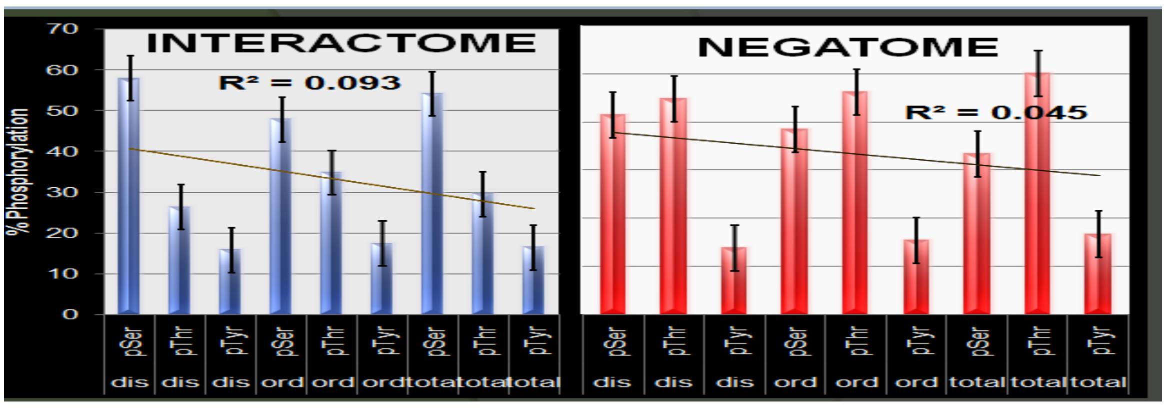


Figure 3. The above graph shows residue wise phosphorylation in interactome (blue bars) and in negatome (Red bars). In The interactome degree of phosphorylation follows the descending order of Ser>Thr>Tyr while no such specific order is present in Negatome.

Further, we performed linear regression analysis in order to confirm whether phopshoryation intensity actually depends on disordered state of the protein for which results were quite convincing (Figure 4). We found significant effect of disorderness on phosphorylation propensity of the protein (R^2= 0.79). The fitness of statistical results therefore confirms that positive correlation effects of diordeness to phosphorylation intensity of the protein.

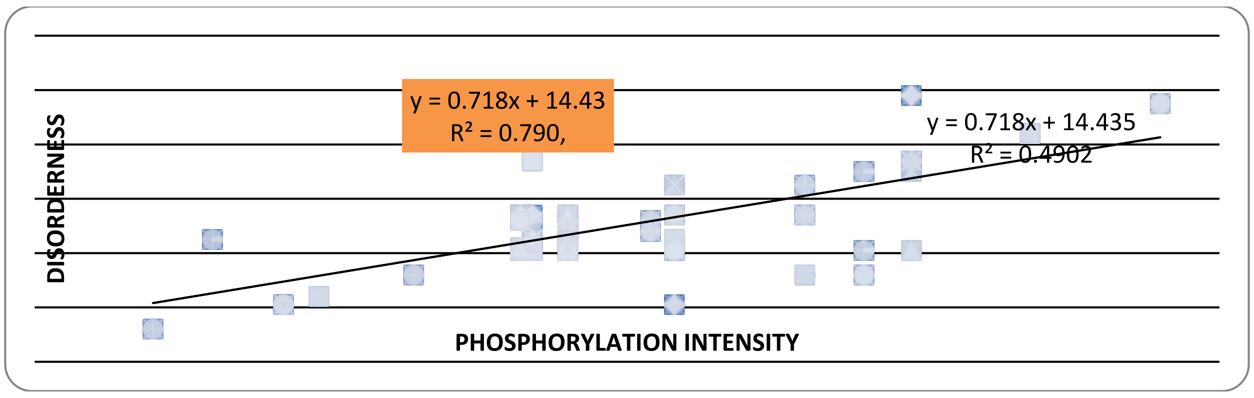


Figure 4. The graph shows the positive correlation (Pearson Corr. Coefficient 0.740) between disorderness and phosphorylation intensity. The graph plotted has the combined data from Interactomes and Negatomes.

. .

4. Conclusions

In conclusion we report that the disordered regions are prominently higher in the interactomes while abruptly low in negatomes. We further observed that the phosphorylated sites which were prominent in disordered regions were significantly higher in interactome than negatome and this observation perhaps explains the interacting nature of proteins. The widespread non-flexible and ordered regions in the negatomes confer the non interacting nature of the protein in turn makes it poor participant in signal transduction that usually involves phosphorylation. Our study sheds light on the importance of phosphorylated sites on disordered regions as a mark to decide whether protein would possibly interact or not.

References and Notes

1. Lu, K. P. (2004). Pinning down cell signaling, cancer and Alzheimer's disease.*Trends in biochemical sciences*, *29*(4), 200-209.
2. Liu, J., Tan, H., & Rost, B. (2002). Loopy proteins appear conserved in evolution. *Journal of molecular biology*, *322*(1), 53-64.
3. Dunker, A. K., Cortese, M. S., & Romero, P. I. LM and Uversky, VN (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks.*FEBS J*, *272*, 5129-5148.
4. Tompa, P. (2005). The interplay between structure and function in intrinsically unstructured proteins. *FEBS letters*, *579*(15), 3346-3354.
5. Dosztanyi, Z., Chen, J., Dunker, A. K., Simon, I., & Tompa, P. (2006). Disorder and sequence repeats in hub proteins and their implications for network evolution. *Journal of proteome research*, *5*(11), 2985-2995.
6. Bellay, J., Han, S., Michaut, M., Kim, T., Costanzo, M., Andrews, B. J., ... & Kim, P. M. (2011). Bringing order to protein disorder through comparative genomics and genetic interactions. *Genome biology*, *12*(2), R14.

7. Nishi, H., Fong, J. H., Chang, C., Teichmann, S. A., & Panchenko, A. R. (2013). Regulation of protein–protein binding by coupling between phosphorylation and intrinsic disorder: analysis of human protein complexes. *Molecular bioSystems*,*9*(7), 1620-1626.
8. Bertolazzi, P., Bock, M. E., & Guerra, C. (2013). On the functional and structural characterization of hubs in protein–protein interaction networks. *Biotechnology advances*, *31*(2), 274-286.
9. Nishi, H., Fong, J. H., Chang, C., Teichmann, S. A., & Panchenko, A. R. (2013). Regulation of protein–protein binding by coupling between phosphorylation and intrinsic disorder: analysis of human protein complexes. *Molecular bioSystems*,*9*(7), 1620-1626.
10. Bray, D. (1995). Protein molecules as computational elements in living cells.*Nature*, *376*(6538), 307-312.

SciForum
Mol2Net

Genetic Algorithms with Fine Tuning

Marcos Gestal[1,*], Julián Dorado[2]

[1] RNASA-IMEDIR Group, Faculty of Computer Science, University of A Coruna, 15071 A Coruña, Spain; Emails: mgestal@udc.es, julian@udc.es

* Correspondent author: Marcos Gestal, Information and Communication Technologies Department, Faculty of Computer Science, University of A Coruna, 15071 A Coruña, Spain; E-Mail: mgestal@udc.es; Tel.: +34-981-167-000; Fax: +34-981-167-160.

Published: 4 December 2015

Abstract: Genetic algorithms are search and optimization techniques which have their origin and inspiration in the world of biology. They provide very good results in different kind of problems, but they are not free of complications. One of the most common problems that may arise with these techniques is that, despite a few generations obtain an approximation to the solution of the problem, they need considerably more to adjust to the final solution. To solve this problem Nature gives us, another time, a valid option. Fine Tuning techniques can model this transmission of knowledge between generations making slight variations in offspring before inserting it into the next generation. For its implementation, a new individual is generated from a solution (non best), changing slightly their genes. It can be performed by means a new Genetic Algorithm, with a lower number of individuals and its own configuration. On this way, solutions avoid local minima and introduce more variability in the global population that increase the possibilities to achieve the best solution. The developed solution uses this approach within a generic tool that makes possible that the user provides their own fitness functions to add any kind of problems. The software will allow to parametrize the execution and will show several graphics to control the evolution. Furthermore, to minimize the time for obtaining solutions the assessment of individuals is made under a distributed scheme. The control of the implementation of Genetic Algorithm will be made from a master computer, which delegated to other slave devices for evaluation and, if necessary, apply fine tuning.

Keywords: genetic algorithms, fine adjustment, fitness evaluation, distributed evaluation

1. Introduction

Genetic Algorithms (GA) [1,2] are one of the techniques included within the field of Evolutionary Computation. Since they were introduced in the work of Holland, they provided excellent results in very different fields of applications [3-10]. The main advantage of this technique is related with the fact that the user only need to know a way to evaluate the goodness of a solution to rank it instead of the need to know a formal/deterministic way to solve it.

But one of the most common issues of this kind of algorithms is related with the local minima: GAs are so good to make an approximation to the optimal solutions, but need a huge number of generations to reach the best solution.

2. Results and Discussion

This paper presents a proposal to prevent the AG can run around a local minimum trying to improve the fitness of genetics individual in each generation

There are different methods to locally improve the fit of an individual [11-14]: hill-climbing, neural networks, fitness sharing... But most of these methods present a similar problem: their dependence with the problem that the GA tries to solve. Our approach is based on perform fine adjustment by a recursive use of GA, so dependency problem is solved. Roughly Fine Tuning technique using GA consists on apply a new (brief) GA evolution to each individual in the population so it attempts to model the transmission of knowledge from one to another generation by means of the acquisition of information of each individual (see Figure 1).

The application of Fine Adjust techniques tries to conquer that during the implementation of GA no stagnation occurs in areas of local minima. Furthermore, these small variations in the offspring attempts to prevent the GA explore with greater profusion to the desired pass these areas and exploring new regions of the search space (increase heterogeneity).

Of course this approach introduces an overhead due the bigger number of evaluations required. To solve this question, a master-slave schema is proposed, as Figure 2 represents.

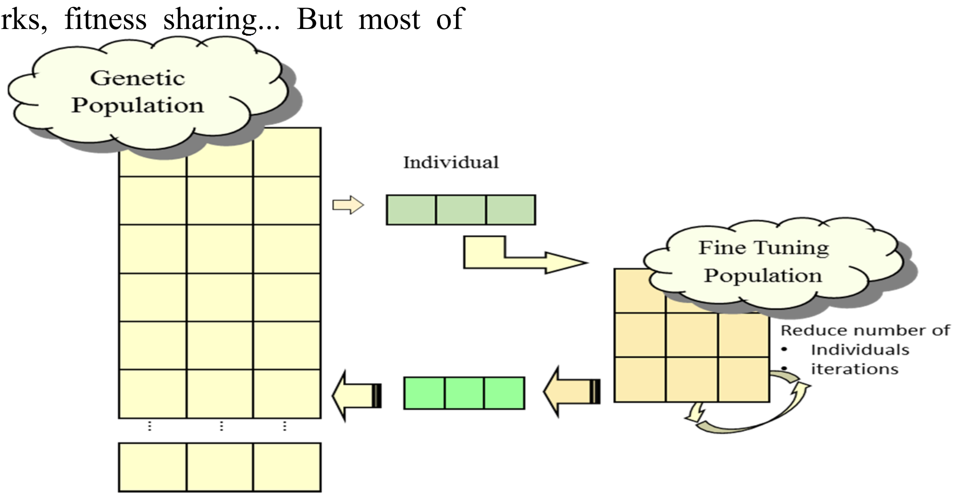


Figure 1. General schema for Fine Tuning

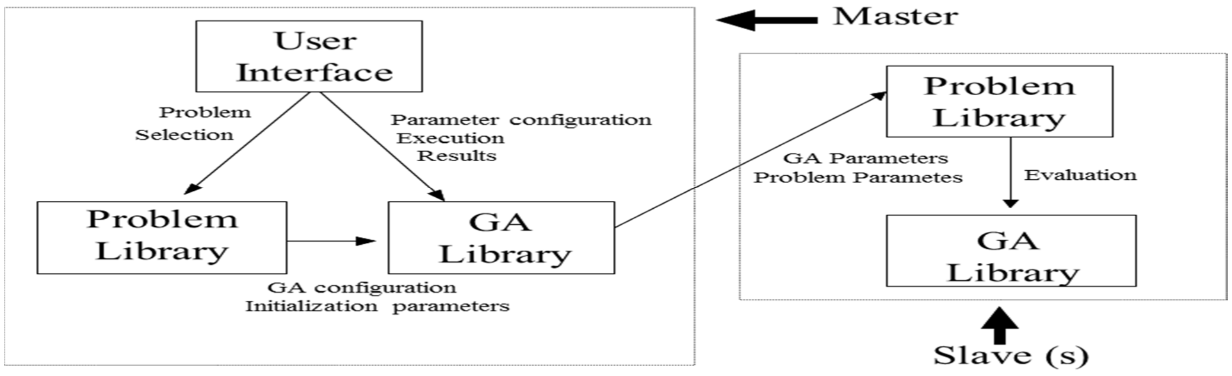
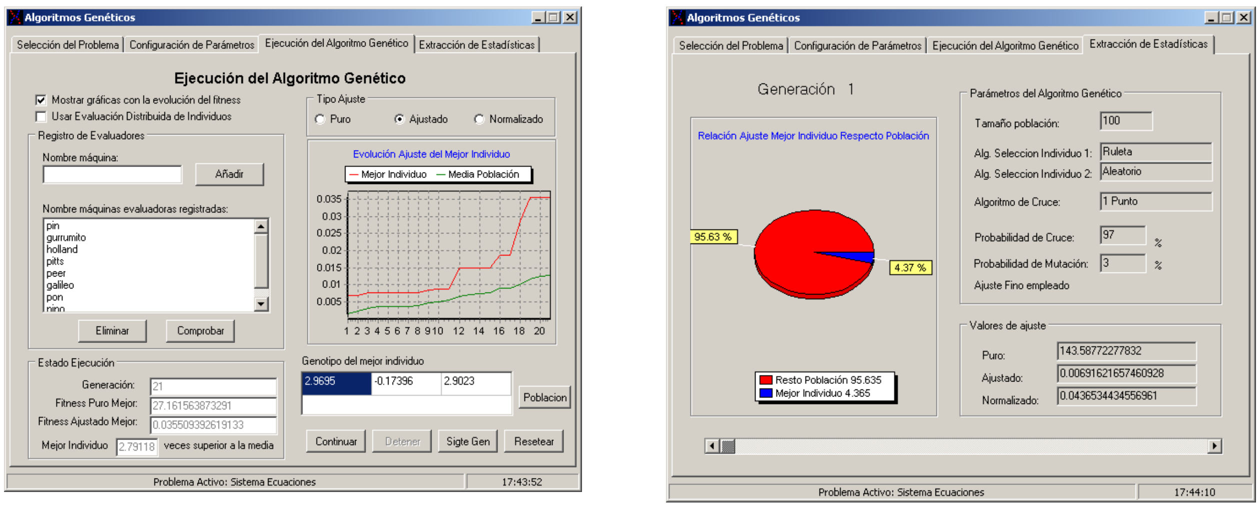

Master
User Interface
Problem Selection
Parameter configuration Execution Results
Problem Library
GA Library
GA configuration Initialization parameters
Problem Library
GA Parameters Problem Parametes
Evaluation
GA Library
Slave (s)

Figure 2. Communication schema between subsystems under distributed evaluations

3. Materials and Methods

The approach explained in this short paper is part of a Phd. Thesis [11] where it was used as support for several methods developed to address the feature selection process under multimodal search spaces [12-15]. The tool was implemented within a library in VisualC++ [16] to leverage its high performance and the user interface was developed in Delphi [17] to produce a good user experience when setting the method, check the evolution, configure the distribution or reviewing the results (see Figure 3)



Ejecución del Algoritmo Genético

☑ Mostrar gráficas con la evolución del fitness
□ Usar Evaluación Distribuida de Individuos

Tipo Ajuste: ○ Puro ● Ajustado ○ Normalizado

Registro de Evaluadores

Nombre máquina:

Nombre máquinas evaluadoras registradas:
pin
gurrumito
holland
pitts
peer
galileo
pon
nino

Estado Ejecución

Generación: 21
Fitness Puro Mejor: 27.161563873291
Fitness Ajustado Mejor: 0.035509392619133
Mejor Individuo 2.79118 veces superior a la media

Genotipo del mejor individuo

2.9695	-0.17396	2.9023

Problema Activo: Sistema Ecuaciones 17:43:52

Algoritmos Genéticos
Selección del Problema
Configuración de Parámetros
Ejecución del Algoritmo Genético
Extracción de Estadísticas
Relación Ajuste Mejor Individuo Respecto Población
95.63 %
4.37 %
Resto Población 95.635
Mejor Individuo 4.365

Generación 1

Parámetros del Algoritmo Genético

Tamaño población: 100
Alg. Seleccion Individuo 1: Ruleta
Alg. Seleccion Individuo 2: Aleatorio
Algoritmo de Cruce: 1 Punto
Probabilidad de Cruce: 97 %
Probabilidad de Mutación: 3 %
Ajuste Fino empleado

Valores de ajuste

Puro: 143.58772277832
Ajustado: 0.00691621657460928
Normalizado: 0.0436534434556961

Problema Activo: Sistema Ecuaciones 17:44:10

Figure 3. Screenshots to check the execution (left) and results (right) phases

One of the advantages of the system is the possibility to add new problems in an easy way.

If a user want to optimize this problem using GAs (and optionally using the Fine Tuning approach) he only needs to provide a library problem (a DLL file) that describes the way how the fitness is evaluated in several functions to allow the communication with the user interface. These functions are the following:

- void FAR PASCAL Inicializar()
  This function is called from the user interface to initialize the problem, so it should read the datasets, initialize the genetic individuals

(number of genes, limits for each one of them…)

- void FAR PASCAL Finalizar()
  Releases memory and other locked resources.

- void FAR PASCAL GetNumeroVariables()
  It is used to pass the number of genes required graphical interface

- __declspec(dllexport) float Evaluar(float* genotype, int nGenes)
  The most important function. The genetic library will call it, so it is necessary to use this nomenclature to allow that remote call. It will return the fitness of the individual passed as argument

- __declspec(dllexport)void SetParametros(char **lArgs, float *argsNum)
  This function, called from the distributed evaluator, will indicate to the slave equipments the problem parameters.

4. Conclusions

This short communication presents a simple modification over the Genetic Algorithm standard functioning to reduce the number of evaluations required to achieve the final solution by means of a fine tuning approach. This new approach is coded within a tool that also allows to distribute the computation requirements along different machines in an asynchronous way to reduce the computational time needed. This tool include a library with the evolutionary algorithm, where the user can add their own functions with a minimum effort and an elemental programming skills. Finally, the tool offers a graphical interface that provides an easy way to parametrize the problem, check the GA evolution and review the results of the execution.

Acknowledgments

The authors acknowledge the support provided by the Galician Network of Drugs R+D REGID (Xunta de Galicia R2014/025) and by the "Collaborative Project on Medical Informatics (CIMED)" PI13/00280 funded by the Carlos III Health Institute from the Spanish National plan for Scientific and Technical Research and Innovation 2013-2016 and the European Regional Development Fund / GAIN (FEDER - CONECTAPEME - INTERCONECTA). This work was partially supported by the Galician Network for Colorectal Cancer Research (Red Gallega de Cáncer Colorrectal - REGICC, Ref.: CN 2012/217), Institute for Biomedical Informatics of A Coruña (INIBIC), and Center for Research of Information and Communication Technologies (CITIC).

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Goldberg, D. *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley Longman Publishing Co.: Boston, MA, 1989.
2. Holland, J. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. University of Michigan Press: 1975.
3. Fernandez-Lozano, C.; Canto, C.; Gestal, M.; Andrade-Garda, J.M.; Rabunal, J.R.; Dorado, J.; Pazos, A. Hybrid model based on genetic algorithms and svm applied to variable selection within fruit juice classification. *Scientific World Journal* **2013**.
4. Gomez-Carracedo, M.P.; Gestal, M.; Dorado, J.; Andrade, J.M. Chemically driven variable selection by focused multimodal genetic algorithms in mid-ir spectra. *Anal. Bioanal. Chem.* **2007**, *389*, 2331-2342.
5. Picado, H.; Gestal, M.; Lau, N.; Reis, L.P.; Tome, A.M. Automatic generation of biped walk behavior using genetic algorithms. In *Bio-inspired systems: Computational and ambient intelligence, pt 1*, Cabestany, J.; Prieto, A.; Sandoval, F.; Corchado, J.M., Eds. 2009; Vol. 5517, pp 805-812.
6. Li, S.; Mehra, R.; Smith, R.; Beard, R. In *Multi-spacecraft trajectory optimization and control using genetic algorithm techniques*, Aerospace Conference Proceedings, 2000 IEEE, 2000, 2000; pp 99-108 vol.107.
7. de Croon, G.C.H.E.; O'Connor, L.M.; Nicol, C.; Izzo, D. Evolutionary robotics approach to odor source localization. *Neurocomputing* **2013**, *121*, 481-497.
8. Gomez-Carracedo, M.P.; Gestal, M.; Dorado, J.; Andrade, J.M. Linking chemical knowledge and genetic algorithms using two populations and focused multimodal search. *Chemom. Intell. Lab. Syst.* **2007**, *87*, 173-184.
9. Ramadan, Z.; Jacobs, D.; Grigorov, M.; Kochhar, S. Metabolic profiling using principal component analysis, discriminant partial least squares, and genetic algorithms. *Talanta* **2006**, *68*, 1683-1691.
10. Cantú-Paz, E. In *Feature subset selection, class separability, and genetic algorithms*, Genetic and Evolutionary Computation Conference, Seattle, Washington, USA, 2004; Springer-Verlag: Seattle, Washington, USA, pp 959–970.
11. Gestal, M. Computación evolutiva para el proceso de selección de variables en espacios de búsqueda multimodales. Universidade da Coruña, A Coruña, 2009.
12. Gestal, M.; Dorado, J. Genetic algorithms in multimodal search space. In *Encyclopedia of information science and technology*, Information Science Reference: Hershey, USA, 2009.
13. Gestal, M.; Gómez-Carracedo, M. Finding multiple solutions with ga in multimodal problems. In *Encyclopedia of artificial intelligence*, Information Science Reference: Hershey, USA, 2009; pp 647-653.
14. Gestal, M.; Vázquez-Naya, J.M.; Ezquerra, N. Genetic algorithms and multimodal search. *Advancing Artificial Intelligence Through Biological Process Applications* **2008**, 231.
15. Rabunal, J.R.; Dorado, J.; Gestal, M.; Pedreira, N. Diversity and multimodal search with a hybrid two-population ga: An application to ann development. In *Computational intelligence and*

bioinspired systems, proceedings, Cabestany, J.; Prieto, A.; Sandoval, F., Eds. 2005; Vol. 3512, pp 382-390.

16. Kruglinski, D.J.; Wingo, S.; Sheperd, G.W. *Programming microsoft visual c++*. Microsoft Press: 1998; p 1153.
17. Pacheco, X.; Teixeira, S. *Delphi 5 developer's guide*. Sams: 2000.

SciForum
Mol2Net

Docking Studies and ADMET Profile of Streblusol E, Anti-Hepatitis B viral Agent of Streblus Asper

Rajeev K Singla [1]*, Rohit Gundamaraju [2], Baishakhi Dey [3], Varadaraj Bhat G [4]

[1] Division of Biotechnology, Netaji Subhas Institute of Technology, Azad Hind Fauz Mar, Sector-3, Dwarka, New Delhi-78, India; Email: rajeevsingla26@gmail.com
[2] Department of Medical Microbiology, Faculty of Medicine, University of Malaya, Kuala Lumpur 50603, Malaysia; Email: rohit.gundamaraju@gmail.com
[3] School of Medical Science & Technology, IIT Kharagpur-721302, India; Email: baishakhidey123@gmail.com
[4] Department of Pharmaceutical Chemistry, Manipal College of Pharmaceutical Sciences, Manipal University, Manipal-576104, India; Email: varad.g@manipal.edu
* Author to whom correspondence should be addressed; E-Mail: rajeevsingla26@gmail.com ; Tel.: +91-98186 03719.

Published: 4 December 2015

Abstract: Background: Streblusol E, a phenolic phytoconstituents of Streblus asper is a potential antihepatitis B viral agent. Objective: Current study is to mechanistically analyze the probable site of action for streblusol E. Material and methods: Streblusol E has been docked with EF3-CaM adenylyl cyclase(1PK0), deoxycytidine kinase(2NOA), human nucleoside diphosphate kinase(3FKB), human Hepatitis B Viral Capsid(1QGT) and hepatitis B X-interacting protein(3MSH) proteins using GRIP docking methodology. Results: Results revealed its preferential intractability towards 1PK0 i.e. EF3-CaM adenylyl cyclase and 1QGT i.e. human hepatitis B viral capsid(HBCAG) compared with reference ligand like adefovir diphosphate(active metabolite of adefovir), lamivudine, tenofovir monophosphate(active metabolite of tenofovir) and tenofovir diphosphate(active metabolite of tenofovir). Drug metabolism and pharmacokinetics studies did affirm that Streblusol E possessed all the desired drug Likeness potential. According to Derek Nexus predictions, Streblusol E did not carry potential toxicities like carcinogenicity, mutagenicity genotoxicity and developmental toxicity, providing further impetus for discovery and clinical development of semi-synthetic analogs of Streblusol E. Conclusion: The present study successfully denotes the docking studies and ADMET profile of streblusol E from Streblus asper.

Keywords: Streblus asper; Streblusol E; Antihepatitis B Viral Activity; GRIP Docking; Viral Receptors.

1. Introduction

Traditional medicinal plants have been recognized for their therapeutic benefits for centuries. However, there is still lack of the evidence for the clarification of their typical mechanism of action (Sripanidkulchai et al., 2009). Streblus asper, family Moraceae, commonly known as Sihora, is a rigid shrub or medium sized tree distributed in South China and South Asia (Aeri et al., 2012; Jun et al., 2012). S. asper was used by traditional healers as remedy for hepatitis B virus in South China.

Streblusol E (Fig. 1), a phenolic phytoconstituent of S. asper was proved to have potential anti-hepatitis B viral activity (anti-HBV activity), but its mechanism is still unknown for the world. Anti-HBV activity was evaluated in HepG2.2.15 cell line stably transfected with the HBV genome. Test concentrations were 4, 20, 100 and 200 µM. The levels of HBV surface antigen (HBsAg) and HBV e antigen (HBeAg) were assayed with ELISA.IC50 value reported was 153.7 µM for HBsAg while for HBeAg, it is 23.1 µM (Jun et al., 2012). In purview of this, we tried to mechanistically analyze the probable site of anti-HBV action and to find out the pharmacophores present.

So aim of current study was to dock streblusol E with specific protein responsible for the virulence of HBV and analyze the site of action and also to evaluate its drug likeness by calculating drug metabolism, pharmacokinetics and toxicity by in silico route.

2. Results and Discussion

Literature revealed the potential of Streblusol E, a phytoconstituent of Streblus asper as anti-hepatitis B viral agent. In purview of this, we tried to mechanistically analyze the probable site of anti-HBV action using various proteins like EF3-CaM adenylyl cyclase(1PK0), deoxycytidine kinase (2NOA), human nucleoside diphosphate kinase (3FKB), human Hepatitis B viral capsid (1QGT), Hepatitis B X-interacting protein(3MSH) and compared it with standard antiviral agents like adefovir diphosphate(active metabolite of adefovir), lamivudine, tenofovir monophosphate (active metabolite of tenofovir) & tenofovir diphosphate (active metabolite of tenofovir)(Table 1). Docking studies revealed that streblusol E of streblus asper have significant anti-hepatitis B viral activity and preferentially interacting with EF3-CaM adenylyl cyclase and human hepatitis B viral capsid protein. In case of EF3-CaM adenylyl cyclase (1PK0), Streblusol E have vanderwaal's interactions with Lys346A, Leu348A, Val350A, His351A, Lys353A, Asp493A, Gly547A, Thr548A, His577A, Thr579A, Glu580A and Asn583A amino acid residues while having hydrophobic interactions with Leu348A and Gly547A amino acid residues of 1PK0. Along with this, streblusol E also have aromatic interactions (Fig. 2) with His351A & His577A and hydrogen bonding with Lys346A. Similarly, reference ligand, Adefovir diphosphate have vanderwaal's interactions with Arg329A, Lys346A, Leu348A, Lys353A, Ser354A, Lys372A, Ala490A, Asp493A, Gly547A, Thr548A, His577A, Gly578A, Thr579A, Glu580A and Asn583A amino acid residues while having hydrophobic interactions with Leu348A, Gly547A, Thr548A, His577A, Gly578A and Asn583A amino acid residues of 1PK0. Along with these, Adefovir diphosphate do exert charge effect with Asp493A, aromatic interactions with His577A and hydrogen bonding with Arg329A, Lys346A, Ser354A, Lys372A and Thr548A amino acid residue of 1PK0. In case of deoxycytidine kinase(2NOA), Streblusol E have vanderwaal's interactions with Ile30A, Glu53A, Val55A, Trp58A, Met85A, Tyr86A, Phe96A, Gln97A, Arg104A, Arg128A, Asp133A, Phe137A, Leu141A, Arg192A, Arg194A, Glu197A and Tyr204A amino acid residues while having hydrophobic interactions with Met85A, Phe137A and Leu141A amino acid residues of 2NOA. In addition to these, Streblusol E also have aromatic interactions (Fig. 3) with Phe96A & Phe137A and hydrogen bonding with Arg104A & Arg194A amino acid residues of 2NOA i.e. deoxycytidine kinase. Similarly, reference ligand, Lamivudine have vanderwaal's interactions with Ile30A, Glu53A, Val55A, Trp58A, Leu82A, Met85A, Tyr86A, Phe96A, Gln97A, Arg104A, Arg128A, Asp133A, Phe137A and Arg194A amino acid residues while having hydrophobic interactions with Ile30A, Val55A, Leu82A, Met85A, Ala100A, Asp133A and Phe137A amino acid residues of 2NOA. Along with these, lamivudine do have charge effect with Asp133A and hydrogen bonding with Gln97A and Arg128A amino acid residues of 2NOA i.e. deoxycytidine kinase. In case of human nucleoside diphosphate kinase(3FKB), Streblusol E have vanderwaal's interactions with Lys16D, Arg92D, Thr98D, Arg109D, Val116D, Gly117D, Asn119D, Gly122D and Gly123D amino acid residues while having hydrophobic interactions with Val116D and Gly117D amino acid residues of 3FKB. In addition to these, streblusol E do exert hydrogen bonding (Fig. 4) with Arg92D, Thr98D, Arg109D and Gly123D amino acid residues of 3FKB i.e. human nucleoside diphosphate kinase. Similarly, reference ligand, tenofovir diphosphate have vanderwaal's interactions with Lys16D, Tyr56D, His59D, Arg62D, Phe64D, Leu68D, Arg92D, Thr98D, Arg109D, Val116D, Gly117D, Asn119D and Gly122D amino acid residues while having hydrophobic interactions with Phe64D, Leu68D, Thr98D, Val116D and Gly117D amino acid residues of 3FKB. Along with these, tenofovir do have charge effect with Glu58D and hydrogen bonding with Lys16D, His59D and Arg92D amino acid residues of human nucleoside diphosphate kinase. In case of human hepatitis B viral capsid(HBCAG)(1QGT), Streblusol E have vanderwaal's interactions with Gln57B, Ala58B, Leu60B, Cys61B, Glu64B, Gln57A, Ala58A, Cys61A, Lys96A, Ile97A and Leu100A amino acid residues while having hydrophobic interactions with Gln57B, Ala58B, Cys61B, Ala58A and Cys61A amino acid residues of 1QGT. In addition to these, streblusol E do have hydrogen bonding (Fig. 5) with Lys96A of human hepatitis B viral capsid. In case of hepatitis B X-interacting protein(3MSH), Streblusol E have vanderwaal's interactions with Glu40A, His41A, Val44A, Ile45A, Leu48A, Leu67A, Ile74A and Ile76A amino acid residues while having hydrophobic interactions with, His41A, Val44A, Ile45A, Leu48A and Ile74A amino acid residues of 3MSH. Along with these, streblusol E do exert aromatic interactions/pi-staking (Fig. 6) with His41A of hepatitis B X-interacting protein. In case of human hepatitis B viral capsid protein, the active site for streblusol E consist amino acid residues of chain A and chain B of this capsid. It has been found that both the aromatic rings played a great role in pi-pi staking, while the three hydroxyl groups possess hydrogen bonding with the amino acids of target proteins.

It was well emphasized in this present study that the crystal structure of the EF3-CaM complexed with PMEApp(1PK0) was docked with various drugs along with streblusol E, adefovir diphosphate, lamivudine, tenofovir mono phosphate and tenofir diphosphate. Interestingly, it was notable that streblusol E could bind with the EF3-CaM significantly like the standard drug adefovir diphosphate. It did dock evidently with other proteins like deoxycytidine kinase complexed with Lamivudine & ADP (2NOA), NDPK H122G and Tenofovir-diphosphate (3FKB), Hepatitis B Viral Capsid (HBCAG) (1QGT), crystal structure of Hepatitis B X-interacting protein at high resolutions (3MSH) (Table 1). From the previous literature, it was well identified that EF3-CaM was well related to viral hepatitis B [14]. The molecular docking of sreblusol E in the present study was cohesive to this protein and it is thus evident that the action against hepatitis B protein can be possible by streblusol E.

As per the Lipinski rule of five, Streblusol E stands a good chance to be drug (Table 2). Further the logs of 3.244 and logP of 2.664 made it a good molecule which will be having better dissolution and bioavailability because of its absorption via human intestine also. hERG channel inhibition is also below 5, which is good signal. As plama protein binding is high, so the half life of the drug, Streblusol E is expected to be high and have a longer duration of action. Streblusol E will not be able to cross blood brain barrier, so side effect will not be related to that of brain like nausea etc. As this molecule is not a P-gp substrate, so likeliness of resistance will be minimal.

Moreover, as per the data of Derek Nexus, Streblusol E doesn't have potential toxicities like carcinogenicity, mutagenicity, genotoxicity and developmental toxicity, but found to be possible hepatotoxic, skin sensitizer and can damage chromosome Table 3).

These results will certainly attract the attention of researchers worldwide to derivatize streblusol E and clinically developed this class to reach bedside as alternative and complementary medicine for the treatment of acute or chronic hepatitis B viral infection.

Table 1. Docking studies of streblusol E and reference ligands with various targeted proteins.

Proteins under Docking Study	**Dock Score**				
	Streblusol E	**Adefovir Diphosphate**	**Lamivudine**	**Tenofovir Mono Phosphate**	**Tenofovir Diphospha-te**
Crystal Structure of the EF3-CaM complexed with PMEApp(**1PK0**)	-68.075706	-85.297596	-	-	-
The structure of deoxycytidine kinase complexed with Lamivudine & ADP(**2NOA**)	-36.970556	-	-79.054872	-	-
Structure of NDPK H122G and Tenofovir-diphosphate(**3FKB**)	-54.745946	-	-	-76.971639	-75.772605
Human Hepatitis B Viral Capsid(HBCAG)(**1QGT**)	-59.566571	-	-	-	-
Crystal structure of Hepatitis B X-interacting protein at high resolution(**3MSH**)	-47.307476	-	-	-	-

Table 2. ADME profile prediction of Streblusol E.

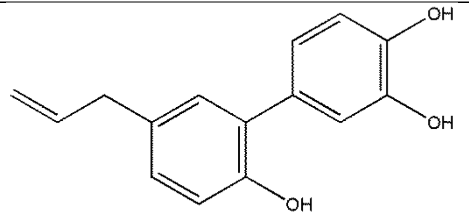
ID	Streblusol E
Structure	
MW	242.3
HBD	3
HBA	3
TPSA	60.69
Flexibility	0.1579
Rotatable Bonds	3
P450_3A4_CSL	0.9901
LogS	3.244
logS @ pH7.4	3.244
logP	2.664
logD	2.664
2C9 pKi	5.039
hERG pIC50	4.867
BBB log([brain]:[blood])	-0.3491
BBB category	-
HIA category	+
P-gp category	no
2D6 affinity category	medium
PPB90 category	high

Table 3. Toxicity Profile Prediction of Streblusol E

Carcinogenicity	No report
Photocarcinogenicity	No report
Hepatotoxicity	**Plausible**
Genotoxicity in vitro	No report
Genotoxicity in vivo	No report
Photogenotoxicity in vitro	No report
Photogenotoxicity in vivo	No report
Chromosome damage in vitro	**Plausible**
Chromosome damage in vivo	No report
Photo-induced chromosome damage in vitro	No report
alpha-2-mu-Globulin nephropathy	No report

Anaphylaxis	No report
Bladder urothelial hyperplasia	No report
Cardiotoxicity	No report
Cerebral oedema	No report
Chloracne	No report
Cholinesterase inhibition	No report
Cumulative effect on white cell count and immunology	No report
Cyanide-type effects	No report
High acute toxicity	No report
Methaemoglobinaemia	No report
Nephrotoxicity	No report
Neurotoxicity	No report
Oestrogenicity	No report
Peroxisome proliferation	No report
Phospholipidosis	No report
Phototoxicity	No report
Pulmonary toxicity	No report
Uncoupler of oxidative phosphorylation	No report
Irritation (of the eye)	No report
Irritation (of the gastrointestinal tract)	No report
Irritation (of the respiratory tract)	No report
Irritation (of the skin)	No report
Lachrymation	No report
HERG channel inhibition in vitro	No report
Thyroid toxicity	No report
Photoallergenicity	No report
Skin sensitisation	**Plausible**
Occupational asthma	No report
Respiratory sensitisation	No report
Developmental toxicity	No report
Teratogenicity	No report
Testicular toxicity	No report
Ocular toxicity	No report
Mutagenicity in vitro	No report
Mutagenicity in vivo	No report
Photomutagenicity in vitro	No report

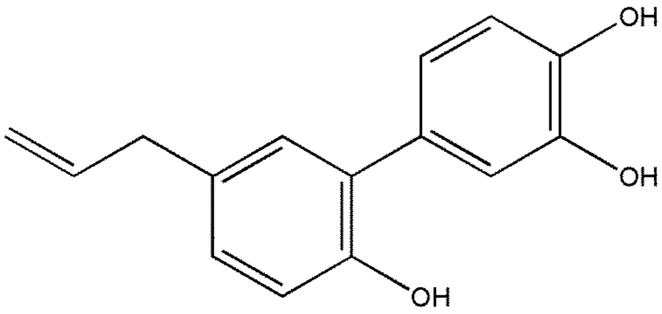


Figure 1. Structure of Streblusol E

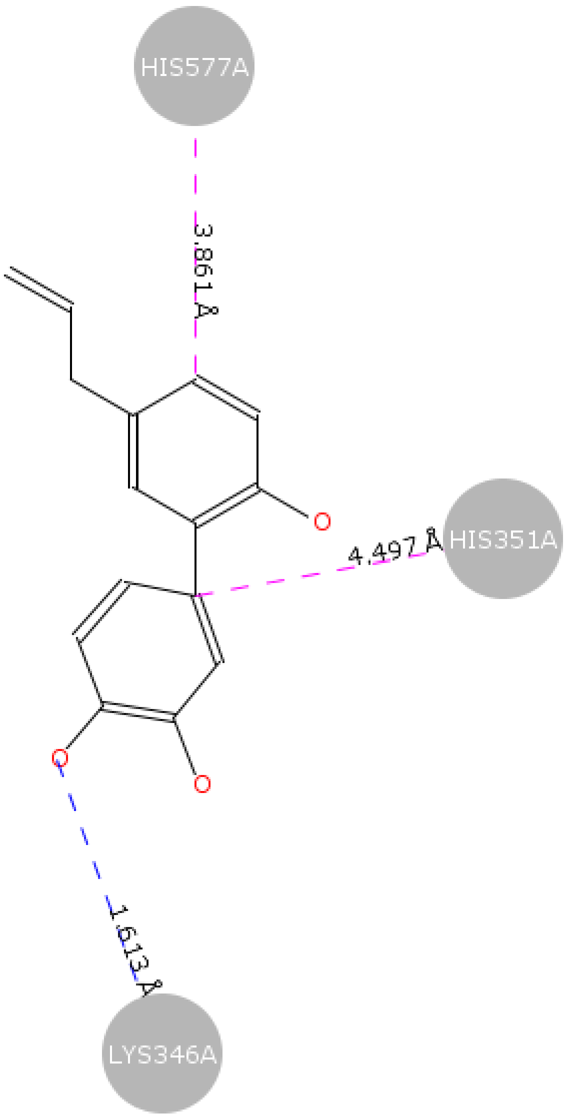


Figure 2. Hydrogen bonding and aromatic interactions of Streblusol E with EF3-CaM adenylyl cyclase(1PK0); Blue Broken Line- Hydrogen Bonding; Pink Broken Line: Aromatic/pi-pi staking.

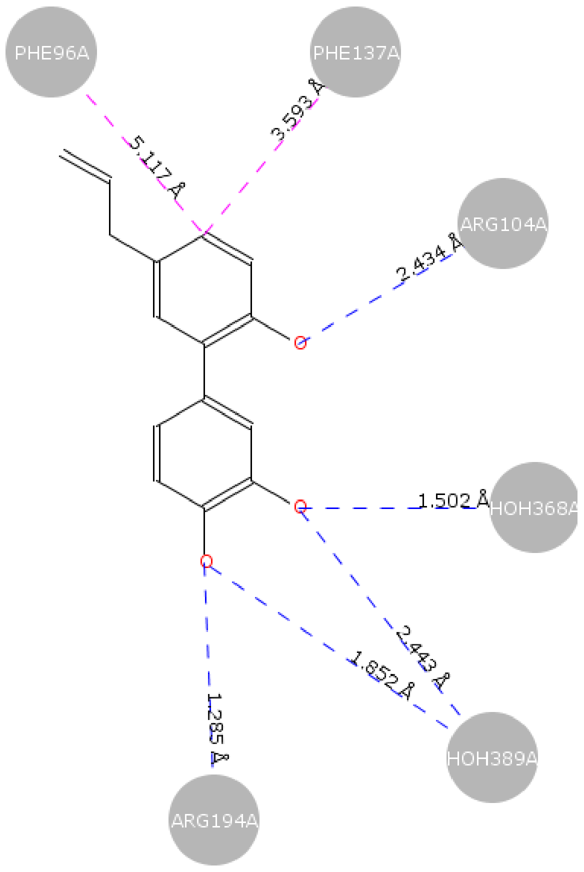


Figure 3. Hydrogen bonding and aromatic interactions of Streblusol E with deoxycytidine kinase(2NOA); Blue Broken Line- Hydrogen Bonding; Pink Broken Line: Aromatic/pi-pi staking.

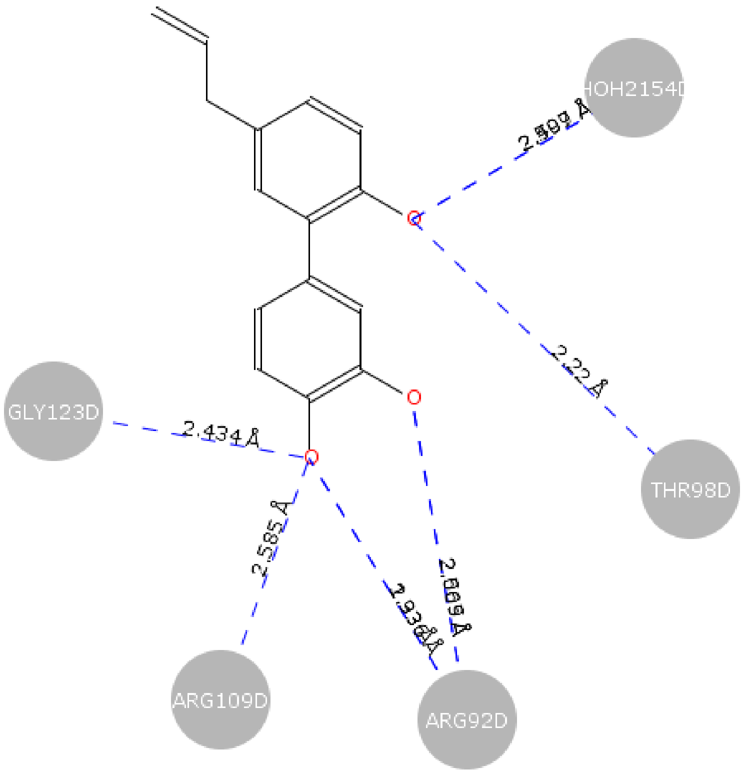


Figure 4. Hydrogen bonding of Streblusol E with human nucleoside diphosphate kinase(3FKB) ; Blue Broken Line- Hydrogen Bonding.

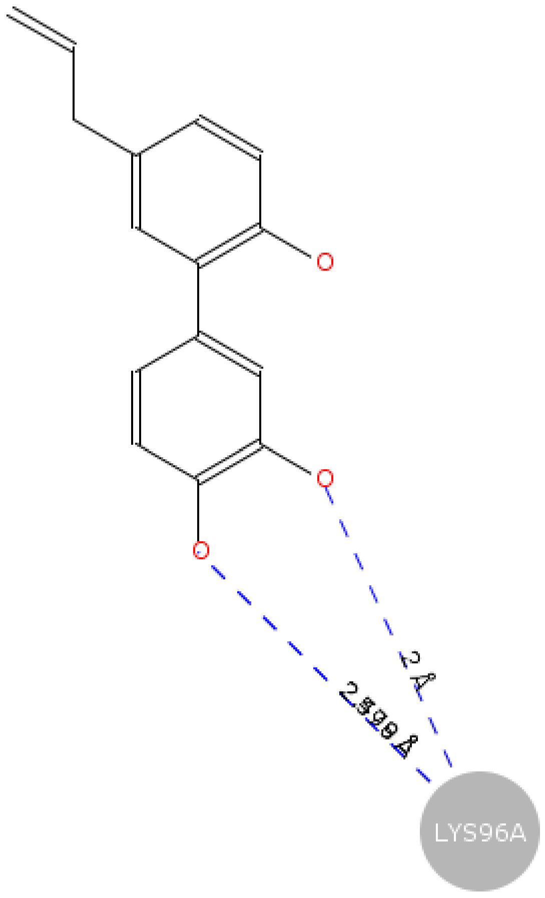


Figure 5. Hydrogen bonding of Streblusol E with Human Hepatitis B Viral Capsid(1QGT); Blue Broken Line- Hydrogen Bonding.

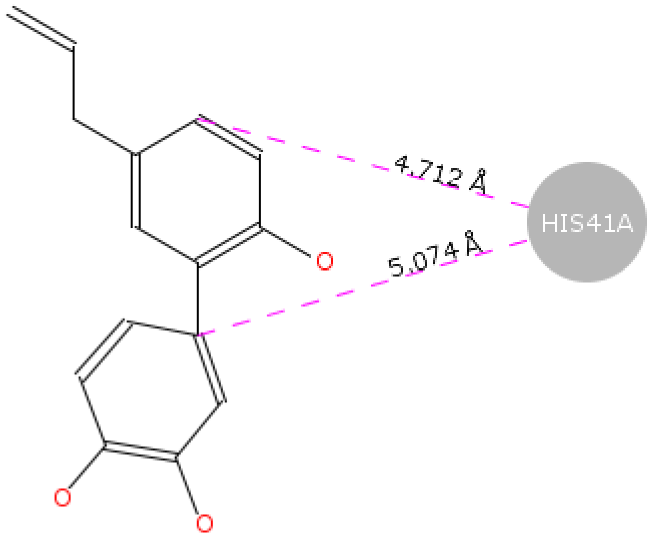


Figure 6. Aromatic/pi-pi staking of Streblusol E with hepatitis B X-interacting protein(3MSH); Pink Broken Line: Aromatic/pi-pi staking.

3. Materials and Methods

Proteins used for GRIP Docking

EF3-CaM complexed with PMEApp (1PK0): Adefovir dipivoxil, a drug approved to treat chronic infection of hepatitis B virus, effectively inhibit EF-induced cAMP accumulation by inhibiting calmodulin(CaM)-activated adenylyl cyclase (Shen et al., 2004).

Deoxycytidine kinase complexed with Lamivudine & ADP (2NOA): L-nucleoside analogs represent an important class of small molecules for treating both viral infections and cancers. These pro-drugs achieve pharmacological activity only after enzyme-catalyzed conversion to their tri-phosphorylated forms. Crystal structure of human deoxycytidine kinase (dCK) in complex with the L-nucleosides (-)-beta-2',3'-dideoxy-3'-thiacytidine (3TC), an approved anti-human immunodeficiency virus (HIV) agent and troxacitabine (TRO), an experimental anti-neoplastic agent was used. The first step in activating these agents is catalyzed by dCK. The capability of dCK to phosphorylate both D- and L-nucleosides and nucleoside analogs derives from structural properties of both the enzyme and the substrates themselves (Sabini et al., 2007).

NDPK H122G and Tenofovir-diphosphate (3FKB): Tenofovir is an acyclic phosphonate analog of deoxyadenylate used in AIDS and hepatitis B therapy. Tenofovir diphosphate, its active form can be produced by human nucleoside diphosphate kinase (NDPK), but with low efficiency, and that creatine kinase is significantly more active (Koch et al., 2009).

Human Hepatitis B Viral Capsid (HBCAG) (1QGT): Hepatitis B is a small enveloped DNA virus that poses a major hazard to human health. The crystal structure of the T = 4 capsid has been used. The monomer fold is stabilized by a hydrophobic core that is highly conserved among human viral variants. Association of two amphipathic alpha-helical hairpins results in formation of a dimer with a four-helix bundle as the major central feature. The capsid is assembled from dimers via interactions involving a highly conserved region near the C terminus of the truncated protein used for crystallization. The major immunodominant region lies at the tips of the alpha-helical hairpins that form spikes on the capsid surface (Wynne et al., 1999).

Hepatitis B X-interacting protein at high resolution (3MSH): Hepatitis B X-interacting protein (HBXIP) is a ubiquitous protein that was originally identified as a binding partner of the hepatitis B viral protein HBx. HBXIP is also thought to serve as an anti-apoptotic cofactor of survivin, promoting the suppression of pro-caspase-9 activation (Garcia-Saez et al., 2011).

Docking studies

Vlife MDS 4.4 is very robust software with inclusion of all the necessary simulation modules. The structure of streblusol E in the study has been drawn in the 2D drawing application (2D Draw app) of MDS 4.3, followed by its conversion into 3D form by using default conversion procedure. Best conformer with the minimum energy was used for the docking analysis (Singla and Bhat, 2010; Singla et al., 2013). Molecular docking energy evaluations are usually carried out with the help of scoring function like dock score, PLP score, potential of mean force (PMF) score, steric and electrostatic score, etc. The PLP function is incorporated by the MDS Vlife Science software in the GRIP docking method which calculates the ligand-receptor binding affinity in terms of the PLP score. The PLP score is designed to enable flexible docking of ligands to perform a full conformational and positional search within a rigid binding site. Streblusol E was docked into the active site of 1PK0, 1QGT, 2NOA, 3FKB and 3MSH that can be obtained in the co-crystallized with adefovir diphosphate, lamivudine, tenofovir monophosphate & tenofovir diphosphate or by the use of cavities. The parameters fixed for docking simulation was like this- number of placements: 50, rotation angle: 10o, exhaustive method, ligand-wise results: 10, scoring function: PLP score. By rotation angle, ligand would be rotated inside the receptor cavity to generate different ligand poses inside the receptor cavity. By placements, the method will check all the 50 possible placements into the active site pocket and will result out best placements out of 50. After docking simulation, the best docked conformer of streblusol E and reference ligands were then checked for their interactions with targeted proteins like hydrogen bonding, hydrophobic, pi-staking/aromatic, charge and vanderwaal's interactions (Igoli et al., 2014a, 2014b; Malleshappa and Patel, 2013; Singla et al., 2012; Singla, 2014; VLife , 2013) .

Drug Metabolism and Pharmacokinetics

Various parameters like hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), total polar surface area (TPSA), flexibility, rotatable bonds, solubility (LogS), solubility at pH 7.4 (LogS @ pH 7.4), lipophilicity (LogP), lipophilicity at pH 7.4 (LogP @ pH 7.4), affinity towards cytochrome P450 2C9 isoform (2C9 pKi), hERG inhibition (hERG pIC50), blood brain barrier crossing capability (BBB log([brain]:[blood]); BBB category) , human intestinal absorption (HIA category), substrate for P-glycoprotein (P-gp category),), affinity towards cytochrome P450 2D6 isoform(2D6 affinity category), plasma protein binding (PPB90 category) and composite site lability of Streblusol E on three isoforms of cytochrome P450 i.e 3A4, 2D6 and 2C9 were calculated using StarDrop software of Optibrium Ltd (Optibrium Ltd).

In Silico Prediction of Toxicity

Using Derek Nexus module in StarDrop(liaison between Optibrium Ltd. and LHASA ltd), probability of Streblusol E to exert toxicity against various toxicological endpoints like carcinogenicity, mutagenicity, genotoxicity etc were calculated and reporting under four reasoning level like

- No Report – no evidence of toxicity or nothing to report
- Probable- there is atleast on strong argument for the proposition and none against it
- Plausible – the weight of evidence supports the proposition
- Equivocal – there is an equal weight of evidence for and against the proposition (Lhasa Ltd.; Segall and Barber, 2014).

4. Conclusions

Streblusol E, bioactive from Streblus asper has potential to inhibit hepatitis B virus. Its in silico studies with various target proteins further strengthens its efficacy as anti-hepatis B viral agents.

Acknowledgments

Docking study was performed on Vlife MDS 4.4, funded by Science & Engineering Research Board (Department of Science & Technology), Government of India. Author Rajeev K Singla expresses his gratitude to Government of India for providing Young Scientist's Fellowship vide Project No. SR/FT/LS-149/2011.

Author Contributions

RKS arranged the financial grant for this project as well as collected the data. RG and BD had helped in the data analysis and manuscript drafting. VBG lead the study, analyse the data, finalize the manuscript. All authors have read and approved the final version of manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1) Aeri, V.; Alam, P.; Ali, M. et al. Isolation of new aliphatic ester linked with δ-lactone cos-11-enylpentan-1-oic-1,5-olide from the roots of Streblus asper Lour. *Indo Global J. Pharm. Sci.* **2012**, *2(2)*, 114-120.

2) Garcia-Saez, I.; Lacroix, F.B.; Blot, D. et al. Structural characterization of HBXIP: the protein that interacts with the anti-apoptotic protein survivin and the oncogenic viral protein HBx. *Mol. Biol.* **2011**, *405*, 331-340.

3) Igoli, J.O.; Gray, A.I.; Clements, C.J.; Kantheti, P.; Singla, R.K. Antitrypanosomal activity & docking studies of isolated constituents from the lichen cetraria islandica : possibly multifunctional scaffolds. *Curr. Top. Med. Chem.* **2014**a, *14*, 1014-1021.

4) Igoli, N.P.; Clements, C.J.; Singla, R.K.; Igoli, J.O.; Uche, N.; Gray, A.I. Antitrypanosomal activity & docking studies of components of crateva adansonii DC leaves: novel multifunctional scaffolds. *Curr. Top. Med. Chem.* **2014**b, *14*, 981-990.

5) Jun, L.; Huang, Y.; Guan, X.L. et al. Anti-hepatitis B virus constituents from the stem bark of Streblus asper. *Phytochemistry*. **2012**, *82*, 100-109.

6) Koch, K.; Chen, Y.X.; Feng, J.Y. et al. Nucleoside diphosphate kinase and the activation of antiviral phosphonate analogs of nucleotides: binding mode and phosphorylation of tenofovir derivatives. Nucleosides Nucleotides *Nucleic Acids*. **2009**, *28*, 776-792.

7) LHASA Ltd. Url: http://www.lhasalimited.org/ Accessed on 25.08.2015

8) Malleshappa, N.N.; Patel, H.M. A comparative QSAR analysis and molecular docking studies of quinazoline derivatives as tyrosine kinase (EGFR) inhibitors: A rationale approach to anticancer drug design. *J. Saudi Chem. Soc.* **2013**, *17(4)*, 361-379.

9) Optibrium Ltd. Url: http://www.optibrium.com/stardrop/ Accessed on 25.08.2015

10) Sabini, E.; Hazra, S.; Konrad, M. et al. Structural basis for activation of the therapeutic L-nucleoside analogs 3TC and troxacitabine by human deoxycytidine kinase. *Nucleic Acids Res.* **2007**, *35*, 186-192.

11) Segall, M.D.; Barber, C. Addressing toxicity risk when designing and selecting compounds in early drug discovery. *Drug Discov. Today*. **2014**, *19(5)*, 688-693.

12) Shen, Y.; Zhukovskaya, N.L.; Zimmer, M.I. et al. Selective inhibition of anthrax edema factor by adefovir, a drug for chronic hepatitis B virus infection. *Proc. Natl. Acad. Sci. USA*. **2004**, *101*, 3242-3247.

13) Singla, R.K.; Bhat, V.G. QSAR model for predicting the fungicidal action of 1,2,4-triazole derivatives against Candida albicans. *J. Enz. Inhib. Med. Chem.* **2010**, *25(5)*, 696-701.

14) Singla, R.K.; Paul, P.; Nayak, P.G.; Bhat, V.G. Investigation of anthramycin analogs induced cell death in MCF-7 breast cancer cells. *Indo Global J. Pharm. Sci.* **2012**, *2(4)*, 383-389.

15) Singla, R.K.; Bhat, V.G.; Kumar, T.N.V.G. 3D-quantitative structure activity relationship: a strategic approach for in silico prediction of anti-candididal action of 1, 2, 4-triazole derivatives. *Indo Global J. Pharm. Sci.* **2013**, *3(1)*, 52-57.

16) Singla, R.K. Mechanistic evidence to support the anti-hepatitis B viral activity of multifunctional scaffold & conformationally restricted magnolol. *Natl. Acad. Sci. Lett.* **2014**, *37(1)*, 45-50.

17) Sripanidkulchai, B.; Junlatat, J.; Wara-aswapati, N.; Hormdee, D. Anti-inflammatory effect of Streblus asper leaf extract in rats and its modulation on inflammation-associated genes expression in RAW 264.7 macrophage cells. *J. Ethnopharmacol.* **2009**, *124*, 566-570.

18) Wynne, S.A.; Crowther, R.A.; Leslie, A.G. The crystal structure of the human hepatitis B virus capsid. *Mol. Cell.* **1999**, *3*, 771-780.

19) VLifeMDS: Molecular Design Suite. VLife Sciences Technologies Pvt. Ltd., Pune, India. 2013. (www.vlifesciences.com)

SciForum
Mol2Net

Homology Modeling, Molecular Dynamic Simulation and in Silico Screening of Activator for the Intensification of Human Sirtuin Type 1 (SIRT1) by novel 1, 3, 4-Thiadiazole Derivatives-A Potential Antiaging Approach

Sudipta Saha, Amit Rai, Mahendra Singh, Vinit Raj *, Durgesh Kumar and Anil Kumar Sahdev
Department of Pharmaceutical Sciences, Babasaheb Bhimrao Ambedkar University, Vidya Vihar, Rai Bareli Road, Lucknow 226025

* Author to whom correspondence should be addressed; E-Mail: raj.vinit24@gmail.com;
Tel.: +918859383897.

Published: 4 December 2015

Abstract: Sirtuin type-1(SIRT1) is a regulator of various biosynthetic pathways via activation of peroxisome proliferator-activated receptor-γ and interacting with adenosine-mono-phosphate kinase. SIRT1 is the important target for various neurodegenerative, cancer and metabolic disorders as well as aging medicine. Keeping in view of the above fact, we considered novel 1,3,4-thiadiazole derivatives series for SIRT1 screening, which was performed through virtual screening, homological modeling, docking and computational studies. On the basis of available molecular structure in protein data bank of SIRT1 protein, we calculated the interaction energy designed molecules. The interaction energy of designed compound VR3 closely better than resveratrol (~ 6.4 kcal/mol). Among of them the VR 3 shown the best conformation fitting stability in the binding site of SIRT1 predicted by MD (Molecular dynamics) simulation for 2.5ns. Therefore, the designed compounds have good binding affinities to SIRT1 target, would serve better lead compound for antiaging screening for future drug design perspective.

.

.

..

Keywords Homology modeling, Molecular docking, ADMET prediction, actives substrate binding domain of SIRT1, Antiaging

1. Introduction

In the last few years, sirtuin (SIRT) has become a large attention to scientific communities for developing lead optimization. Interesting in this protein family is to its crucial role in genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient sensing and mitochondrial dysfunction (López-Otín, Blasco, Partridge, Serrano, & Kroemer, 2013). SIRT1 downregulates pro-inflammatory factors like p53[2-3] and nuclear factor-kappa B (NF-κB), whereas upregulates peroxisome proliferator-activated receptor-gamma coactivator 1 alpha (PGC-1α)(Amat et al., 2009; Wareski et al., 2009) and forkhead box class O transcription factors (FOXOs). SIRT 1 is the main target for diverse pharmacological properties including neurodegenerative disorders, cancer and metabolic disorders as well as aging medicine (Pallàs et al., 2009).

Because of that, the discovery of SIRT1 activator is an important target for drug discovery. In this study, we focused on developing a new scaffold which can be able to have potent activation effect. During searching of new lead for SIRT1 target, we used 1, 3, 4-thiadiazole moiety as it has one hydrogen binding domain and two-electron donor system. The previous literature survey suggested that 1, 3, 4-thiadiazole is the important pharmacophore than other isomers for binding to the receptor and it has multiple pharmacological actions as well. This ring exhibited antimicrobial (Demirbas, Karaoglu, Demirbas, & Sancak, 2004; Karegoudar et al., 2008), anticancer (Chou et al., 2003), antanxiety, anti-depressant (Clerici et al., 2001), anti-oxidant properties (Martinez et al., 1999), anticonvulsant activity (Yusuf, Khan, Khan, & Ahmed, 2013) and antitubercular activities (Alegaon et al., 2012). Thiadiazole ring expressed diverse biological activities, might be due to the presence of =N-C-S moiety (Oruç, Rollas, Kandemirli, Shvets, & Dimoglo, 2004).

In view of the above fact, the question arose whether 1, 3, 4-thiadiazole might be an important activator for SIRT1 target. To prove this hypothesis, homology modeling was performed using one or more known protein structures that are resembling to the structural sequence of SIRT1. Later, all these sequences collapsed together to reach the desired template sequence. Finally, docking studies was carried out between newly designed protein and prepared ligand to get interaction energy.

After that, the pharmacokinetics parameters (ADME, BBB and toxicity) were also measured with that designed compound to rule out whether these compounds might be suitable for *in vivo* biological system. We hypothesized that these compounds may be a lead target for antiaging as a SIRT1 agonist and also suitable for *in vivo* screening in the future.

Material and methods

In this present study, National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov) and Protein Data Bank (PDB) (http://www.rcsb.org) were used as chemical sources. The software which was used to experiment, tabulated in table (1). Resveratrol, VR1, VR2, and VR3 structures were drawn through Chemdraw Ultra 10.0 figure (1) and their geometry was optimized six times with Gauss view 5.0. However, these structures represent a minimum energy optimization, selected for *in-silico* study and 3-D structure of sirtuin type-1 Protein structure was not available in the PDB and NCBI Protein database. The 3-D structure of the protein was prepared through the run blast of protein on the database of PDB that had shown the description of sequences producing significant alignments of query cover and identity (37 and 97%, respectively). Finally, prepared 3-D protein structure was used for homological modeling figure (2).

..
.
.

.

Table 1 Softwares used for modeling and their Purposes

Softwares	**Purposes**
Chemdraw Ultra 10.0 and open babel GUI	For drawing the chemical structure and convert into PDB format
Argus Lab software,Gauss view 5.0	For optimizing the geometry of derivatives
1: 3-D pssm sever, phyre 2.0 server, easy moduller 2.0, swiss pdb viewer (spdbv 4.1.0), Modbase server and saves server plot.	For the homological modeling of Sirtuin type 1
Autodock 4.0,Discovery studio and autodock vina	For docking studies
Molinspiron software toolkit, Med Chem Designer and Lasar toxicity prediction service	For characterization of the derivatives

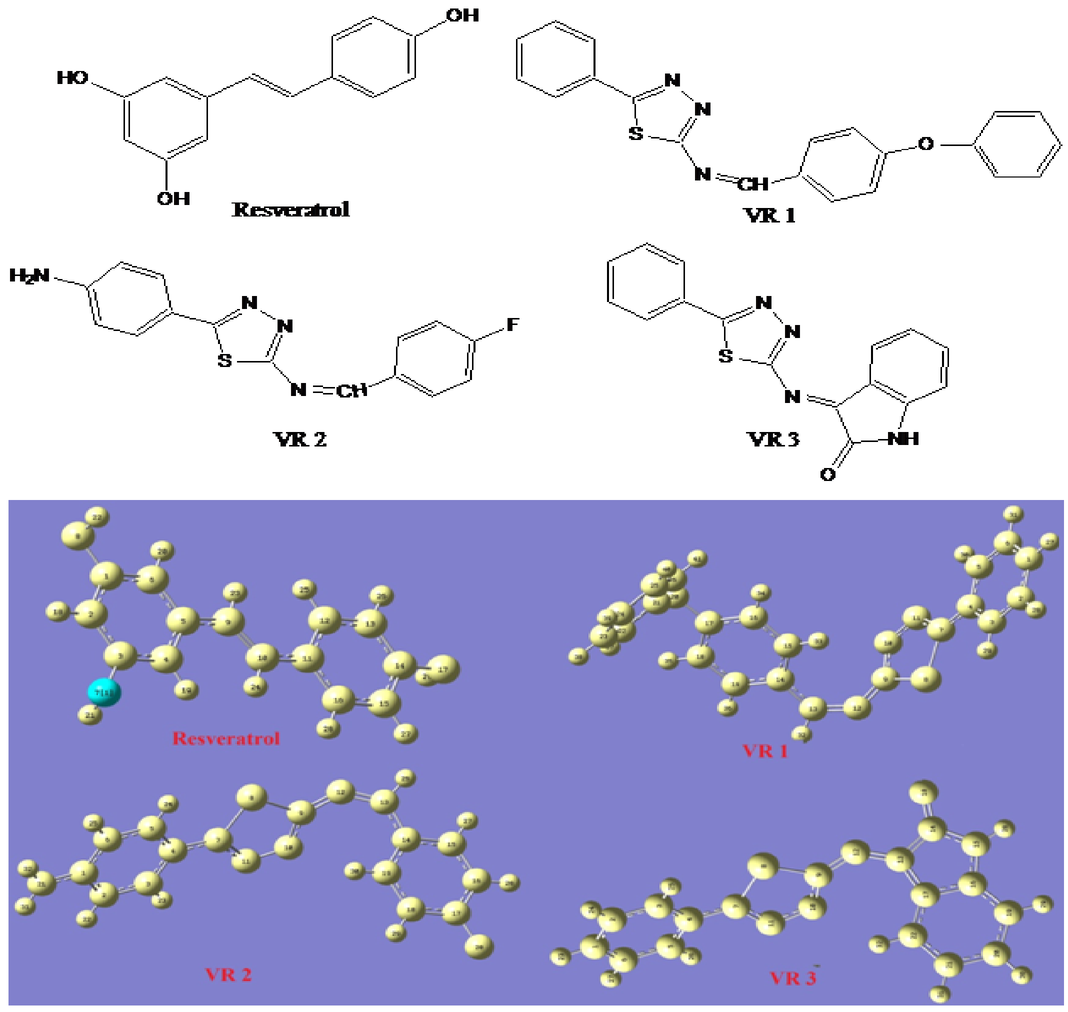


Figure 1: Structures of standard and designed compounds with optimized geometry

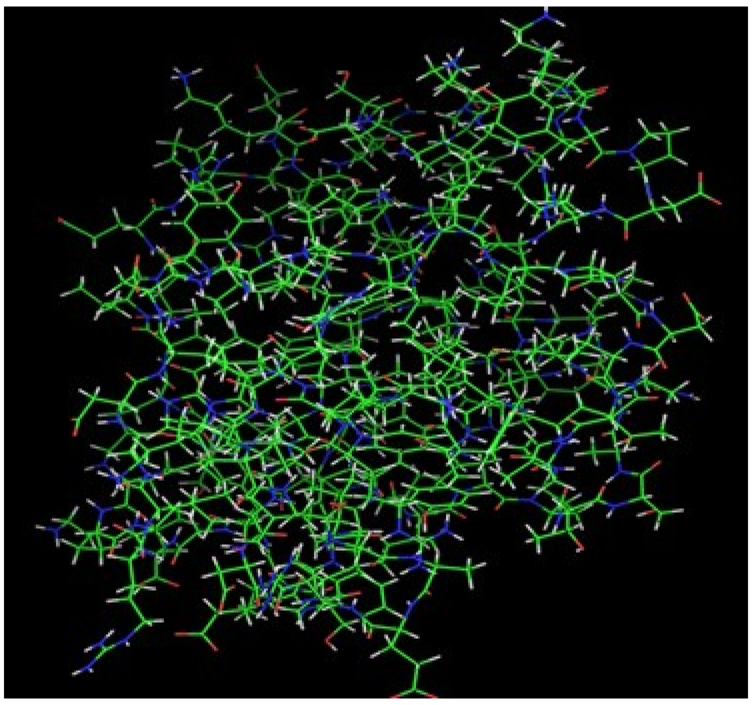


Figure 2: 3-Dimensional Structure Prediction of Sirtuin type 1 Protein of Homo sapiens

Amino acid sequences of Sirtuin type 1 Protein

GenBank: AAD40849.2
GenPept Graphics
>gi|7555471|gb|AAD40849.2|AF083106_1
sirtuin type 1 [Homo sapiens]

```
MADEAALALQPGGSPSAAGADREAASSPA
GEPLRKRPRRDGPGLERSPGEPGGAAPERE
VPAAARGCPGAAAAALWREAEAEAAAAG
GEQEAQATAAAGEGDNGPGLQGPSREPPL
ADNLYDEDDDDEGEEEEEAAAAAIGYRDN
LLFGDEIITNGFHSCESDEEDRASHASSSDW
TPRPRIGPYTFVQQHLMIGTDPRTILKDLLP
ETIPPPELDDMTLWQIVINILSEPPKRKKRK
DINTIEDAVKLLQECKKIIVLTGAGVSVSCG
IPDFRSRDGIYARLAVDFPDLPDPQAMFDIE
YFRKDPRPFFKFAKEIYPGQFQPSLCHKFIA
LSDKEGKLLRNYTQNIDTLEQVAGIQRIIQC
HGSFATASCLICKYKVDCEAVRGDIFNQVV
PRCPRCPADEPLAIMKPEIVFFGENLPEQFH
RAMKYDKDEVDLLIVIGSSLKVRPVALIPSS
IPHEVPQILINREPLPHLHFDVELLGDCDVII
NELCHRLGGEYAKLCCNPVKLSEITEKPPR
TQKELAYLSELPPTPLHVSEDSSSPERTSPPD
SSVIVTLLDQAAKSNDDLDVSESKGCMEEK
PQEVQTSRNVESIAEQMENPDLKNVGSSTG
EKNERTSVAGTVRKCWPNRVAKEQISRRL
DGNQYLFLPPNRYIFHGAEVYSDSEDDVLS
SSSCGSNSDSGTCQSPSLEEPMEDESEIEEFY
NGLEDEPDVPERAGGAGFGTDGDDQEAIN
EAISVKQEVTDMNYPSNKS
```

Homology modeling

SIRT1 was modeled by using Easy Modeler software and self-developed commends. The amino acid sequence of human SIRT1 was retrieved from Gen Bank (accession number: AF083106.2) in NCBI (Frye, 1999). It consists of 747 amino acids, among which residues 250-500 belong to SIRT1. The SIRT1 was then subjected to a PSI-BLAST search in order to identify the homologous proteins. 3-D pssm and phyre 2.0 servers were shown to produce a number of potential templates and identify better templates for sequences, were observed through distant relationships to any solved structure. Easy modeler software and self-developed command was used to generate the nine probabilities of query. Ramachandran plot checked and viewed in the swiss pdb reviewer 4.1.0.

Validation of the modeled structure

The backbone conformation of the modeled structure was calculated by analyzing the phi (Φ) and psi (ψ) torsion angles using Saves server and evaluations of the modeled 3-D structure of sirtuin was performed using PROCHECK by calculating the Ramachandran plot. This plot represented the distribution of the Φ and ψ angles for the amino acid residues.

Docking studies

Docking study of desiged compounds were performed with anti aging, molecular targets, namely SIRT1 by using Autodock 4.0 along with Autodock Vina. Before the docking study, we identified the active site domain with the help of Dog P/Castp active Site recognizer server of protein, wherein the legend showed the best configuration figure (3). Keeping in view active site amino acid sequence, Grid box was set. Their binding affinity (kcal/mol) and count of probable hydrogen bonds also evaluated in the similar experiment.

Prediction of pharmacokinetic properties

The designed compounds assessed for pharmacokinetic properties through medchem designer software. Later, the pharmacokinetic parameters of the lead molecules analyzed, including their absorption, distribution, metabolism and excretion (ADME), using Molinspiration property online calculator (Bratislava). The percentage of absorption (%ABS) was calculated using TPSA by the following formula: % ABS = 109-(0.345×TPSA) (Zhao et al., 2002). Oral bioavailability and blood brain barrier (BBB) penetration of all compounds have performed by the ACD/ Lab-I online server.

Bioactivity prediction and Toxological comparative studies

For prediction of bioactivity and toxicological properties of titles compounds evaluated by Molinspiration property online calculator and the Lasar toxicity prediction server. The designed derivatives and original drug bioactivity predictions had been compared along with some selected activity GPCR (G-Protein coupled receptor).

MD Simulation study

Molecular dynamics simulation has been performed for the higher affinity complex with the help of Yasara tools. Hence, the complex placed in a cubic box and filled with solvent (HOH) by applying AMBER 99 force field, temperature 298 K that controlled through rescale velocities and pressure reached 1.000 bar these parameters were applied in order to check the stability of their complex. figure (4) shows the solvated structure when visualized in MD. After energy minimization of the solvated and electroneutral system its Potential Energy had been analyzed and plotted by using Sigma Plot 11.0 tools. MD simulation was run for 2.5ns. The following parameter evaluated, including: (i) Complex binding energy vs time which indicated that complex stability under the MD simulation figure(5), (ii) Potential energy of complex with respective time figure (6) and (iii) Average RMSD (Root Mean Square Deviation) graph which indicated convergence of the simulated structure towards an equilibrium state with respect to a reference structure (starting structure) figure (7).

Results and discussion

Homology modeling of SIRT1 and its evaluation

Before going to interaction energy analysis through docking studies, it is necessary to prepare an amino acid sequence of targeted SIRT1 in our experiment. Homology modeling and database searching are the key essential part for lead optimization. Total three molecules were used for this study and resveratrol were used as positive control.

For homology modeling, 3-D pssm and phyre 2.0 produced a larger number of potential templates. The selected templates were LJ8F, LMA3, LS5P, 2RCR, LQ2W, LICI, UEK, LRIA, LETP and LOGY with identity 39, 31, 28, 21, 18, 18, 17, 19, 16 and 16%, respectively. These best identical templates were downloaded from the protein data bank (PDB) with consideration of the x-ray diffraction and resolution (R). R value was within the range (not more than 3.0 A^0, and 0.5 obs). Again, Easy modeler software and self-developed commend were used to generate the nine probabilities of query or model. The best dope score was selected which resided with phi\psi out of core regions. Later these data were checked through Ramachandran plot and viewed in the swiss pdb reviewer 4.1.0. Very few numbers of amino acid laid outside the core region. Therefore, loop modeling was done by the mod loop server and checked the output through the Ramachandran plot.

Validation of the modeled structure

The modeled structure of SIRT1 was calculated by analyzing the phi (Φ) and psi (ψ) torsion angles using the Saves server software and evaluations of the modeled 3D structure of certain were performed using PROCHECK by calculating the Ramachandran plot figure (8) and table (2). The percentage Φ and ψ angles were 92.3% of core residues, whereas this percentage was 0.2% for the disallowed residues.

Docking studies

Docking study of designing compounds was performed with antaging molecular targets SIRT1. We detected the active site domain with the help of DogP active Site recognizer server of protein where the legend showed the best configuration. Later, Grid box was set according to an active site sequence of amino acid. Their binding affinity (kcal/mol) and count of probable hydrogen bonds were evaluated table (3) through docking studies. Docking images of resveratrol, VR1, VR2 and VR3 with the target receptors was shown in figure (9). All Compounds exhibited good binding properties with SIRT1 receptor (affinity value -6.4, -7.0, -7.3 and -7.7 kcal/mol and 1, 0, 0, and 0, H-bonds, respectively for resveratrol, VR1, VR2, and VR3). Addition, the interaction of ligand to the receptor has concluded that PRO 419, GLU410 and VAL 412 common essential amino acids, which may be involved in enhancing the efficacy of SIRT1. Hence, this observation could be attributed as potential antigens with SIRT1 mimetic/facilitator mode of action.

Prediction of ADME properties

The ADME properties of the designed compounds were assessed by evaluating their physicochemical properties using the medchem designer software. Their molecular weights were <500 Da; they had <5 hydrogen bond donors and <10 hydrogen bond acceptors, and logP values of <5 table (4). These properties are within the acceptable range of Lipinski's rule of five. Furthermore, the pharmacokinetic parameters of the lead molecules were analyzed, including their ADME using Molinspiration property online calculator and Lasar toxicity prediction server. For the designed compounds, the partition coefficient (QPlogPo/w) and water solubility (QPlogS) values, the %ABS for the compounds ranged from approximately 80 to 95%. These pharmacokinetic parameters are well within the acceptable range defined for human use, thereby indicating their potential as drug-like molecules.

All designed compounds had shown the better BBB penetration power table (5) and the CNS active properties of all compounds were shown in the graph figure (10). Oral bioavailability of all designed compounds were calculated theoretically which lied in accepting a range (more than 70%). BBB penetration and oral bioavailability essential key for the pharmacokinetic profile of the compound to enhance the pharmacological activity. These all theoretically parameter of the designed compounds supported our hypothesis.

MlogP, Moriguchi estimation of logP. S+ log P logP calculated using Simulations Plus' highly accurate internal model; S+logD, logD at user-specified pH (default 7.4), based on S+logP;n-OHNH donor, Number of Hydrogen bond donor protons; M_NO, Total number of Nitrogen and Oxygen atoms; T_PSA, Topological polar surface area in square angstroms; Rule Of Five, Lipinski's Rule of Five: a score indicating the number of potential problems a structure might have with passive oral absorption;miLog P, logarithm of compound partition coefficient between n-octanol and water; log D, logarithm of compound distribution coefficient; n-ROTB, number of rotatable bonds; MV, molecular volume; n-ON acceptor, number of Hydrogen bond acceptor protons.

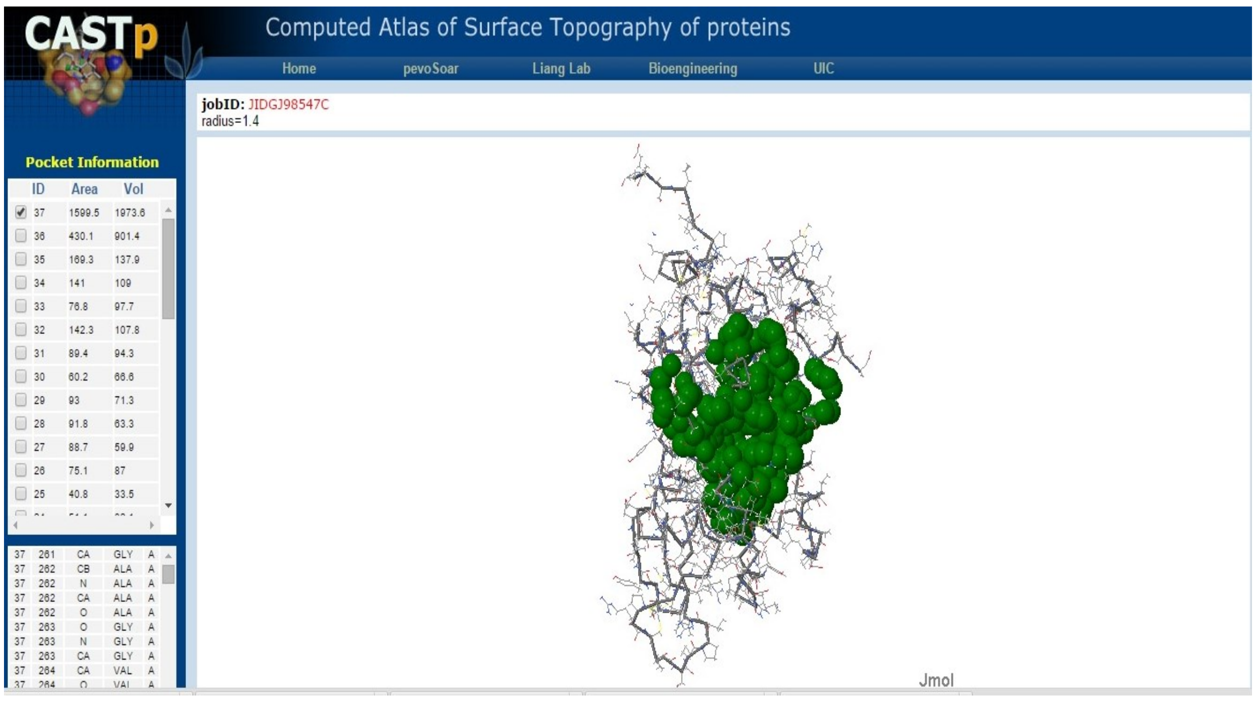
Bioactivity prediction and Toxological comparative studies

In this study, for prediction of bioactivity and toxicological properties of titled compounds was also determined in our study. From all calculated parameters, it can be observed that all titled compounds expressed less affinity to GPCR (G-Protein coupled receptor) ligand, ion channel modulator, kinase inhibitor, nuclear receptor ligand, protease enzyme inhibitor and the toxicological as compared to resveratrol. All this investigation suggested that activation of SIRT1 through our compounds has great importance for providing neuroprotection in various neurodegenerative disorders including the temporal lobe epilepsy (TLE) (Shetty, 2011). The Bioactivity and Toxological data are given in table (6) and (7).

Computational details

A computational study for prediction of docking, energy minimization and ADMET properties of title compounds was performed. From all these parameters, it can be observed that all titled compounds exhibited a good ADMET and BBB properties. None of the compounds violated Lipinski,^{s} parameters, making them potentially promising agents for antiaging durg. From the MD simulation study of compound VR 3 shown the stability of complex at 2.3ns with average energy -388.981 kcal/mol. In addition, the complex didn't show more fluctuation in potential energy in respectively with time. Whereas, binding energy of compound at 0 ps time was founded -189.469 Kcal/mol which increased -645.930 kcal/mol at 700 ps under MD simulation, whereas in the figure (5) the complex exhibited the fluctuation before the 800ps. Later, the complex shows the stability near 800 ps with -400(kcal/mol) compound binding energy. However, the RMSD of the backbone structure shown the stability near 200 ps. These findings suggested that the complex structure of VR 3 with SIRT1 shown the best stable fitting (affinity) in the MD simulation study.

.



Pocket Information

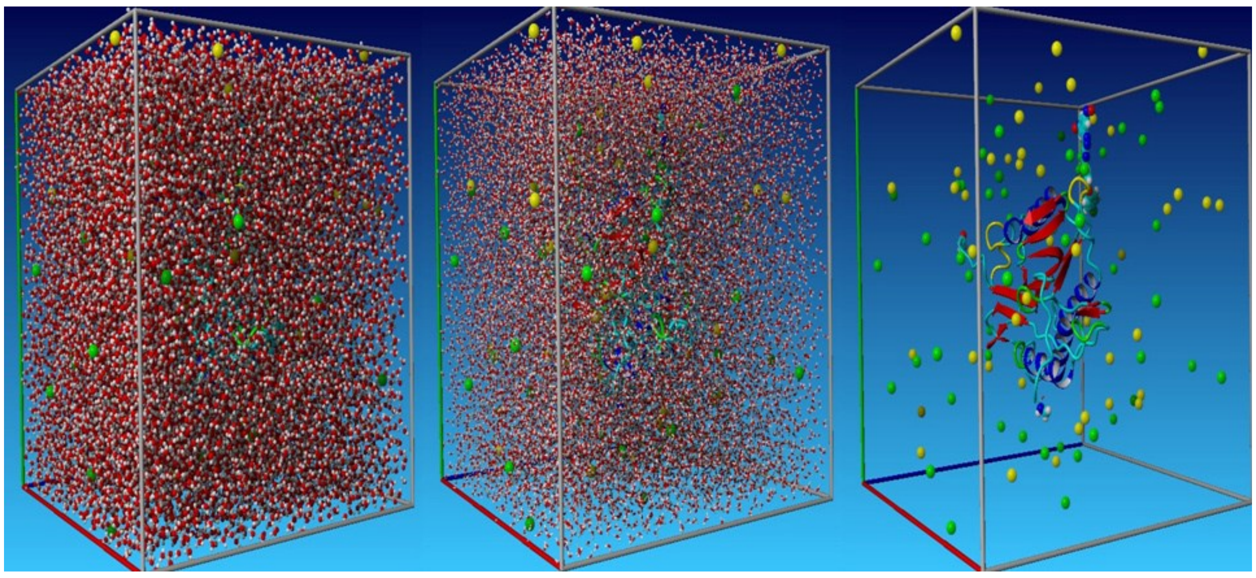
	ID	Area	Vol
☑	37	1599.5	1973.6
□	36	430.1	901.4
□	35	169.3	137.9
□	34	141	109
□	33	76.8	97.7
□	32	142.3	107.8
□	31	89.4	94.3
□	30	60.2	66.6
□	29	93	71.3
□	28	91.8	63.3
□	27	88.7	59.9
□	26	75.1	87
□	25	40.8	33.5

37	261	CA	GLY	A
37	262	CB	ALA	A
37	262	N	ALA	A
37	262	CA	ALA	A
37	262	O	ALA	A
37	263	O	GLY	A
37	263	N	GLY	A
37	263	CA	GLY	A
37	264	CA	VAL	A
37	264	O	VAL	A

jobID: JIDGJ98547C
radius=1.4

Figure 3: Amino acid present in the active site are labeled with green

.



.

Figure 4: Shows solvated structure visualized in MD simulation. Here red color represents the solvent. graph after Energy minimization step.Protein shown in greenish-yellow-red-blue color and ligand shown as white-sky in blue color.

.

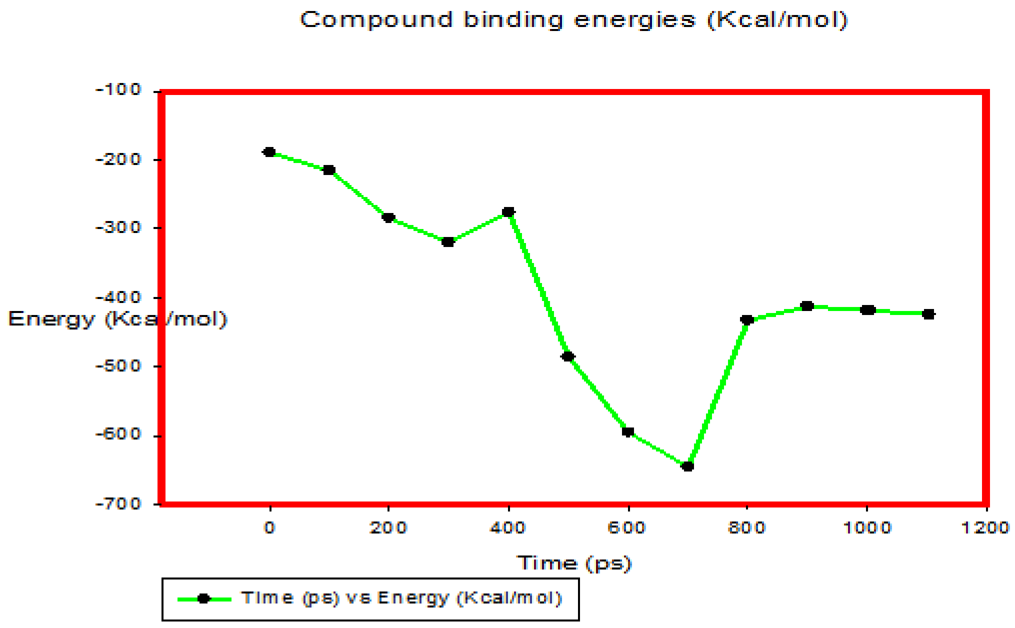


Figure 5: Shown the Time vs compound binding energy, which indicated that the energy stabilized at 1000 ps

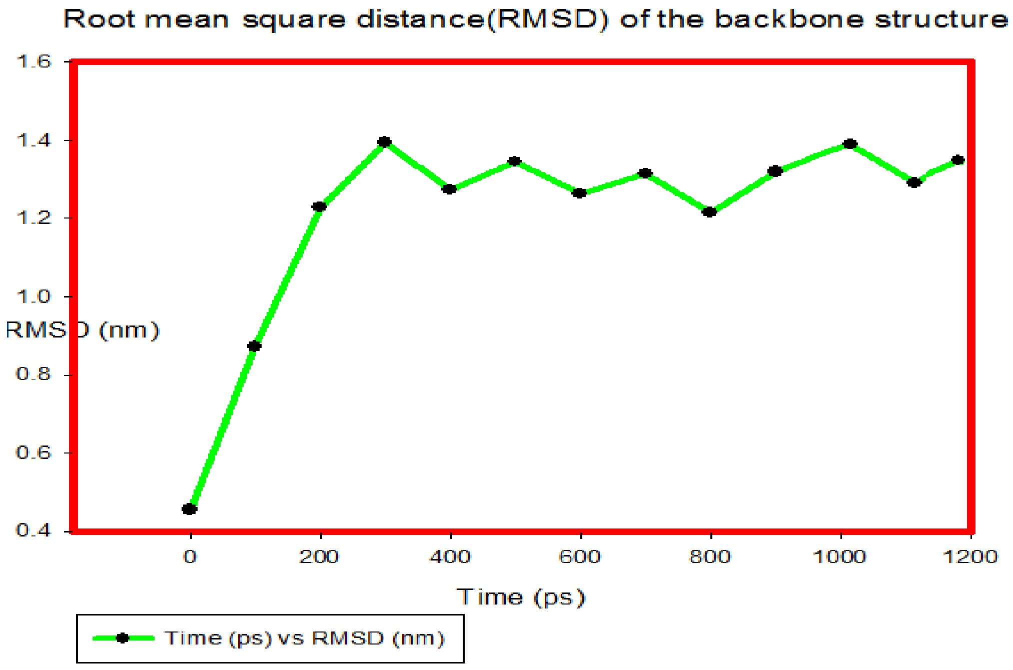


Figure 6: Root mean square distance (RMSD) of the backbone of the structure simulated over 2.5 nanoseconds.

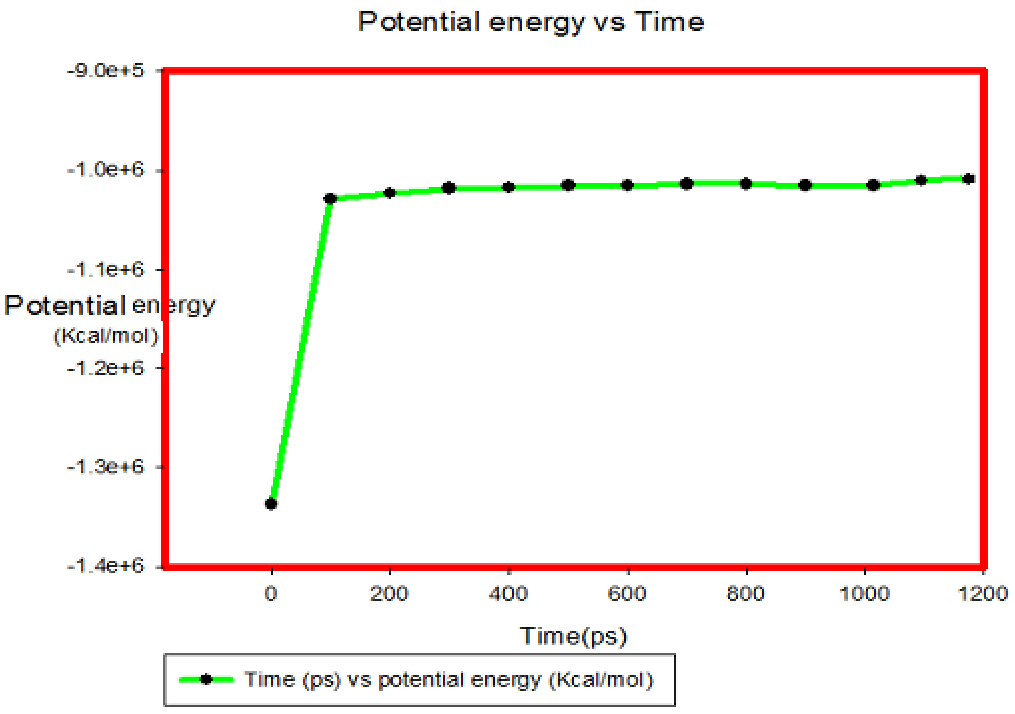


Figure 7: Shows Time (1200ps) Vs Potential Energy which indicated that very small fluctuation observed

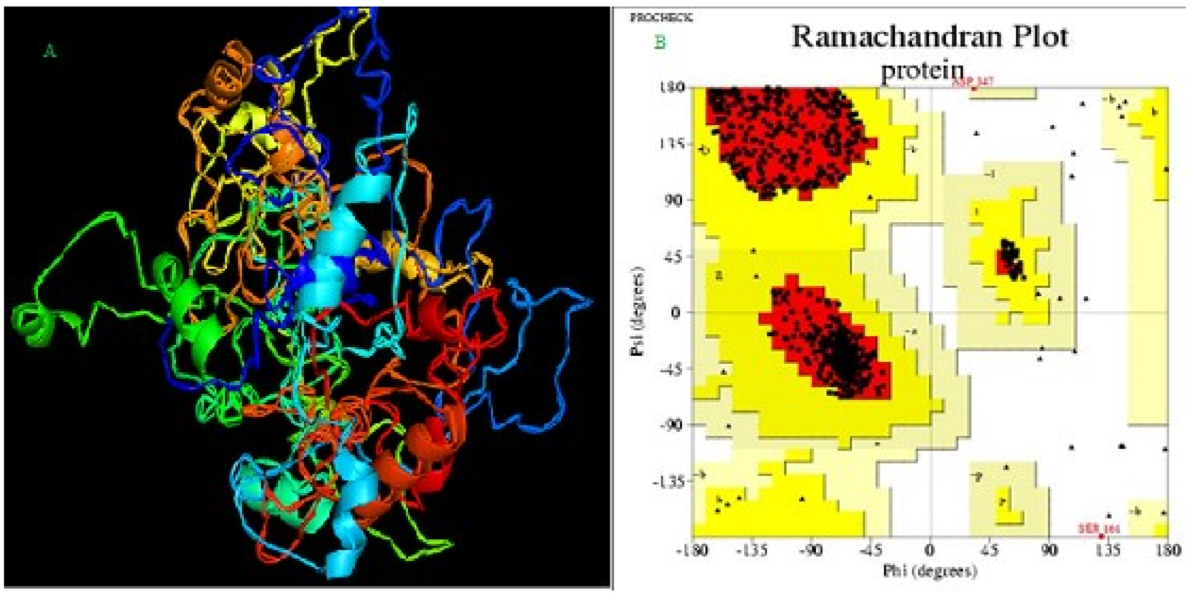


Figure 8: Protein structure validation: (A) Modeled structure of the SIRT1obtained from Easy Modeler. The structure is shown in secondary structure mode using Pymol. **(B)** Ramachandran plot for the modeled SIRT1. The most favored regions are colored red; additional allowed, generously allowed and disallowed regions are shown as yellow, light yellow and white fields, respectively.

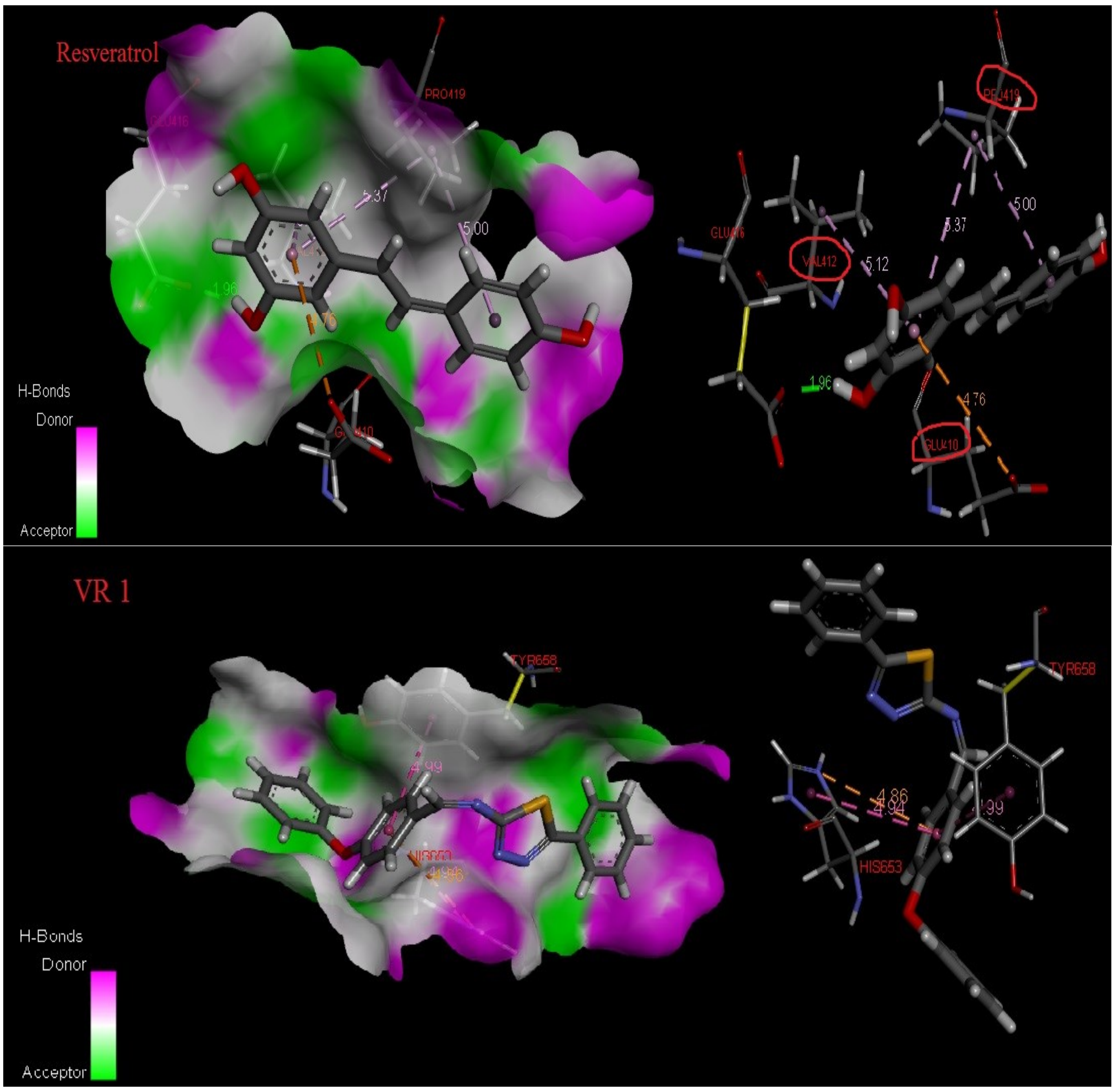
Table 2: Ramachandran plot statistics for the 3D model of SIRT1, calculated using PROCHECK

Parameter	**Value**
Core %	92.3
Allowed %	7.4
Disallowed %	0.2
General %	0.2

.

Table 3: Binding affinities of standard and designed compounds

Ligand	**Receptor**	**Affinity(Kcal/Mol)**	**Amino acids involved in interactions**	**H- bonds**	**Pi bonds**
Resveratrol	Sirtuin type1 (SIRT1)	-6.4	GLN A 361, GLY A 364, SER A 365, ALA A 367, LYS A 408, GLU A 410, ILE A 411, VAL A 412, PHE A 413, GLU A 416, ASN A 417, LEU A 418, PRO A 419, GLN A 421	1	4
VR1	Sirtuin type1 (SIRT1)	-7.0	ARG A 466, ASP A 481, GLN B 641, TYR B 642, LEU B 643, ILE B 651, PHE B 652, HIS B 653, GLY B 654, ALA B 655, GLU B 656,TYR B 658, SER B 659	0	3
VR2	Sirtuin type1 (SIRT1)	-7.3	ILE A 360, GLN A 361, GLY A 364, LYS A 408, GLU A 410, ILE A 411, VAL A 412, GLU A 416, ASN A 417, PRO A 419, GLN A 421	0	7
VR3	Sirtuin type1 (SIRT1)	-7.7	ILE A 359, ILE A 360, GLN A 361, GLY A 364, LYS A 408, GLU A 410, ILE A 411, VAL A 412, GLU A 416, ASN A 417, PRO A 419, GLN A 421, PHE A 422	0	7


Resveratrol
PRO419
GLU416
5.37
5.00
1.96
4.76
GLU410
H-Bonds
Donor
Acceptor
PRO419
5.00
5.37
GLU416
VAL412
5.12
1.96
4.76
GLU410
VR 1
TYR658
4.99
HIS653
4.96
H-Bonds
Donor
Acceptor
TYR658
4.86
4.94
4.99
HIS653

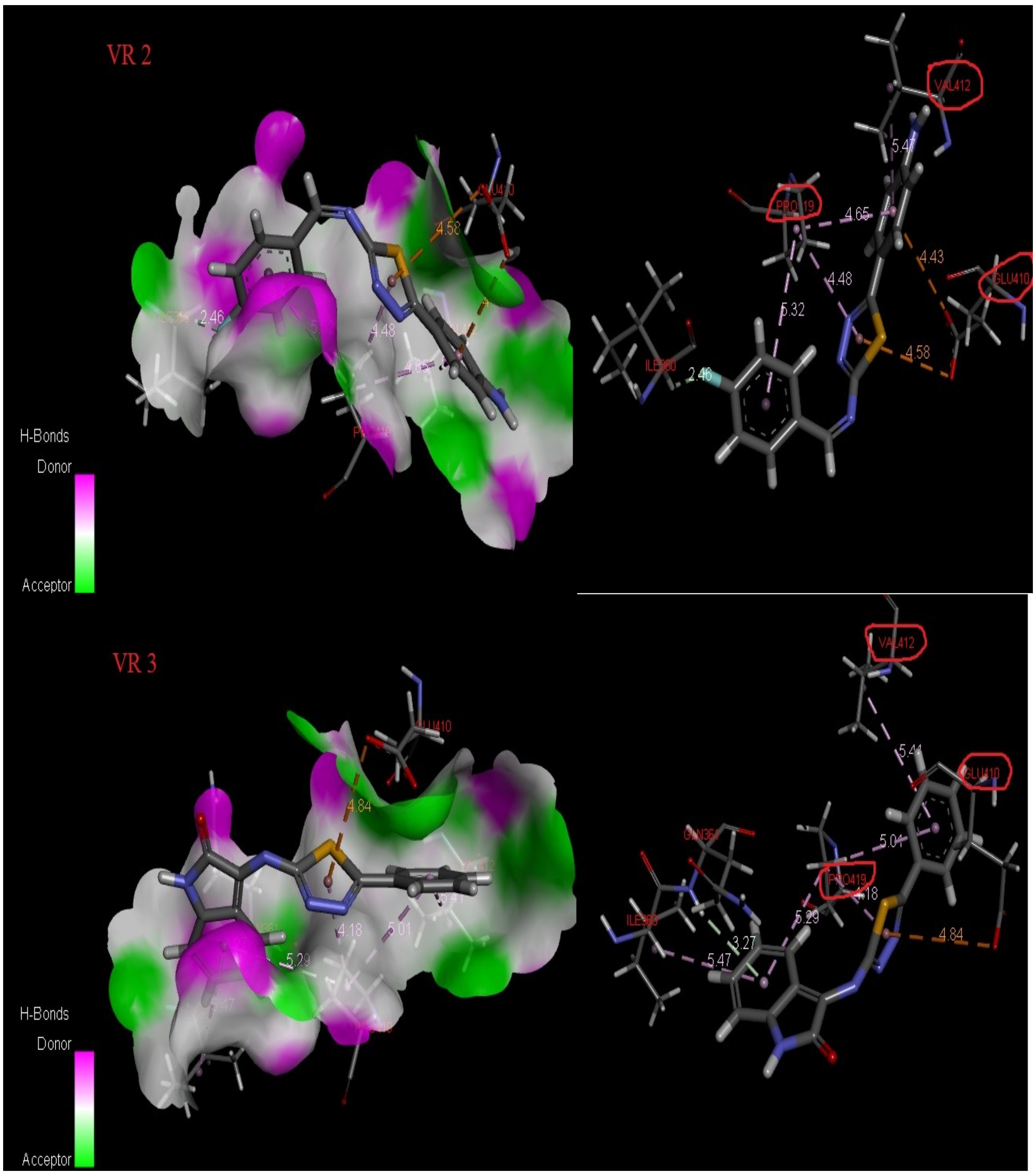


Figure 9: Docking images (a) Resveratrol, (b) VR 1, (c) VR 2 and (d) VR 3 with SIRT1, the green color dotted line shows hydrogen bonding and yellowish, light blue or whitish dotted line show Pi Donor, Acceptor and Alkyl bond respectively with amino acids involved in binding poses.

.

Table 4. The theoretical ADME properties of resveratrol and all designed 1,3,4-thiadiazole derivatives.

S. No.		**Rule**	**Resveratrol**	**VR 1**	**VR 2**	**VR 3**
1.	S+ log P	−2.0 to 6.5)	2.907	4.679	2.892	2.454
2.	S +log D	–	2.897	4.679	2.892	2.454
3.	M log P	–	2.402	3.987	2.917	2.585
4.	T_PSA	–	60.690	47.370	64.160	67.240
5.	n-OHNH donor	<5	3.000	0.000	2.000	1.000
6.	M_NO.	–	3.000	4.000	4.000	5.000
7.	Rule of 5	≤ 1	0.000	0.000	0.000	0.000
8.	%ABS (% of absorption)	_	88.07	92.66	86.87	85.81
9.	MV	–	206.922	329.474	265.301	253.064
10.	n-ON acceptor	<10	3	4	4	5
11.	n-ROTB	–	2	5	3	2
12.	M. Wt.	< 500	228.249	371.465	298.343	306.347

.

Table 5: BBB penetration of Resveratrol, VR1, VR2 and VR3

Parameter	**Resveratrol**	**VR1**	**VR2**	**VR3**
Rate of blood penetration Log PS	-1.6	-1.1	-1.1	-1.1
Extent of brain penetration Log PB	0.37	-0.15	-0.15	-0.15
Brain/plasma equilibration rate Log (PS*FU brain)	-2.5	-2.9	-3.0	-3.0

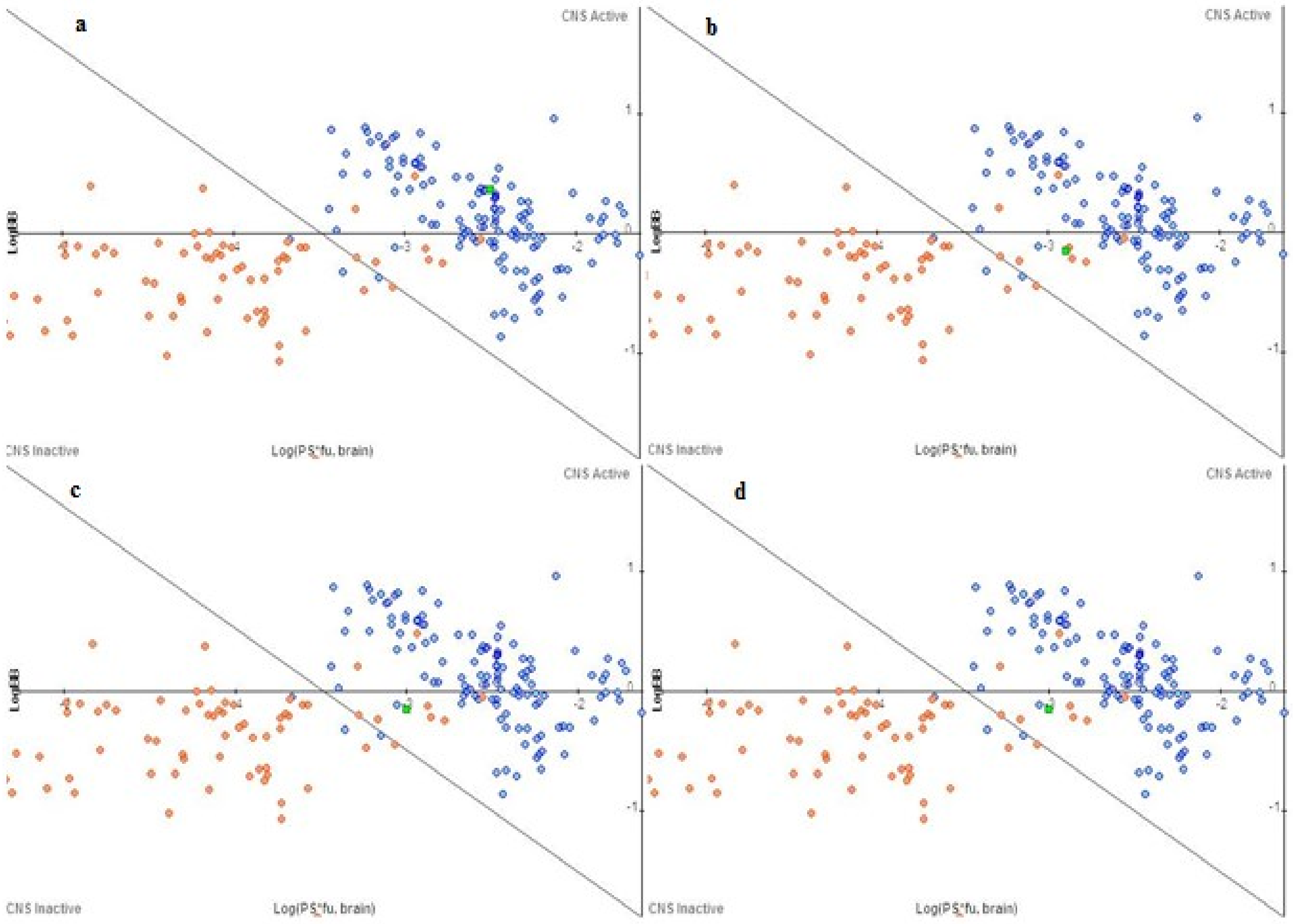


Figure 10: The BBB penetration power of the all active compounds (a, b, c, d, represents the following drug profile graph as resveratrol, VR1, VR2 and VR3 respectively). All graphs represent two region 1.CNS inactive 2.CNS active

Table 6: Score of bioactivity prediction of resveratrol and thiadiazole derivatives

S.No.	**Receptors**	**Resveratrol**	**VR 1**	**VR 2**	**VR 3**
1.	GPCR ligand	-0.20	-0.37	-0.46	-0.48
2.	Ion channel Modulator	0.02	-0.68	-0.79	-0.90
3.	Kinase inhibitor	-0.20	-0.11	-0.09	0.04
4.	Nuclear receptor ligand	0.01	-0.16	-0.39	-0.74
5.	Protease inhibitor	-0.42	-0.39	-0.55	-0.94
6.	Enzyme inhibitor	0.02	-0.24	-0.25	-0.32

Table 7: Topological comparative studies of resveratrol and thiadiazole derivatives

S.No.	DSSTox toxicity origin	Resveratrol	VR 1	VR 2	VR 3
1.	DSSTox Carcinogenic Potency DBS MultiCellCall: non-carcinogen	0.0127	0.0136	0.0123	0.00723
2.	DSSTox Carcinogenic Potency DBS Mutagenicity: non-mutagenic	0.162	0.101	0.00678	0.00723
3.	DSSTox Carcinogenic Potency DBS Rat: non-carcinogen	0.0517	0.0614	0.0417	0.0495
4.	Kazius-Bursi Salmonella mutagenicity: non-mutagenic	0.089	0.0335	0.0534	0.0419
5.	FDA v3b Maximum Recommended Daily Dose mmol: 0.0152722115276765	0.136	0.106	0.0834	0.0884
6.	DSSTox Carcinogenic Potency DBS SingleCellCall: non-carcinogen	0.011	0.0463	0.0126	0.0131
7.	EPA v4b Fathead Minnow Acute Toxicity LC50_mmol: 0.00359162218026281	0.207	0.184	0.203	0.19
8.	DSSTox ISSCAN v3a Canc: carcinogen	0.121	0.0869	0.243	0.000
9.	DSSTox Carcinogenic Potency DBS Hamster: non-carcinogen	0.137	0.237	0.179	0.131
10.	DSSTox Carcinogenic Potency DBS Mouse: non-carcinogen	0.0661	0.0146	0.105	0.0692

Conclusion

Homology modeling approach was used in our study to developed 3D structure of SIRT1. According to exist literature and analysis of the results from the our research of the homology modeling, docking and computational study indicated that the designed novel 1,3,4-thiadiazole derivatives has a potent activation effect on SIRT1 receptor at potential antigen target as well as treatment of a number of life threading diseases. All compounds displayed significant binding affinity compared with resveratrol. Among of them compound VR 3 has shown significant efficacy as well as the complex stability in the MD simulation. The other parameters like toxicity, ADME, oral bioavailability and BBB penetration of all designed compounds showed similar trends. The docking study data strongly support the assumption that SIRT1 may be involved in antiaging activity of 1, 3, 4-thiadiazole derivatives. However, the interaction of compounds with the receptor has concluded that PRO 419, GLU410 and VAL 412 common essential amino acids those may be involved in enhancing the efficacy of SIRT1. Thus, all data compared with the Resveratrol drug supported our antiaging hypothesis as a SIRT1 agonist (Kelly, 2010) .Hence, this observation could be attributed that the compound VR3 among of them as potential antiaging with SIRT1 mimetic/facilitator mode of action.

However, further studies, like synthesis, *in-vivo* evaluation and mechanism of action of these compounds are necessary to support this hypothesis. These titled compounds emerged as a lead for SIRT1 drug screening for future.

Conflict of interest statement

We wish to confirm that there are no known conflicts of interest associated with this publication.

Acknowledgements

The authors would like to express their gratitude to Babasaheb Bhimrao Ambedkar University (a Central University) Lucknow for providing the software and research data facilities.

References

Alegaon, S. G., Alagawadi, K. R., Sonkusare, P. V., Chaudhary, S. M., Dadwe, D. H., & Shah, A. S. (2012). Novel imidazo [2, 1-b][1, 3, 4] thiadiazole carrying rhodanine-3-acetic acid as potential antitubercular agents. *Bioorganic & medicinal chemistry letters, 22*(5), 1917-1921.

Amat, R., Planavila, A., Chen, S. L., Iglesias, R., Giralt, M., & Villarroya, F. (2009). SIRT1 controls the transcription of the peroxisome proliferator-activated receptor-γ co-activator-1α (PGC-1α) gene in skeletal muscle through the PGC-1α autoregulatory loop and interaction with MyoD. *Journal of Biological Chemistry, 284*(33), 21872-21880.

Bratislava.). Molinspiration Cheminformatics, from http://www.molinspiration.com/services/properties.html

Chou, J.-Y., Lai, S.-Y., Pan, S.-L., Jow, G.-M., Chern, J.-W., & Guh, J.-H. (2003). Investigation of anticancer mechanism of thiadiazole-based compound in human non-small cell lung cancer A549 cells. *Biochemical pharmacology, 66*(1), 115-124.

Clerici, F., Pocar, D., Guido, M., Loche, A., Perlini, V., & Brufani, M. (2001). Synthesis of 2-amino-5-sulfanyl-1, 3, 4-thiadiazole derivatives and evaluation of their antidepressant and anxiolytic activity. *Journal of medicinal chemistry, 44*(6), 931-936.

Demirbas, N., Karaoglu, S. A., Demirbas, A., & Sancak, K. (2004). Synthesis and antimicrobial activities of some new 1-(5-phenylamino-[1, 3, 4] thiadiazol-2-yl) methyl-5-oxo-[1, 2, 4] triazole and 1-(4-phenyl-5-thioxo-[1, 2, 4] triazol-3-yl) methyl-5-oxo-[1, 2, 4] triazole derivatives. *European journal of medicinal chemistry, 39*(9), 793-804.

Frye, R. A. (1999). Characterization of five human cDNAs with homology to the yeast SIR2 gene: Sir2-like proteins (sirtuins) metabolize NAD and may have protein ADP-ribosyltransferase activity. *Biochemical and biophysical research communications, 260*(1), 273-279.

Karegoudar, P., Prasad, D. J., Ashok, M., Mahalinga, M., Poojary, B., & Holla, B. S. (2008). Synthesis, antimicrobial and anti-inflammatory activities of some 1, 2, 4-triazolo [3, 4-b][1, 3, 4] thiadiazoles and 1, 2, 4-triazolo [3, 4-b][1, 3, 4] thiadiazines bearing trichlorophenyl moiety. *European journal of medicinal chemistry, 43*(4), 808-815.

Kelly, G. (2010). A review of the sirtuin system, its clinical implications, and the potential role of dietary activators like resveratrol: part 1. *Altern Med Rev, 15*(3), 245-263.

López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The Hallmarks of Aging. *Cell, 153*(6), 1194-1217. doi: http://dx.doi.org/10.1016/j.cell.2013.05.039

Martinez, A., Alonso, D., Castro, A., Aran, V. J., Cardelus, I., Banos, J. E., & Badia, A. (1999). Synthesis and potential muscarinic receptor binding and antioxidant properties of 3-(thiadiazolyl) pyridine 1-oxide compounds. *Archiv der Pharmazie, 332*(6), 191-194.

Oruç, E. E., Rollas, S., Kandemirli, F., Shvets, N., & Dimoglo, A. S. (2004). 1, 3, 4-Thiadiazole derivatives. Synthesis, structure elucidation, and structure-antituberculosis activity relationship investigation. *Journal of medicinal chemistry, 47*(27), 6760-6767.

Pallàs, M., Casadesús, G., Smith, M. A., Coto-Montes, A., Pelegri, C., Vilaplana, J., & Camins, A. (2009). Resveratrol and neurodegenerative diseases: activation of SIRT1 as the potential pathway towards neuroprotection. *Current neurovascular research, 6*(1), 70-81.

Shetty, A. K. (2011). Promise of resveratrol for easing status epilepticus and epilepsy. *Pharmacology & therapeutics, 131*(3), 269-286.

Wareski, P., Vaarmann, A., Choubey, V., Safiulina, D., Liiv, J., Kuum, M., & Kaasik, A. (2009). PGC-1α and PGC-1β regulate mitochondrial density in neurons. *Journal of Biological Chemistry, 284*(32), 21379-21385.

Yusuf, M., Khan, R. A., Khan, M., & Ahmed, B. (2013). An Interactive Human Carbonic Anhydrase-II (hCA-II) Receptor – Pharmacophore Molecular Model & Anti-Convulsant Activity of the Designed and Synthesized 5-Amino-1,3,4-Thiadiazole-2-Thiol Conjugated Imine Derivatives. *Chemical Biology & Drug Design, 81*(5), 666-673. doi: 10.1111/cbdd.12113

Zhao, Y. H., Abraham, M. H., Le, J., Hersey, A., Luscombe, C. N., Beck, G., . . . Cooper, I. (2002). Rate-limited steps of human oral absorption and QSAR studies. *Pharmaceutical research, 19*(10), 1446-1457.

SciForum
Mol2Net

Law, Software, & Cheminformatics: Copyright, Taxes, and Legal Issues

Aliuska Duardo-Sanchez*, and Antonio López-Díaz

Department of Especial Public Law, Financial and Tributary Law Area, Faculty of Law, University of Santiago de Compostela, 15782, Spain; aliuska.duardo@usc.es

Published: 4 December 2015

Abstract: In this communication we summarized our previous in-depth review (Current Topics in Medicinal Chemistry, 2008, Vol. 8, No. 18.) on the legal aspects of the use of software and computational models in Chemoinformatics. An overview of relevant international tax issues on the use of software is also presented.

Keywords: Legal issues; software; computational models; Bioinformatics; Copyright; Taxes; Regulatory uses; QSAR

Original Source: Duardo- Sanchez, A.; Patlewicz, G, López-Díaz, A., Current Topics on Software Use in Medicinal Chemistry: Intellectual Property, Taxes, and Regul,tory, Current Topics in Medicinal Chemistry, 2008, Vol. 8, No. 18.

1. Introduction

The uses of Bioinformatics software is strongly related, but not limited, to the development of Quantitative Structure-Activity Relationship (QSAR) models [3]. These QSAR models connect the structure of compounds of low molecular weight, nucleic acids (DNA and RNA), and proteins with the biological function. The use of this type of software allows researchers designing and/or predicting new promising compounds [5].

The relationship between Bioinformatics and the law spans a wide range of different issues including intellectual property, licensing legislation, regulation, product development as well as corporate legal issues [9]. In our previous work, we described the various legal procedures that are available to protect software, the acceptance and legal treatment of scientific results and techniques derived from such software, as well as some of the specific tax issues from the computer programs field [9a].

2. Results and Discussion

LEGAL PROTECTION OF SOFTWARE

There is no single method of legally protecting software. In fact, no unique international regime exists to address software protection [10].

COPYRIGHT

The most significant international treaties relating to copyright protection are the Berne Convention, the Universal Copyright Convention [14] and certain provisions of the TRIPS (Trade-Related Aspects of Intellectual Property Rights) agreement [15]. Software is protected by copyright throughout the E.U under the Community Directive on Software Copyright (91/250/EEC Directive). On a practical level, software vendors use multiple levels of Protection: trade secret rights, publishing "object code", binding users by contract, and increasingly- seeking patent protection [13].

PATENT PROTECTION

A patent is an exclusive right granted for an invention, which is a product or a process that provides a new way of doing something, or offers a new technical solution to a problem. So, patents require an inventive step, an assessment of industrial applicability and should undergo an examination procedure [18]. The law relating to the patentability of software is still not harmonized internationally [13]. The most widely followed doctrine governing the scope of patent protection for software related inventions is the "technical effects" doctrine, first promulgated by the European Patent Office (EPO) [10].

TRADE SECRET PROTECTION

“Given that access or non-access to the source code is such a key computing issue and that most proprietary software owners make great efforts to protect such code as confidential (non-disclosed) information, the relevance of trade secret law is immediately obvious” [19].

TRADEMARK PROTECTION

A Trademark is understood to be a brand or medium that is capable of being represented graphically, and which is capable of distinguishing goods or services of one undertaking from those of other undertaking. It may consist of words, designs, letters, numerals or the shape of goods or packaging [20]. The inherent limitations of the territorial application of trademark laws have been mitigated by various intellectual property treaties, foremost amongst which is the TRIPS. It establishes legal compatibility between member jurisdictions by requiring harmonization of applicable laws. International trademark issues are also governed by the European Community Trademark (ECT), the Madrid Agreement and the Madrid Protocol [21, 22].

TAX ISSUES FOR SOFTWARE

Tax systems depend of each internal country regulation, and for that reason taxation of cross border transactions relating to computer software, has always been a matter of debate. The payment for the use of computer programs is classified as a `royalty´. According to the OECD Model Tax Convention on Income and on Capital (Article 12.2), the term "royalties" means payments of any kind received as a consideration for the use of, or the right to use, any copyright of literary, artistic or scientific work including cinematograph films, any patent, trade mark, design or model, plan, secret formula or process, or for information concerning industrial, commercial or scientific experience [20, 26]].

REGULATORY USES OF QSAR

The challenge for bioinformatics and computational biology is in the validation and acceptance of new scientific methods and the results derived from software. The impact of the EU regulatory framework for QSAR provides some experience on how to address the issue of validation and acceptance of new approaches. Under Registration, Evaluation, Authorization and Restriction of Chemicals (REACH), all chemicals produced or imported in quantities of more than a one tons per annum (tpa) in the European Union, need to be assessed for human and environmental hazards. [29]. The REACH text refers to the need to demonstrate the validity of the QSAR used [33]. It is anticipated that validity will make reference to the internationally agreed OECD principles for QSAR validation already described [29, 31].

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

[3] Hansch, C.; Hoekman, D.; Leo, A.; Weininger, D.; Selassie, C.D., Chem. Rev, 2002, 102, (3) 783-812.

[5] González-Díaz, H.; Vilar, S.; Santana, L.; Uriarte, E., Curr. Top. Med. Chem. 2007, 7, (10)1025-1039.

[9] Kennedy, D. Bioinformatics Law Resources. http://www. denniskennedy.com/index.aspx (11/9/2007),

[9a] Duardo- Sanchez, A.; Patlewicz, G, López-Díaz, A., Current Topics on Software Use in Medicinal Chemistry: Intellectual Property, Taxes, and Regul,tory, Current Topics in Medicinal Chemistry, 2008, Vol. 8, No. 18.

[10] International Legal Protection for Software. http://www. softwareprotection.com/ (25/9/2007),

13] Steering Committee for Intellectual Property Issues in Software Computer Science and Telecommunications Board Commission on Physical Sciences, M., and Applications National Research Council. Intellectual Property Issues In software. National Academy Press: Washington, D.C, 1991.

[14] WIPO. Berne Convention for the Protection of Literary and Artistic Works of September 9, 1886, completed at PARIS on May 4, 1896, revised at BERLIN on November 13, 1908, completed at BERNE on March 20, 1914, revised at ROME on June 2, 1928, at BRUSSELS on June 26, 1948, at STOCKHOLM on July 14, 1967, and at PARIS on July 24, 1971, and amended on September 28, 1979. http://www.wipo.int/treaties/en/ip/berne/trtdocs_wo001.html

[15] WTO. Trade-Related Aspects of Intellectual Property Rights. Annex 1C of the Marrakech Agreement Establishing the World Trade Organization, signed in Marrakech, Morocco on 15 April 1994. http://www.wto.org/english/docs_e/legal_e/27-trips_01_e.htm

[18] Westkamp, G.N., IPR-Helpdesk Bulletin, 2005, (19).

[19] Story, A. In Intellectual Property Rights and Sustainable Develop-ment. ICTSD-UNCTAD, Ed.; Imprimerie Typhon: Chavanod, 2004, p 12.

[20] Morcon, C.; Roughton, A.; Gaham, J. The Modern Law of Trade Marks. Butterworth: London, 2005.

[21] WIPO. Madrid Agreement Concerning the International Registration of Marks of April 14, 1891, as revised at Brussels on December 14, 1900, at Washington on June 2, 1911, at The Hague on November 6, 1925, at London on June 2, 1934, at Nice on June 15, 1957, and at Stockholm on July 14, 1967,1 and as amended on September 28, 1979. http://www.wipo.int/madrid/en/legal_texts/trtdocs_wo015.html

[22] WIPO. Protocol Relating to the Madrid Agreement Concerning the International Registration of Marks adopted at Madrid on June 27, 1989 and amended on October 3, 2006. http://www.wipo.int/madrid/en/legal_texts/trtdocs_wo016.html

[23] Rowland, D.; Campbell, A., International Journal of Law and Information Technolog, 2002, 10, (1), 23-40(18).

[26] OECD. Organization for Economic Co-operation and Development. Articles of The Model Convention Whit Respect to Taxes on Income and on Capital. http://www.oecd.org/dataoecd/50/49/35363840.pdf (15/9/2007),

[29] OECD. Organisation for Economic Cooperation and Development. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships [(Q)SARs] Models. http://www.oecd.org/dataoecd/55/35/38130292.pdf (9/10/2007),

[31] OECD. Organisation for Economic Cooperation and Development. The Report from the Expert Group on (Quantitative) Structure-Activity Relationships [(Q)SARs] on the Principles for the validation of (Q)SARs. http://www.oecd.org/document/23/0,2340,en_2649_34365_33957015_1_1_1_1,00.html (9/10/2007),

[33] EC. In Official Journal of the European Union; Official Journal of the European Union, 2006; Vol. L396/1.

SciForum
Mol2Net

Multi-Target Prediction of Neuroprotective Drugs, Synthesis, Assay, and Theoretical Study of Rasagiline Carbamates

Francisco J Romero Durán[1], Nerea Alonso [1], Olga Caamaño [1], Xerardo García-Mera [1,*], Matilde Yañez [2]

[1] Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela (USC), 15782, Santiago de Compostela, Spain.
[2] Department of Pharmacology, USC, 15782, Santiago de Compostela, Spain.

Published: 4 December 2015

Abstract: In this work, we developed a multi-target model for neuroprotective compounds reported in the CHEMBL database. The model predicted correctly >8300 experimental outcomes with Accuracy, Specificity, and Sensitivity above 80-90% in training and external validation series. This is model can different outcomes for >30 experimental measures in >400 different experimental protocols and related to >150 molecular and cellular targets present in 11 different organisms (including human). After that, we reported by the first time, the synthesis, characterization, and experimental assays of new series of chiral 1,2-rasagiline carbamate derivatives; not reported in previous works. This work is a synopsis of the results presented in our previous paper: Int J Mol Sci. 2014 Sep 24;15(9):17035-64. doi: 10.3390/ijms150917035.

Keywords: CHEMBL; Neuroprotective agents; Rasagiline derivatives; Asymmetric synthesis

*** Corresponding authors:** X García-Mera (xerardo.garcía@usc.es)

1. Introduction

The discovery of new drugs for the treatment of neurodegenerative diseases such as Alzheimer's, Parkison's, and Huntington's diseases, Friedreich ataxia and others an important goals of medicinal chemistry [1-4]. In order to develop such computational models we need to use modeling techniques to process chemical information from public databases. These databases have accumulated immense datasets of experimental results of pharmacological trials for many compounds. For instance, CHEMBL[5, 6], https://www.ebi.ac.uk/chembldb, is one of the biggest with more than 11,420,000 activity data for >1,295,500 compounds, and 9,844 targets. In our previous work, we reported the first multi-target, multi-output, and multi-scale ALMA model for CHEMBL data of neuroprotective / neuro-toxic effect of drugs. After that, we reported by the first time, the synthesis, characterization, and experimental assays of new series rasagiline carbamate derivatives; not reported in previous works.

RESULTS AND DISCUSSION

1.1. Development of New Model for Prediction of Drug-Target Networks

Model training and validation. We report a model to predict when the ith compound may present a high ($L_{ij}(c_q)$ = 1) or not ($L_{ij}(c_q)$ = 0) value of the experimental parameter used to characterize interaction with a molecular or cellular target involved in neuroprotective/neurodegenerative process. The output $S_{ij}(c_q)$ of our multi-output model depend on both chemical structure of the ith drug d_i and the set of conditions selected to perform the biological assay (c_q) including the jth target, of course. In consonance, the ALMA model should predict different probabilities if we change the organisms (c_1), the biological assays (c_2), the molecular / cellular target (c_3), or the standard experimental parameter measured (c_4), for the same compound [7]. The best ALMA-entropy model found in this work was:

$$S_{ij}\left(c_q\right)=1.1396-0.4039\cdot p(c_l)?\theta_1^i+0.1993\cdot\Delta\theta_1^i\left(s_x\right)+0.4349\cdot\Delta\theta_1^i\left(a_u\right) \quad (4)$$
$$-0.0202\cdot\Delta\theta_1^i\left(o_t\right)-0.0017\cdot\Delta\theta_1^i\left(t_e\right)$$
$$N=2661 \quad R_c=0.72 \quad \chi^2=1913.007 \quad p<0.005$$

The statistical parameters for the above equation in training are: Number of cases used to train the model (N), Canonical Regression Coefficient (Rc), Chi-square (χ^2), and p-level [8]. The probability cut-off for this LDA model is $^ip_1(c_q)$ > 0.5 => $L_{ij}(c_q)$ = 1. It means that the drug d_i predicted by the model with probability > 0.5 are expected to give a positive outcome in the qth assays carry out under the given set of conditions c_q. This ALMA-entropy model present excellent performance in both training and external validation series with Sensitivity (Sn), Specificity (Sp), and Accuracy (Ac) > 80%. Values higher than 75% are acceptable for LDA-QSAR models, according to previous reports [9-13].

1.2. Experimental and Theoretical Study of New Compounds

Synthesis and experimental assay of new 1,2-rasagiline derivatives. The compounds **2**, **3**, **4**, **5**, **6**, **7**, **8** and **9** were synthesized according to the strategy given in **Figure 2**. As shown in this scheme, they were synthesized from the aminoalcohol **1** [(1*R*,2*S*)-(+)-1-amino-2-indanol], a commercial product. The alkylation of **1** with propargyl bromide and potassium carbonate in hot acetonitrile provided, in a global yield of 92%, a mixture of the corresponding mono- and dipropargylated derivatives (**2** and **3**), which were separated by flash column chromatography using hexane/EtOAc (3:1) as eluent. Compound **3** was converted to the corresponding acetate (**4**) and benzoate (**5**) by treatment with acetic anhydride or benzoyl chloride, Et_3N and catalytic amounts of DMAP in MeCN. The carbamate derivatives (**6**, **7**, **8** and **9)** were synthesized, from the hidroxy mono- or dipropargylaminoindans (**2** and **3**), by reaction with the corresponding dialkylcarbamyl chloride in NaH and acetonitrile following the procedure described in the literature [14].

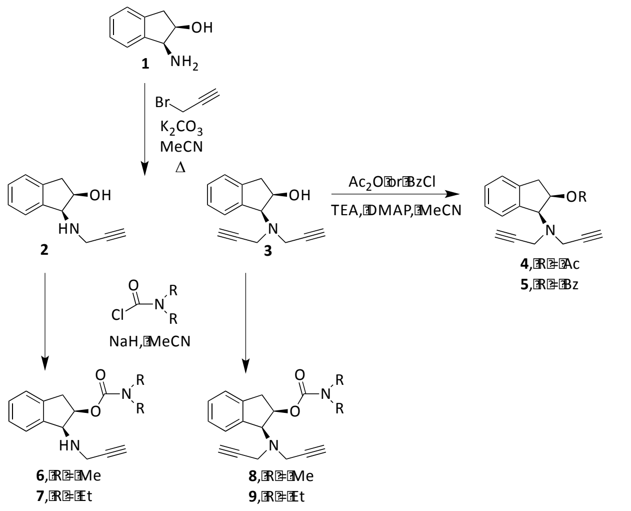


Figure 1. Synthesis of compounds **2-9**

The new compounds synthesized in this work (**2**, **3**, **4**, **5**, **6**, **7**, **8**, and **9**) were subjected to an initial study to determinate its neuroprotective ability in both the presence and the absence of neurotoxic agents (ANA). The method of reduction of the 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) was used to ascertain the cell viability, given by the number of cells present in the culture. The ability of cells to reduce MTT is an indicator of the integrity of mitochondria, and its functional activity is interpreted as a measure of cell viability [15]. Three assays were conducted in a culture of motor cortex neurons of 19-day-old Sprague-Dawley rat embryos. Firstly, we studied the ability to induce a neuroprotective effect in the absence of any neurotoxic stimulation. Secondly, we studied the neuroprotective effect in the presence of glutamate, a compound that causes a pathological process in which neurons are damaged leading to apoptosis when its receptors such as the NMDA and AMPA are over-activated. Lastly, the ability of the compounds synthesized to protect from damage by H_2O_2, that causes neuronal death by Oxidative stress, was analyzed. The results obtained allow to deduce the existence of a moderate neuroprotective effect in the absence of any toxic stimulus, presenting the best results type **6** and **9** carbamate derivatives, with values of 11.5% and 8.4% respectively, followed by the compound **3**, **4**, and **7** with values slightly above 4% (see **Figure 2**).

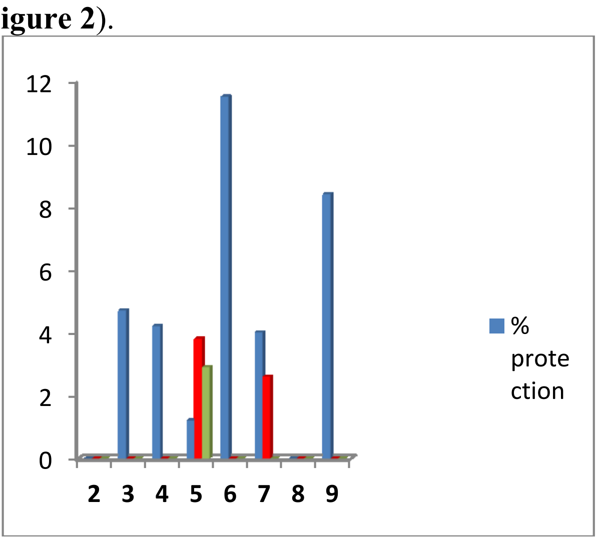


Figure 2. Results of the experimental assay of Neuroprotective effect of the new compounds.

Predict new drugs in other assays. We used the ALMA-entropy model to predict the more probable results for all the new rasagiline derivatives, synthesized in this work, in >500 assays not carried out experimentally. When the molecular descriptors (entropy indices) of the new rasagiline derivatives were introduced in our model we obtained the probable interaction with different targets. The model predicts that most of them could interact with the subunits A and B of the 5-hidroxy-tryptamine type 3 receptors (5-HT3Rs). These results seem to be consistent with the literature, since the antagonists of 5-HT3Rs have been related to neuroprotective properties *in vitro* and *in vivo* [16].

Acknowledgements

The authors thank the Xunta de Galicia for financial support of this work under project 07CSA008203PR.

Conflict of Interest

The authors declare no conflict of interest

References

1. Allegri RF, Guekht A: **Cerebrolysin improves symptoms and delays progression in patients with Alzheimer's disease and vascular dementia**. In: *Drugs Today (Barc).* vol. 48 Suppl A. United States: 2012 Prous Science, S.A.U. or its licensors; 2012: 25-41.
2. Park NH: **Parkinson disease**. *JAAPA* 2012, **25**(5):73-74.
3. Morris HR, Waite AJ, Williams NM, Neal JW, Blake DJ: **Recent advances in the genetics of the ALS-FTLD complex**. *Curr Neurol Neurosci Rep* 2012, **12**(3):243-250.
4. Trushina E, McMurray CT: **Oxidative stress and mitochondrial dysfunction in neurodegenerative diseases**. In: *Neuroscience.* vol. 145. United States; 2007: 1233-1248.
5. Heikamp K, Bajorath J: **Large-scale similarity search profiling of ChEMBL compound data sets**. *J Chem Inf Model* 2011, **51**(8):1831-1839.
6. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B *et al*: **ChEMBL: a large-scale bioactivity database for drug discovery**. *Nucleic Acids Res* 2012, **40**(Database issue):D1100-1107.
7. Gerets HH, Dhalluin S, Atienzar FA: **Multiplexing cell viability assays**. *Methods Mol Biol* 2011, **740**:91-101.
8. Hill T, Lewicki P: **STATISTICS Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining**, vol. 1. Tulsa: StatSoft; 2006
9. Patankar SJ, Jurs PC: **Classification of inhibitors of protein tyrosine phosphatase 1B using molecular structure based descriptors**. *J Chem Inf Comput Sci* 2003, **43**(3):885-899.
10. Garcia-Garcia A, Galvez J, de Julian-Ortiz JV, Garcia-Domenech R, Munoz C, Guna R, Borras R: **New agents active against Mycobacterium avium complex selected by molecular topology: a virtual screening method**. *J Antimicrob Chemother* 2004, **53**(1):65-73.
11. Marrero-Ponce Y, Castillo-Garit JA, Olazabal E, Serrano HS, Morales A, Castanedo N, Ibarra-Velarde F, Huesca-Guillen A, Sanchez AM, Torrens F *et al*: **Atom, atom-type and total molecular linear indices as a promising approach for bioorganic and medicinal chemistry: theoretical and experimental assessment of a novel method for virtual screening and rational design of new lead anthelmintic**. *Bioorg Med Chem* 2005, **13**(4):1005-1020.
12. Casanola-Martin GM, Marrero-Ponce Y, Khan MT, Ather A, Sultan S, Torrens F, Rotondo R: **TOMOCOMD-CARDD descriptors-based virtual screening of tyrosinase inhibitors: evaluation of different classification model combinations using bond-based linear indices**. *Bioorg Med Chem* 2007, **15**(3):1483-1503.
13. Casanola-Martin GM, Marrero-Ponce Y, Khan MT, Khan SB, Torrens F, Perez-Jimenez F, Rescigno A, Abad C: **Bond-based 2D quadratic fingerprints in QSAR studies: virtual and in vitro tyrosinase inhibitory activity elucidation**. *Chemical biology & drug design* 2010, **76**(6):538-545.

14. Sterling J, Herzig Y, Goren T, Finkelstein N, Lerner D, Goldenberg W, Miskolczi I, Molnar S, Rantal F, Tamas T *et al*: **Novel dual inhibitors of AChE and MAO derived from hydroxy aminoindan and phenethylamine as potential treatment for Alzheimer's disease**. In: *J Med Chem.* vol. 45. United States; 2002: 5260-5279.

15. Mosmann T: **Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays**. In: *J Immunol Methods.* vol. 65. Netherlands; 1983: 55-63.

16. Fakhfouri G, Rahimian R, Ghia JE, Khan WI, Dehpour AR: **Impact of 5-HT(3) receptor antagonists on peripheral and central diseases**. *Drug Discov Today* 2012, **17**(13-14):741-747.

SciForum
Mol2Net

3D Hierarchically Scaffolds for Bone Repair: At the Crossroads of Experimental and Computational Outlooks

Paula Messina [1] and Juan M. Ruso [2,*]

[1] Department of Chemistry, INQUISUR-CONICET, Universidad Nacional del Sur, 8000 Bahía Blanca, Argentina; E-Mail: pmessina@uns.edu.ar

[2] Soft Matter and Molecular Biophysics Group, Department of Applied Physics, University of Santiago de Compostela, Santiago de Compostela E-15782, Spain.

* Author to whom correspondence should be addressed; E-Mail: juanm.ruso@usc.es
Tel.: +34- 881-811-000 (ext. 14042); Fax: +34-881-141-112.

Published: 4 December 2015

Abstract: A combination of experimental and theoretical approaches is proposed to determine the best template and conditions to synthesized a bio-active open mesopore structure material with a 3D-sponge-like network similar than those existed in trabecular bone with a consequent saving of time and products. Specifically, we discuss the possibility of combining experimental methods with theoretical methods that have been published in J Mater Sci 2012, 47, 2837-2844 and Langmuir 2015, doi: 10.1021/acs.langmuir.5b03074, respectively. The strategy proposed opens a new gate to the rational design of 3D hierarchically scaffolds for bone repair in the future.

Keywords: Self-Assembly; Scaffolds; QSPR; Biomaterials; Biomineralization; Bones; Biomaterials; Nanoparticles.

1. Introduction

Nanomedicine goes beyond the use of classical medicine at cell scales, currently faces an exceptional opportunity to improve global healthcare[1,2]. One of the main goals of nanomedicine is to develop new biocompatible scaffolds[3]. These materials in order to perform effectively must have an appreciation of the complex interrelationships between humans and the physical/chemical components of the environment [4].

Around 45% of the population aged 50 years and older will have an osteoporosis-related fracture in their lifetime. Such statistics demonstrate the vast need for new developments in this area. Bone is a complex composite material, with living and nonliving matter. The stiff nature of bone clearly enables it to form a semirigid framework, enable motion, and provide organ protection. Because of the trabecular bone exhibit spongelike bicontinuity at the millimeter scale, there is a growing interest in the synthesis of spongy-like materials[5]. However, there are too many factors to consider: immune response, vascularization, chemotactic and so on [6]. To overcome these challenges particular interest must be paid in the molecular sieve [7]. Here we propose a combination of experimental and computational tools to characterize and optimize the properties of sieves based on bicontinuous pore silica materials templated with bile salts mixtures.

2. Results and Discussion

Self-aggregation of binary systems has received considerable attention recently due to their potential as sustainable templates. In different works, we have studied, both experimentally and with molecular modeling, the optimization of the conditions and routes for the development of the best bio-active implantable materials. Such studies allows for the direct assessment of the role that molecular architecture and experimental conditions plays in approaches to simulate the trabecular bone organization

2.1. Computational models

On one side, in a previous work published this year we have built the first model that combines perturbation theories (PT) and linear free energy relationships (LFER) ideas [8]. The model uses as input covariance PT operators (CPTOs). CPTOs are calculated as the difference in covariance ΔCov($^1\mu_k$) functions before and after multiple perturbations in the binary system[9]. In turn, we can calculate the covariances calculated as the product of two Box−Jenkins operators (BJO) operators [10]. BJOs are used to measure the deviation of the structure of different chemical compounds from a set of molecules measured in a given subset of experimental conditions (b_j). The best CPT-LFER model found predicted the effects of 25000 perturbations over 9 different properties of binary systems. The best CPT-LFER model found with this algorithm was the following:

$$
{}^{0}f\left({}^{p}\varepsilon_{ij}\right)_{new} = -0.275152\,{}^{1}f\left({}^{p}\varepsilon_{ij}\right)_{ref} - \\
-0.158186\Gamma_{pi}(dip)^{0} + 0.037112\Gamma_{pi}(solv)^{0} + \\
+0.017595\Gamma_{pi}(part)^{0} - 0.150110\Gamma_{pi}(solv)^{-} + \\
+0.095564\Gamma_{pi}(solv)^{+} + 0.181357
$$

where the output function ${}^{0}f\left({}^{p}\varepsilon_{ij}\right)_{new}$ is a multi-output function that quantifies the numerical values (ε) of different *p*th physicochemical properties of the *i*th binary system that have been experimentally determined under a certain set of *j*th boundary conditions (c_j). $\Gamma_{jp}(k)^{q}$ are the covariance perturbation functions. The types of potentials were the electrostatic dipole potential (*dip*), electrostatic potential in solution (*solv*) and thermodynamic potential for water−nonpolar phase partitioning (*part*). The notation for q is 0 when both species are molecules, + and − when both species are cations and anions, respectively.

2.1. Experimental methods

On the other hand, in other recent work [11] we reported the experimental study of the physicochemical properties of the nanostructures. The nanostructures studied are formed within aqueous mixtures of bile salts sodium glycodeoxycholate (NaGDC), dehydrocholic acid (HDHC), Sodium deoxycholate (NaDC) and surfactant didodecyldime thylammonium bromide (DDAB) as a function of total concentration and mixed ratio. The experimental study has been carried out by means of thermodynamic analysis, fluorescence spectroscopy, field emission-scanning electron microscopy (FE-SEM) and energy dispersive X-ray microanalysis (EDX) experiments [11].

Experimental measurements were perfectly reproduced by our CPT-LFER model, which was further used to examine different systems under other conditions not studied experimentally with a consequent saving of time and products.

The siliceous materials (SM) were prepared using hydrothermal synthesis. Tetraethyl orthosilicate (TEOS) was added to different bile salts mixture solutions. The mixture was stirred and left for 24 h in an autoclave at 100 ºC. The obtained materials were calcined [12]. The material final structures take place through a vesicle to sponge-like phase transformation whose driving force seems to be the interaction of bile salts and silica species during the material polymerization synthesis step. Depending of the type and amount of BS in the template mixture, the film can curve toward the a-polar or toward the polar side leading to different final morphologies. The highly hydrophobic steroid backbone of NaDHC molecule causes a great disturbance in the templated liquid crystal mixture. The final mesophases obtained depending on the materials arises in different structures.

It was demonstrated that the requirement for an artificial material to bond to living bone (bioactivity) is the formation of bonelike apatite

(HA) on its surface when implanted in the living body [13],[14], and that this in vivo apatite formation can be reproduced in a SBF with ion concentrations nearly equal to those of human blood plasma.

The existence of a highly porous surface can accelerate the biomimetic process. In agreements, the apatite deposition was observed only in such specimens which content mesopores and high proportion of siloxane bridges in their structures. The time evolution of apatite growth on the synthesized materials was followed by FT-IR measurements (ESM) and confirmed by scanning electron microscopy. Figure 1 show the FE-SEM microphotographs of the material surfaces after soaking in 1.5 SBF. It must be noticed that calcium phosphate coatings grow not only on the material surface but also in the pore interior and the progress of this covering increases as a function of the soaking time SEM revealed the presence of preformed calcium phosphate coatings to be composed entirely of straight platelike units with sharp edges, with a change in crystal geometry as the time of soaking increased. The definite crystalline structure is achieved after soaking for 20 days in SBF. The thickness of the apatite-like coating increases with time and reaches a saturated point after 10 days of soaking. Assuming that the growth rate of apatite coating is controlled by the calcium and phosphorous ions diffusion rates from the SBF to the material surface, the growth kinetics of apatite-like coatings on porous materials can be expressed by an empirical relationship

$$d^2 = Kt$$

where d is the thickness of the coating evaluated from SEM photos, t is the soaking time for the biomimetic deposition, and K is the growth rate constant. The growth rate constant obtained was , K = 2.0208 9× 10^{-18} m^2 seg^{-1}.

.

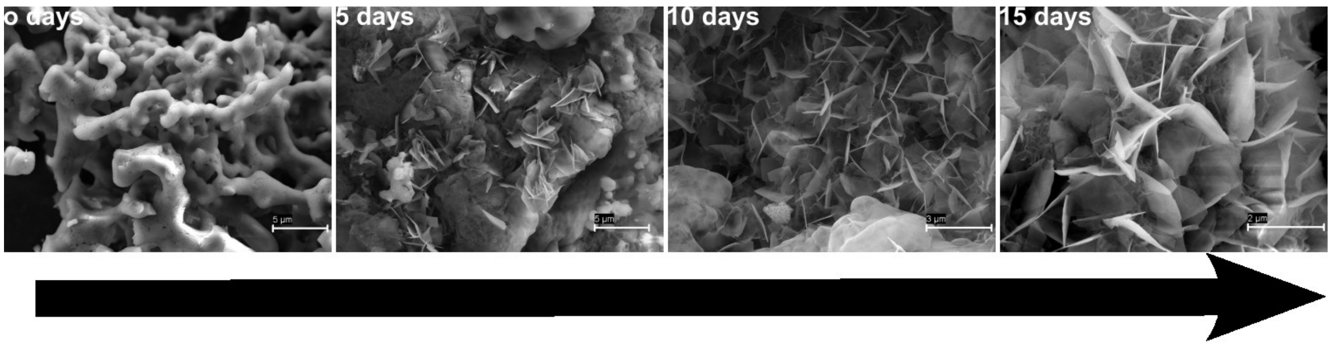


Figure 1. FE-SEM microphotographs showing the time evolution of HA layer formation on materials. Scale bars 3-5µm.

3. Materials and Methods

Dihidroxy bile salt sodium glycodeoxycholate (NaGDC), dehydrocholic acid (HDHC), sodium deoxycholate (NaDC), didodecyldimethyl ammonium bromide (DDAB) and tetraethyl orthosilicate (TEOS) were purchased from Sigma-Aldrich and used as received.

Fluorescence Spectroscopy. Fluorescence spectra of pyrene were obtained with a Cary Eclipse spectrophotometer equipped with a temperature-control device and a multicell sample holder (Varian 198 Instruments Inc.). All samples were prepared with saturated solutions of pyrene (3×10^{-7} mol dm^{-3}). Measurements were performed at 298 K, and the fluorescence intensities ratios (I1/I3) of the first (I1, 373 nm) and third (I3, 384 nm) peaks from the short wavelength in the spectra of pyrene were obtained with excitation at λ = 335 nm. The excitation and emission slit widths were set to be 5 and 1.5 nm, respectively.

Field emission-scanning electron microscopy (FE-SEM) and energy dispersive X-ray microanalysis (EDX) were performed using a

FE-SEM ULTRA PLUS microscope. Microanalysis EDX: resolution 129 eV and wavelength-dispersive (WD) 8.5 nm. Transmission electron microscopy (TEM) was performed using a Philips CM-12 transmission electron microscope equipped with a digital camera MEGA VIEWII DOCU.

Infrared spectra were collected using a Nicolet-Nexus 470 Fourier-Transform infrared Spectrometer (FT-IR) equipped with a pneumatic motion interferometer.

4. Conclusions

We successfully presented a new computational method for predicting key physicochemical properties of compounds for biomaterials templating. For this purpose we developed the first model which combines perturbation theory (PT) and linear free energy relationship (LFER) ideas. We show that a single model may have a good quality of predicting multiple properties. Results obtained by our model were compared with published data, and our experimental results obtained standard errors less than 0.02%.

We also demonstrated that controlling molecular architecture is an efficient platform to manipulate the material bioactivity. The presence of siloxane bridges and its distribution results in HA growth not only at the surface but also at pore interior. Finally, the kinetics of HA formation was also tested and we concluded that the definite crystalline structure is achieved after 20 days.

The detailed analysis and theoretical model provided in these studies is expected to be useful as a reference to design in a fast and economical way new bioactive scaffolds for bone regeneration based on surfactant mixtures.

Acknowledgments

Authors acknowledge Fundación Ramón Areces, Universidad Nacional del Sur (PGI 24/Q064), and Concejo Nacional de Investigaciones Cientificas y Técnicas de la República Argentina (CONICET, PIP-11220130100100CO) for their financial support.

Author Contributions

All authors have contributed equally to this work.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Schillmeier, M. Caring for social complexity in nanomedicine. *Nanomedicine* **2015**, *10*, 3181-3193.
2. Ruso, J.M.; Deo, N.; Somasundaran, P. Complexation between dodecyl sulfate surfactant and zein protein in solution. *Langmuir* **2004**, *20*, 8988-8991.
3. V. Messina, P.; Miguel Besada-Porto, J.; M. Ruso, J. Self-assembly drugs: From micelles to nanomedicine. *Current Topics in Medicinal Chemistry* **2014**, *14*, 555-571.
4. Fu, S.; Ni, P.; Wang, B.; Chu, B.; Peng, J.; Zheng, L.; Zhao, X.; Luo, F.; Wei, Y.; Qian, Z. In vivo biocompatibility and osteogenesis of electrospun poly(ε-caprolactone)–poly(ethylene glycol)–poly(ε-caprolactone)/nano-hydroxyapatite composite scaffold. *Biomaterials* **2012**, *33*, 8363-8371.
5. M. Ruso, J.; Sartuqui, J.; V. Messina, P. Multiscale inorganic hierarchically materials: Towards an improved orthopaedic regenerative medicine. *Current Topics in Medicinal Chemistry* **2015**, *15*, 2290-2305.
6. Sartuqui, J.; D' Elía, N.; Gravina, A.N.; Messina, P.V. Analyzing the hydrodynamic and crowding evolution of aqueous hydroxyapatite-gelatin networks: Digging deeper into bone scaffold design variables. *Biopolymers* **2015**, *103*, 393-405.

7. Ruso, J.M.; Pardo, V.; Sartuqui, J.; Gravina, N.; D'Elía, N.L.; Pieroni, O.I.; Messina, P.V. Photoluminescent sba-16 rhombic dodecahedral particles: Assembly, characterization, and ab initio modeling. *ACS Applied Materials & Interfaces* **2015**, *7*, 12740-12750.
8. Messina, P.V.; Besada-Porto, J.M.; González-Díaz, H.; Ruso, J.M. Self-assembled binary nanoscale systems: Multioutput model with lfer-covariance perturbation theory and an experimental–computational study of nagdc-ddab micelles. *Langmuir* **2015**.
9. Casañola-Martin, G.; Le-Thi-Thu, H.; Pérez-Giménez, F.; Marrero-Ponce, Y.; Merino-Sanjuán, M.; Abad, C.; González-Díaz, H. Multi-output model with box–jenkins operators of linear indices to predict multi-target inhibitors of ubiquitin–proteasome pathway. *Mol Divers* **2015**, *19*, 347-356.
10. Casañola-Martin, G.M.; Le-Thi-Thu, H.; Pérez-Giménez, F.; Ponce, Y.M.; Sanjuán, M.M.; Abad, C.; Díaz, H.G. Multi-output model with box-jenkins operators of quadratic indices for prediction of malaria and cancer inhibitors targeting ubiquitin-proteasome pathway (upp) proteins. *Curr Protein Pept Sci* **2015**.
11. Fernández-Leyes, M.; Verdinelli, V.; Hassan, N.; Ruso, J.; Pieroni, O.; Schulz, P.; Messina, P. Biomimetic formation of crystalline bone-like apatite layers on spongy materials templated by bile salts aggregates. *J Mater Sci* **2012**, *47*, 2837-2844.
12. Ruso, J.M.; Verdinelli, V.; Hassan, N.; Pieroni, O.; Messina, P.V. Enhancing cap biomimetic growth on tio2 cuboids nanoparticles via highly reactive facets. *Langmuir* **2013**, *29*, 2350-2358.
13. Hassan, N.; Verdinelli, V.; Ruso, J.M.; Messina, P.V. Mimicking natural fibrous structures of opals by means of a microemulsion-mediated hydrothermal method. *Langmuir* **2011**, *27*, 8905-8912.
14. Gravina, N.; Ruso, J.M.; Mbeh, D.A.; Yahia, L.H.; Merhi, Y.; Sartuqui, J.; Messina, P.V. Effect of ceria on the organization and bio-ability of anatase fullerene-like crystals. *RSC Advances* **2015**, *5*, 8077-8087.

SciForum
Mol2Net

HPLC-qTOF-MS Platform as Valuable Tool for the Exploratory Characterization of Phenolic Compounds in Guava Leaves at Different Oxidation States

Elixabet Díaz-de-Cerio [1,*], Vito Verardo [2], Ana María Gómez-Caravaca [1], Alberto Fernández-Gutiérrez [1] and Antonio Segura-Carretero [1]

[1] Department of Analytical Chemistry, Faculty of Sciences, University of Granada, Avd. Fuentenueva s/n, 18071, Granada, Spain; and Functional Food Research and Development Center, Health Science Technological Park, Avd. del Conocimiento, Bioregion building, 18100, Granada, Spain. E-Mail: anagomez@ugr.es (A.M.G.C.); albertof@ugr.es (A.F.G.); ansegura@ugr.es (A.S.C.)

[2] Department of Chemistry and Physics (Analytical Chemistry Area) and Research Centre for Agricultural and Food Biotechnology (BITAL), Agrifood Campus of International Excellence, ceiA3, University of Almería, Carretera de Sacramento s/n, E-04120 Almería, Spain.; E-Mails: vito.verardo@cidaf.es (V.V.);

* Author to whom correspondence should be addressed; E-Mail: ediazdecerio002@correo.ugr.es; Tel.: +34 958 637 206; Fax: +34- 958 637 083.

Published: 4 December 2015

Abstract: *Psidium guajava* L. is widely used like food and in folk medicine all around the world. Many studies have demonstrated that guava leaves have anti-hyperglycaemic and anti-hyperlipidemic activities, among others. The biological activity of guava leaves belongs mainly to phenolic compounds. Andalusia is one of the regions in Europe where guava is grown, thus, the aim of this work was to study the phenolic compounds present in Andalusian guava leaves at different oxidation state (low, medium and high). The phenolic compounds in guava leaves were determined by HPLC-DAD-ESI-qTOF-MS. We identified seventy-two phenolic compounds and, to our knowledge, twelve of them were determined for the first time in guava leaves in negative ionization mode. Moreover, positive ionization mode allowed the identification of the cyanidin-glucoside. To our knowledge this compound has been identified for the first time in guava leaves.

The results obtained by chromatographic analysis reported that guava leaves with low degree of oxidation have a higher content gallic and ellagic derivatives and flavonols compared to the other two guava leaf samples. Contrary, high oxidation state guava leaves reported the highest content of cyanidin-glucoside that was 2.6 and 15 times higher than guava leaves with medium and low oxidation state, respectively.

The qTOF platform permitted the determination of several phenolic compounds and provided new information about guava leaf phenolic composition that could be useful for nutraceutical production.

Keywords: Psidium guajava L.; HPLC-DAD-ESI-qTOF-MS; phenolic compounds; gallic and ellagic derivatives; flavonols; cyanidin-glucoside.

Mol2Net YouTube channel: *http://bit.do/mol2net-tube*

1. Introduction

Psidium guajava L., from the Myrtaceae family, is common throughout tropical and subtropical areas[1] and Andalusia is one of the regions in Europe where guava is grown. Guava tree shows different phenological stages throughout its vegetative period in response to environmental conditions[2]. Moreover, is widely used like food and in folk medicine all around the world. Many studies have demonstrated that guava leaves have anti-hyperglycaemic and anti-hyperlipidemic activities[3], among others. The biological activity of guava leaves belongs mainly to phenolic compounds[4].

Different analytical techniques are commonly used to characterize the bioactive present in plant extracts. LC/MS technique has opened up new approaches for the qualitative and the quantitative analysis of target compounds. LC/TOF/MS can provide tentative identification of unknown peaks, due to accurate-mass measurement[5].

Thus, the aim of this work was to study the phenolic compounds present in Andalusian guava leaves at different oxidation state (low, medium and high) by HPLC-DAD-ESI-qTOF-MS.

.

2. Results and Discussion

Negative and positive mode LC-ESI/MS conditions were optimized for the analysis of all the phenolics. To identify compounds for which no commercial standards were available, data generated by TOF analysis were checked. HPLC and mass spectral data obtained are summarized in Table 1. A total of sixty-nine phenolic compounds were characterized in negative mode, twelve of them were determined for the first time in guava leaves. In positive mode, only cyanidin-3-O-glucoside was detected.

Quantification of the extracts by HPLC-DAD-ESI-qTOF-MS revealed that the three samples showed significant differences ($p < 0.05$). Low oxidation state provided the highest content of total phenolic compounds (103 ± 2 mg/g leaf d.w.), followed by medium and high oxidation state (92.0 ± 0.4 and 87.91 ± 0.04 mg/g leaf d.w., respectively).

In terms of concentration of the different families present in leaves, the extracts reported the same trend as TPC, lowest content of different classes of phenolics compound were found in high oxidation state, whereas the highest content was found at low oxidation state (Figure 1). The major class of phenolic compounds in guava leaves samples was flavonols, ranged between 48.1 and 50.6 mg/g leaf d.w. The second class of polar compounds was represented by flavan-3-ols (24.2 - 24.7 mg/g leaf d.w.), succeeded by gallic and ellagic acid derivatives (14.8 - 15.8 mg/g leaf d.w.) and finally, flavanone, that varied from 0.49 to 0.63 mg/g leaf d.w.

In contrast, and as was expected, in positive mode, opposite results were found (Figure 2). Greater amount of cyanidin-3-O-glucoside was determined when the oxidation state was higher (varying from 441.28 ± 0.04 to 29.5 ± 0.2 µg/g leaf d.w.).

These changes in composition could be due to the different synthesis of secondary metabolites as response to oxidative state[6]. It may happen due to a reaction between them and anthocyanins that cause the dramatic red coloration of thee leaves, decreasing in this way its concentration[7].

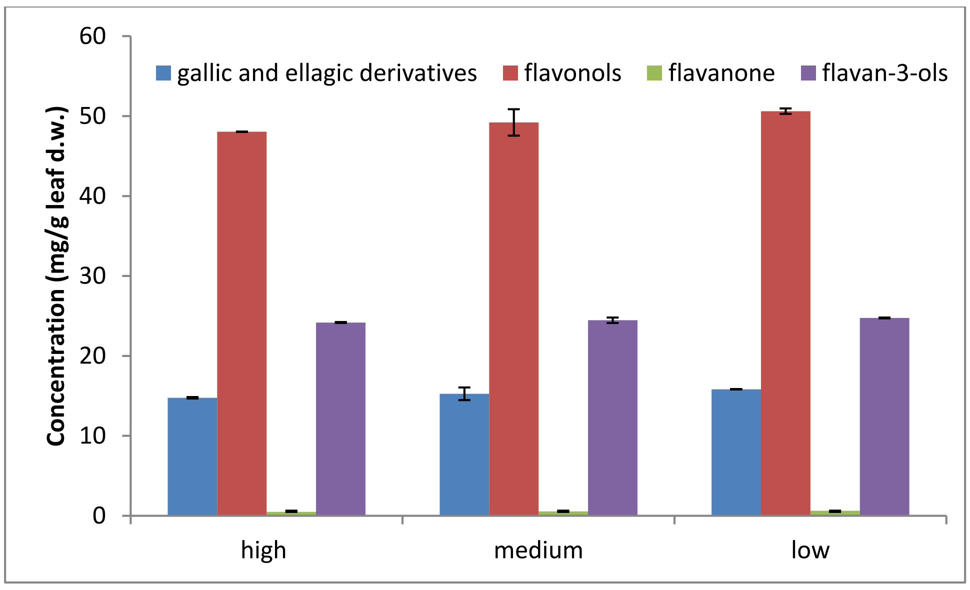


Figure 1. Quantification of different families of phenolic compounds present in guava leaves.

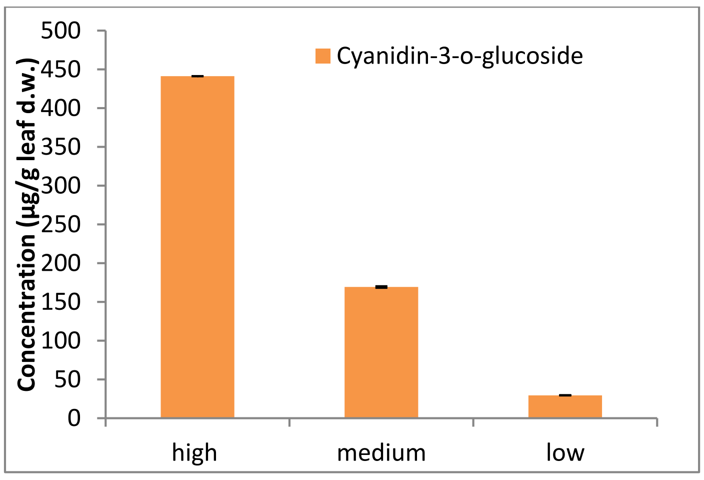


Figure 2. Quantification of cyanidin-3-O-glucoside

No.	Compound	tr (min)	m/z exp	m/z calculated	Molecular Formula	λ (nm)	Fragments	Score	error(ppm)
	Negative mode								
1	HHDP glucose Isomer 1	1.929	481.064	481.3406	C20H18O14	290	421.0406, 300.9991, 275.0202	96.51	-2.55
2	HHDP glucose Isomer 2	2.139	481.0638	481.3406	C20H18O14	290	421.0406, 300.9991, 275.0202	99.09	-0.19
3	HHDP glucose Isomer 3	2.516	481.0639	481.3406	C20H18O14	290	421.0406, 300.9991, 275.0202	97.21	-2.24
4	Prodelphinidin B Isomer	3.85	609.1276	609.5111	C30H26O14	272, 225	441.0838, 423.0701, 305.0687, 125.0226	97.84	-1.7
5	Gallic acid	4.022	169.0142	169.1116	C7H6O5	280, 360	125.0243	99.27	0.37
6	Pedunculagin/ Casuariin Isomer	5.865	783.0699	783.5332	C34H24O22	253	481.0606, 391.0307, 300.9999, 275.0191	98.57	-1.29
7	Prodelphinidin Dimer Isomer	7.272	593.1311	593.5117	C30H26O13	280, 340	425.0875, 289.0715,	96.51	-2.35
8	Gallocatechin	7.814	305.0698	305.2595	C15H14O7	270	125.0241, 179.0347, 219.0661, 261.0774	95.55	-3.32
9	Vescalagin/castalagin Isomer	7.953	933.0649	933.6216	C41H26O26	260, 280	466.0299, 300.9968	99.19	-0.79
10	Prodelphinidin Dimer Isomer	8.119	593.1316	593.5117	C30H26O13	280, 340	305.0667, 423.0719, 441.0841	96.51	-2.35
11	Uralenneoside	9.387	285.0624	285.2268	C12H14O8	270	153.0193, 109.0279	97.8	-2.69
12	Geraniin Isomer	9.497	951.0749	951.6369	C41H28O27	270	907.0825, 783.0785, 481.0606, 300.9999	99.56	-0.2
13	Pedunculagin/ Casuariin Isomer	9.536	783.0699	783.5332	C34H24O22	253	481.0606, 391.0307, 300.9999, 275.0191	98.39	-1.36
14	Geraniin Isomer	9.652	951.0752	951.6369	C41H28O27	270	907.0825, 783.0785, 481.0606, 300.9999	99.56	-0.2
15	Procyanidin B Isomer	10.018	577.1367	577.5123	C30H26O12	278	425.0881, 407.0777, 289.0718, 125.0243	95.68	-2.55
16	Galloyl(epi)catechin-(epi)gallocatechin	10.345	745.142	745.6160	C37H30O17	280, 340	593.1302, 575.1214, 423.0694, 305.0688	96.9	-0.62
17	Procyanidin B Isomer	10.356	577.1367	577.5123	C30H26O13	278	425.0881, 407.0777, 289.0718, 125.0243	99.41	-0.61
18	Tellimagrandin I Isomer	10.738	785.0851	785.5491	C34H26O22	279, 340	615.0674, 392.0396, 300.9985, 169.0144	99.13	-0.96
19	Pterocarinin A	10.998	1067.122	1067.7521	C46H36O30	280	533.0585, 377.0313, 301.0330, 249.0377	99.82	-0.11
20	Pterocarinin A Isomer	11.208	1067.122	1067.7521	C46H36O30	280	533.0585, 377.0313, 301.0330, 249.0377	98.39	-1.26
21	Stenophyllanin A	11.247	1207.1495	1207.8903	C56H40O31	278	917.0763, 603.0735	98.64	-1.08
22	Procyanidin trimer Isomer 1	11.247	865.1998	865.7645	C45H38O18	278	739.1593, 577.1337, 449.0888, 289.0745	97.53	-1.59
23	Catechin	11.258	289.0727	289.2601	C15H14O6	281	245.0821, 203.0718, 179.0349, 125.0242	96.76	-3.18
24	Procyanidin tetramer	11.336	1153.2612	1155.0246	C60H50O24	280	576.1291	99.6	-0.5
25	Procyanidin trimer Isomer 2	11.413	865.1998	865.7645	C45H38O18	278	739.1593, 577.1337, 449.0888, 289.0745	97.53	-1.59
26	Guavin A	11.496	1223.1423	1223.8897	C56H40O32	277	611.0724	99.05	0.85

27	Casuarinin/ Casuarictin Isomer	11.895	935.081	935.6375	$C_{41}H_{28}O_{26}$	275	783.0637, 633.0735, 300.9979, 275.0189	97.67	-1.43
28	Galloyl(epi)catechin-(epi)gallocatechin	12.1	745.142	745.6160	$C_{37}H_{30}O_{17}$	280, 340	593.1302, 575.1214, 423.0694, 305.0688	96.9	-0.62
29	Procyanidin pentamer	12.144	1441.3234	1442.2688	$C_{75}H_{62}O_{30}$	280	720.1604	95.66	1.97
30	Galloyl-(epi)catechin trimer Isomer 1	12.166	1017.2097	1017.8687	$C_{52}H_{42}O_{22}$	280	508.104	99.72	-0.01
31	Gallocatechin	12.327	305.0702	305.2595	$C_{15}H_{14}O_{7}$	270	125.0241, 179.0347, 219.0661, 261.0774	95.55	-3.32
32	Tellimagrandin I Isomer	12.504	785.0855	785.5491	$C_{34}H_{26}O_{22}$	277, 338	615.0674, 392.0396, 300.9985, 169.0144	98.44	-1.38
33	Vescalagin	12.758	933.0649	933.6216	$C_{41}H_{26}O_{26}$	260, 280	466.0295, 457.0781, 300.9968	96.33	-0.8
34	Stenophyllanin A Isomer	12.925	1207.1472	1207.8903	$C_{56}H_{40}O_{31}$	280	917.0763, 603.0735	98.37	0.89
35	Galloyl-(epi)catechin trimer Isomer 2	12.985	1017.2097	1017.8687	$C_{52}H_{42}O_{22}$	280	508.104	98.17	-1.35
36	Myricetin hexoside Isomer	13.284	479.0836	479.3678	$C_{21}H_{20}O_{13}$	261, 358	317.0294, 316.0226, 271.0255	98.36	-0.92
37	Stachyuranin A	13.412	1225.1587	1225.9055	$C_{56}H_{42}O_{32}$	276	612.0779	95.54	1.35
38	Procyanidin gallate Isomer	13.517	729.1476	729.6166	$C_{37}H_{30}O_{16}$	280	577.1356, 559.1226, 425.0874, 407.0798, 298.0716	96.89	-1.91
39	Myricetin hexoside Isomer	13.677	479.0835	479.3678	$C_{21}H_{20}O_{13}$	261, 358	317.0294, 316.0226, 271.0255	97.89	-0.08
40	Vescalagin/castalagin Isomer	13.844	933.0645	933.6216	$C_{41}H_{26}O_{26}$	260	466.0299, 300.9968	88.32	-1.57
41	Myricetin -arabinoside/ xylopyranoside Isomer	13.988	449.0728	449.3418	$C_{20}H_{18}O_{12}$	264, 356	317.0291, 316.0241, 271.0249	98.39	-1.65
42	Myricetin -arabinoside/ xylopyranoside Isomer	14.214	449.0726	449.3418	$C_{20}H_{18}O_{12}$	264, 357	317.0291, 316.0241, 271.0249	98.02	-1.65
43	Procyanidin gallate Isomer	14.563	729.6356	577.5123	$C_{30}H_{26}O_{12}$	280	577.1356, 559.1226, 425.0874, 407.0798, 298.0716	98.17	-1.73
44	Myricetin -arabinoside/ xylopyranoside Isomer	14.99	449.0726	449.3418	$C_{20}H_{18}O_{12}$	264, 356	317.0291, 316.0241, 271.0249	98.66	-1.65
45	Myricetin hexoside Isomer	15.034	479.0839	479.3678	$C_{21}H_{20}O_{13}$	261, 358	317.0294, 316.0226, 271.0255	97.08	-1.92
46	Myricetin hexoside Isomer	15.217	479.0841	479.3678	$C_{21}H_{20}O_{13}$	264, 356	317.0288, 316.0241, 271.0253	97.08	-1.92
47	Myricetin -arabinoside/ xylopyranoside Isomer	15.604	449.0743	449.3418	$C_{20}H_{18}O_{12}$	264, 356	317.0291, 316.0241, 271.0249	98.39	-1.65
48	Quercetin -galloylhexoside Isomer	15.626	615.1008	615.4726	$C_{28}H_{24}O_{16}$	268, 350	463.0886, 300.0283	99.16	-0.98
49	Ellagic acid deoxyhexoside	15.837	447.0578	447.3259	$C_{20}H_{16}O_{12}$	265, 350	300.9974,	91.25	-3.19
50	Quercetin -galloylhexoside Isomer	16.036	615.0999	615.4726	$C_{28}H_{24}O_{16}$	280, 345	463.0886, 300.0283	99.16	-0.98
51	Myricetin -arabinoside/ xylopyranoside Isomer	16.191	449.0736	449.3418	$C_{20}H_{18}O_{12}$	256, 356	317.0291, 316.0241, 271.0249	98.39	-1.65
52	Morin	16.28	301.0362	301.2278	$C_{15}H_{10}O_{7}$	257, 374	178.9978, 151.0032	97.46	-2.5
53	Myricetin -arabinoside/ xylopyranoside Isomer	16.462	449.0735	449.3418	$C_{20}H_{18}O_{12}$	257, 356	317.0291, 316.0241, 271.0249	98.39	-1.65

54	Ellagic acid	16.507	300.9996	301.1847	$C_{14}H_6O_8$	254, 360	283.9921, 257.0088, 229.0169, 185.0233	98.88	-1.71
55	Hyperin	16.616	463.0895	463.3684	$C_{21}H_{20}O_{12}$	259, 355	301.0350, 300.0279, 178.9980, 151.0032	96.41	-2.65
56	Quercetin glucoronide	16.723	477.0659	477.3519	$C_{21}H_{18}O_{13}$	265, 355	301.0359, 151.0026	98.1	-1.83
57	Isoquercitrin	16.95	463.0893	463.3684	$C_{21}H_{20}O_{12}$	258, 355	301.0353, 300.0281, 178.9983, 151.0090	97.04	-2.33
58	Procyanidin gallate Isomer	17.038	729.1476	729.6166	$C_{37}H_{30}O_{16}$	280	577.1356, 559.1226, 425.0874, 407.0798, 298.0716	96.89	-1.91
59	Reynoutrin	17.498	433.0792	433.3424	$C_{20}H_{18}O_{11}$	258, 356	301.0356	95.94	-2.9
60	Guajaverin	17.802	433.0795	433.3424	$C_{20}H_{18}O_{11}$	257, 356	301.0352	97.99	-1.91
61	Guavinoside A	17.985	543.1159	544.4610	$C_{26}H_{24}O_{13}$	218, 288	313.0568, 229.0503, 169.0148	98.1	-1.77
62	Avicularin	18.206	433.0803	433.3424	$C_{20}H_{18}O_{11}$	257, 355	301.0359	96.7	-2.2
63	Quercitrin	19.194	447.0947	447.3690	$C_{21}H_{20}O_{11}$	264, 353	301.0348, 271.0247, 178.9988, 151.0028	95.23	-3.02
64	Myrciaphenone B	19.208	481.0999	481.3836	$C_{21}H_{22}O_{13}$	280, 340	313.0570, 169.0141	97.2	-2.23
65	Guavinoside C	19.768	585.0898	585.4466	$C_{27}H_{22}O_{15}$	265, 355	433.0757, 301.0351, 283.0449, 169.0142	97.19	-1.92
66	Guavinoside B	20.77	571.147	571.5062	$C_{28}H_{28}O_{13}$	218, 283	313.057, 257.0829, 169.0142	97.26	-2.05
67	Guavinoside A Isomer	20.702	543.1159	543.4530	$C_{26}H_{24}O_{13}$	218, 288	313.0568, 229.0503, 169.0148	98.1	-1.77
68	Guavinoside B Isomer	21.667	571.147	571.5062	$C_{28}H_{28}O_{13}$	218, 283	313.057, 257.0829, 169.0142	97.26	-2.05
69	2,6-dihydroxy-3-methyl-4-O-(6″-O-galloyl-β-D-glucopyranosyl)-benzophenone	21.971	557.1318	557.4796	$C_{27}H_{26}O_{13}$	280	313.0575, 243.0670, 169.0146	96.93	-2.12
70	Guavin B	22.237	693.111	693.5414	$C_{33}H_{26}O_{17}$	283	391.1468	97.82	-1.67
71	Quercetin	22.314	301.0358	301.2278	$C_{15}H_{10}O_7$	257, 374	178.9985, 151.0036	98.9	-1.34
72	Naringenin	26.738	271.0622	271.2448	$C_{15}H_{12}O_5$	280	118.6395, 150.5022	96.09	-3.67
	Positive mode								
73	Cyanidin-3-o-glucoside	3.661	449.1089	449.3911	$C_{21}H_{21}O_{11}$	287, 288	517, 280	96.97	-2.34

Table 1. Tentatively identified compounds in guava leaves.

3. Materials and Methods

3.1 Plant Material and Sample Preparation

Fresh guava leaves were harvested in Motril, Spain, at different oxidation states (low, medium, and high). The samples were air-dried, grounded and extracted with ethanol:water 80/20 (v/v) by ultrasonics[8].

3.2 HPLC-DAD-ESI-qTOF-MS Analysis

Chromatographic analyses were performed using an HPLC Agilent 1260 series (Agilent Technologies, Santa Clara, CA, USA) equipped with a binary pump, an online degasser, an autosampler and a thermostatically controlled column compartment, and a UV-Vis Diode Array Detector (DAD). The column was maintained at 25ºC. Phenolic compounds from *Psidium guajava* L. leaves were separated at room temperature using a method previously reported by Gómez-Caravaca et al.[9] for positive mode.

MS analyses were carried out using a 6540 Agilent Ultra-High-Definition Accurate-Mass Q-TOF-MS coupled to the HPLC, equipped with an Agilent Dual Jet Stream electrospray ionization (Dual AJS ESI) interface. In negative ionization mode at the following conditions: drying gas flow (N2), 12.0 L/min; nebulizer pressure, 50 psi; gas drying temperature, 370°C; capillary voltage, 3500 V; fragmentor voltage and scan range were 3500 V and m/z 50-1500, respectively. Automatic MS/MS experiments were carried out using the followings collision energy values: m/z 100, 30 eV; m/z 500, 35 eV; m/z 1000, 40 eV; and m/z 1500, 45 eV. In positive mode: auto MS/MS experiments were carried out using the followings collision energy values: m/z 100, 40 eV; m/z 500, 45 eV; m/z 1000, 50 eV; and m/z 1500, 55 eV.

.

4. Conclusions

HPLC coupled to qTOF-MS detector, which provides a molecular formula and the MS/MS data, permitted the analysis of the major phenolic compounds of guava leaves. The method performed in negative mode has proven to be successful to identify 72 compounds in guava leaves. Moreover, in positive mode, the analysis with TOF analyser and the co-elution with a standard solution allowed the identification of the cyanidin-glucoside. To our knowledge twelve compounds from the negative mode, and also the cyaniding-glucoside, were identified for the first time in guava leaves.

Quantification data, in negative mode, reported that leaves with low oxidation state presented the highest concentration of these compounds and decreased when the oxidation state raised. On the contrary, the state of oxidation affected significantly the cyanidin content. In fact, highest amount was detected in the leaves with high oxidation state.

Acknowledgments

The authors Ana María Gómez-Caravaca and Vito Verardo thank the Spanish Ministry of Economy and Competitiveness (MINECO) for “Juan de la Cierva” post-doctoral contracts.

Author Contributions

EDdC carried out the experimental analyses, data interpretation and manuscript writing; AMGC and VV design the experimental plan and were involved in the data interpretation and manuscript redaction; AFG and ASC were the responsibly of the project and founded the financial sources, moreover, they helped in the data interpretation.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1 Morton, J. F. Guava. In *Fruits of Warm Climates*, Miami, FL, 1987, pp. 356–363.

2 Salazar, D. M.; Melgarejo, P.; Martínez, R.; Martínez, J. J.; Hernández, F.; Burguera, M. *Sci. Hortic. (Amsterdam).* **2006**, *108*, 157–161.

3 Deguchi, Y.; Miyazaki, K. *Nutr. Metab. (Lond).* **2010**, *7* (9), 1–10.

4 Gutiérrez, R. M. P.; Mitchell, S.; Solis, R. V. *J. Ethnopharmacol.* **2008**, *117* (1), 1–27.

5 Heng, M. Y.; Tan, S. N.; Yong, J. W. H.; Ong, E. S. *TrAC Trends Anal. Chem.* **2013**, *50*, 1–10.

6 Vargas-Alvarez, D.; Soto-Hernández, M.; González-Hernández, V. A.; Engleman, E. M.; Martínez-Garza, Á. *Agrociencia* **2006**, *40*, 109–115.

7 Coley, P. D.; Barone, J. a. *Annu. Rev. Ecol. Syst.* **1996**, *27* (1), 305–335.

8 Díaz-de-Cerio, E.; Verardo, V.; Gómez-caravaca, A. M.; Fernández-Gutiérrez, A.; Segura-carretero, A. *J. Chem.* **2015**, *2015*, 1–9.

9 Gómez-Caravaca, A. M.; Verardo, V.; Toselli, M.; Segura-Carretero, A.; Fernández-Gutiérrez, A.; Caboni, M. F. *J. Agric. Food Chem.* **2013**, *61* (22), 5328–5337.

SciForum
Mol2Net

Nanoparticulate Fe_2O_3 and Fe_2O_3/C Composites as Anode Materials for Li-Ion Batteries

Amaia Iturrondobeitia [1], Aintzane Goñi [2] and Luis Lezama [2,*]

[1] Parque Tecnológico de Álava, CIC energiGUNE, Albert Einstein 48, 01510 Miñano, Álava, Spain; E-Mail: a.iturrondobeitia@cicenergigune.com

[2] Departamento de Química Inorgánica, Universidad del País Vasco UPV/EHU, P.P. Box. 644, E-48080 Bilbao, Spain; E-Mail: aintzane.goni@ehu.es; luis.lezama@ehu.es;

* Author to whom correspondence should be addressed; E-Mail: luis.lezama@ehu.es; Tel.: +34-946-012-703 ; Fax: +34-946-013-500.

Published: 4 December 2015

Abstract: Nanoparticulate α-Fe_2O_3 and γ-Fe_2O_3/C composites with different carbon proportions have been prepared by a freeze-drying synthesis procedure. The characterization has been performed by elemental analysis, X-ray powder diffraction (XRD), transmission electron microscopy (TEM), electron paramagnetic resonance (EPR) and magnetic susceptibility measurements. Morphological studies revealed that nanoparticles of γ-Fe_2O_3 in the composites are well-dispersed in the matrix of amorphous carbon. The magnetic and spectroscopic measurements corroborated that described composition and morphology. The electrochemical study showed that composites with carbon have promising electrochemical performances. These samples yielded specific discharge capacities of 1200 mAh/g after operating for 100 cycles at 1C. These excellent results could be explained by the homogeneity of particle size and structure as well as the uniform distribution of γ-Fe_2O_3 nanoparticles in the in situ generated amorphous carbon matrix.

.

.

Keywords: Nanoparticle; composite; Li-ion batteries; anode; Fe_2O_3; freeze-drying

Mol2Net YouTube channel*: http://bit.do/mol2net-tube*

1. Introduction

Lithium ion batteries (LIBs) have attracted extensive attention in various fields such as portable electronic devices or electric/hybrid electric vehicles owing to its high energy density, long cycle life, rapid charge capability, and no memory effect[1]. For commercial LIBs, the positive electrode material is typically a metal oxide or phosphate while graphite is the most widely used anode material. However, to develop higher energy density systems new materials are required[2]. In this sense, 3d transition metal oxides (MO_x) are one of the most promising next-generation anode materials under consideration[3]. In these compounds, contrary to conventional intercalation or alloying mechanisms, the high Li^+ storage capacity is derived from conversion reactions.

Among the metal oxides, Fe_2O_3 has been regarded as a suitable candidate because of its high theoretical capacity (1007 mAh/g), nontoxicity, high corrosion resistance, and low materials and processing costs[4]. However, further optimization of iron oxides as anode materials is needed due to their poor cycling performance and limited rate capability. Introducing carbonaceous materials the electronic contact between the iron oxide particles is increased, and the volume and structural changes associated with the transformation of iron oxide particles into metallic iron during the discharge/ charge process are buffered[5-6]. This improves the cycling stability and rate capability of iron oxides, but a facile synthesis method for large-scale production of these materials for anode materials of LIBs is required. Here, a freeze-drying synthesis procedure, which can be easily implemented as part of an industrial process, has been employed. This synthesis method has already been proven in our group to be useful to obtain nanosized cathodic materials with exceptional electrochemical performance[7].

In this paper we report on the structural, morphologic, spectroscopic, magnetic and electrochemical characterization of nanoparticulate Fe_2O_3 and Fe_2O_3/C composites prepared by the freeze-drying method.

.

2. Results and Discussion

For the nanoparticulate Fe_2O_3 sample calcined in air, hereafter **Fe_2O_3_air**, the X-ray diffraction peaks could be indexed to hematite α-Fe_2O_3 structure. XRD patterns of composites prepared with different carbon content are very similar, and all of the diffraction peaks could be indexed to maghemite γ-Fe_2O_3. The calculated size of the crystals from the deconvolution of (311) diffraction maxima is in the range of 10 to 20 nm for samples with a carbon content determined by elemental analysis of 20 to 40 %. Transmission electron micrograph of these samples show that they consist of spherical nanoparticles embedded in the in situ generated carbon matrix. The nanometric particle size can be attributed to the carbon source that prevents the growth of the particle. The TEM image of the most homogeneous sample, **Fe_2O_3_23%C**, is presented in Fig. 1. This sample (D=15 nm) was chosen to carry out the comparative electrochemical study.

The EPR spectra of the samples were recorded at X-band (9.4 GHz) at room temperature. A similar pattern was observed in all cases: a very broad and intense signal centered at a g-value close to 2.18−2.20, characteristic of ferri/ ferromagnetic systems. It is also found that composites with higher amount of carbon have a more isotropic EPR signal, probably due to the smaller γ-Fe_2O_3 particle size.

The magnetization of the Fe_2O_3_air sample is close to saturation under relatively small applied fields (<5 kOe). However, there is also quasi-linear contribution that is almost independent of temperature. This feature is compatible with the presence of an antiferromagnetic phase of high ordering temperature as it is α-Fe_2O_3. After subtraction of this antiferromagnetic component, a relative content for the γ- Fe_2O_3 ferromagnetic phase of at least 13% in weight is estimated. In contrast, the magnetization of Fe_2O_3_23%C perfectly saturates at 80 emu/g, which is very close to the value of pure maghemite in bulk. The coercive field is ~250 Oe, smaller than in Fe_2O_3_air (450 Oe), in agreement with values reported in the literature for maghemite nanoparticles of similar sizes[8]. In composites with higher carbon content, the lack of saturation at high fields (>20 kOe) can be explained as originating from very small and isolated magnetic nanoparticles, whose magnetization does not saturate at 5 K and consequently also reduces the average magnetization of the sample[9]. Magnetization versus temperature measurements for these samples showed the presence of para- and superparamagnetic contributions ascribed to very small superparamagnetic γ-Fe_2O_3 nanoparticles.

The theoretical capacity of Fe_2O_3 for complete reduction of Fe^{3+} to Fe^0 is 1007 mAh/g, corresponding to a maximum lithium uptake of 6 Li^+ per Fe_2O_3 formula unit. However, our samples exhibit higher initial values (1216 and 1451 mAh/g for Fe_2O_3_air and Fe_2O_3_23%C, respectively) due to the partial decomposition of the electrolyte[10] and the extra capacity provided by the carbon in the composite[11].

The cycling stability of the electrodes based on Fe_2O_3_air and Fe_2O_3_23%C, was investigated at a current rate of 1C. As shown in Figure 2, the capacity retention was better for the carbon composite. This fact could be attributed to the heterogeneity of the Fe_2O_3_air sample as well as to the lack of an in situ generated carbon matrix in this compound. The carbon matrix enhances the cycling performance of the samples since it improves the electrical connection between the nanoparticles and facilitates the accommodation of the structure change associated with successive charge and discharge cycles. Figure 2 also shows the Coulombic efficiencies at C/10 (defined as the ratio between charge and discharge capacity) of the cells for the first 20 cycles. The Coulombic efficiencies in the first cycles were only 70% and 53% for the Fe_2O_3_air and Fe_2O_3_23%C cells, respectively. However, in the successive cycles these values increased noticeably. The highest Coulombic efficiency after the first cycle was that of the cell containing the Fe_2O_3_23%C composite, probably in large part resulting from the homogeneous particle size distribution of the γ-Fe_2O_3.

Thus, composite materials show a promising electrochemical performance providing a specific discharge capacity of around 1200 mAh/g after operating for 100 cycles at 1C. These excellent results could be explained by considering the structural and particle size homogeneity as well as the uniform distribution of the NPs in the in situ generated amorphous carbon matrix.

.

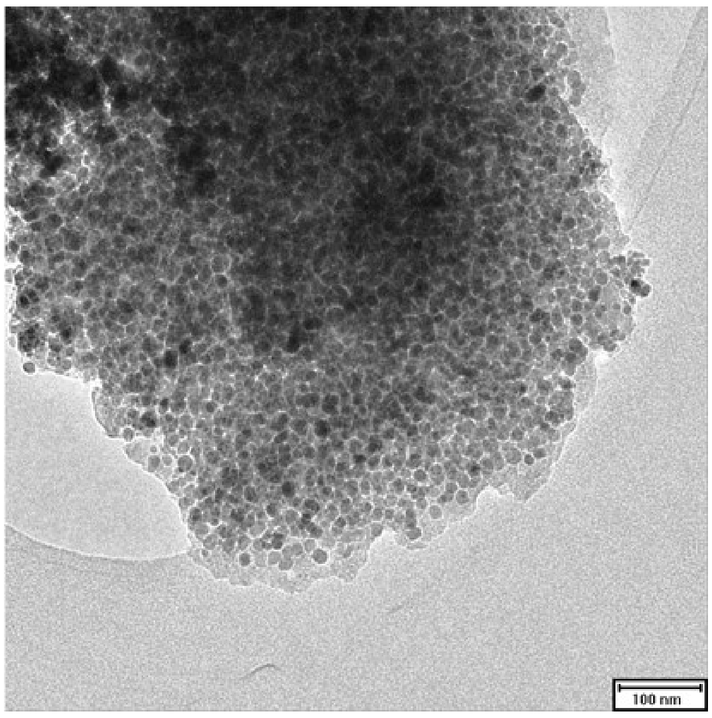


Figure 1. TEM image of Fe_2O_3_23%C.

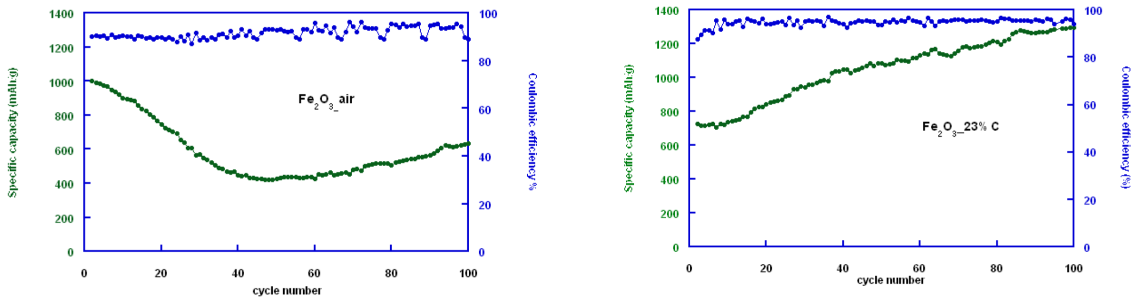


Figure 2. Ciclability at 1C rate and Coulombic efficiencies at C/10 for Fe_2O_3_air and Fe_2O_3_23%C.

3. Materials and Methods

Citric acid monohydrate (99.5 %) and iron(III) citrate (98%) were purchased from Sigma-Aldrich and used as received without further purification. $C_6H_8O_7 \cdot H_2O$ and $C_6H_5FeO_7$ reagents were dissolved in water in required molar ratios. The resulting solutions were subsequently frozen in a round-bottom flask that contained liquid nitrogen. Afterward, the round-bottom flasks were connected to the freeze-dryer for 48 h at a pressure of 3 °— 10^{-1} mbar and a temperature of −80 °C to sublime the solvent. The as-obtained precursors were subjected to a single heat treatment at 400 °C for 6 h in air or nitrogen atmosphere and ball-milled for 30 min.

Structural characterization of the samples was carried out using X-ray powder diffraction with a Bruker D8 Advance Vario diffractometer. The morphologies of the materials were studied by transmission electron microscopy (TEM) using a FEI TECNAI F30. A Bruker ELEXSYS 500 spectrometer operating at the X-band was used to record the EPR polycrystalline spectra. Magnetic susceptibility measurements were carried out between 5 and 300 K with a Quantum Design SQUID magnetometer. CR2032 coin cells were assembled to evaluate the electrochemical performances of the samples. All the electrochemical measurements were carried out on a Bio-Logic VMP3 potentiostat/galvanostat at room temperature. The galvanostatic charge/discharge experiments were performed between 0.1 and 3 V at 0.1C and 1C current rates.

.

4. Conclusions

Fe_2O_3/C composites with different carbon content were successfully prepared by a freeze-drying method, crystallizing in the γ-Fe_2O_3 maghemite structure. TEM micrographs have shown that the Fe_2O_3 particles are well dispersed in the matrix of amorphous carbon that was generated in situ during the synthesis procedure. Electron paramagnetic resonance spectra are in good agreement with the composition and morphology of each sample. The magnetic measurements showed a lack of saturation of the magnetization at high magnetic fields (>20 kOe) due to the presence of very small isolated maghemite nanoparticles. The galvanostatic experiments revealed that composite materials show a promising electrochemical performance. This could be attributed to several factors, such as the suitable morphology of the samples and the good connection that provides the in situ generated carbon matrix without blocking the lithium ion diffusion pathways..

Acknowledgments

This work was financially supported by the Ministerio de Economía y Competitividad (MAT2013-41128-R) and the Gobierno Vasco/Eusko Jaurlaritza (IT570-13). Technical and human support: SGIker (UPV/EHU)

Author Contributions

All authors contributed equally to this work.

Conflicts of Interest

The authors declare no conflict of interest

References and Notes

1. Pistoia G. *Lithium-Ion Batteries: Advances and Applications*, 1st ed.; Elsevier, Amsterdam, The Netherlands, 2014.
2. Van Noorden, R. The rechargeable revolution: A better battery. *Nature* 2014, 507, 26-28.
3. Poizot, P.; Laruelle, S.; Grugeon, S.; Dupont, L.; Tarascon, J. M. Nano-sized transition-metal oxides as negative-electrode materials for lithium-ion batteries *Nature* 2000, 407, 496−499.
4. Chen, J.; Xu, L. N.; Li, W. Y.; Gou, X. L. α-Fe_2O_3 nanotubes in gas sensor and lithium-ion battery applications. *Adv. Mater.* 2005, 17, 582−586.
5. Shi, W.; Zhu, J. X.; Sim, D. H.; Tay, Y. Y.; Lu, Z. Y.; Zhang, X. J.; Sharma, Y.; Srinivasan, M.; Zhang, H.; Hng, H. H.; Yan, Q. Y. Achieving high specific charge capacitances in Fe_3O_4/reduced graphene oxide nanocomposites. *J. Mater. Chem.* 2011, 21, 3422−3427.
6. Ban, C. M.; Wu, Z. C.; Gillaspie, D. T.; Chen, L.; Yan, Y. F.; Blackburn, J. L.; Dillon, A.C. Nanostructured Fe_3O_4/SWNT electrode: Binder-Free and high-rate Li-Ion anode. *Adv. Mater.* 2010, 42, E145−E149.

7. Palomares, V.; Goñi, A.; Gil de Muro, I.; de Meatza, I.; Bengoechea, M.; Cantero, I.; Rojo, T. Influence of carbon content on $LiFePO_4$/C samples synthesized by freeze-drying process. *J. Electrochem. Soc.* 2009, 156 (10), A817.
8. Ramos Guivar, J. A.; Bustamante, A.; Flores, J.; Mejía Santillán, M.; Osorio, A. M.; Martinez, A. I.; De Los Santos Valladares, L.; Barnes, C. H. W. Mössbauer study of intermediate superparamagnetic relaxation of maghemite (γ-Fe_2O_3) nanoparticles. *Hyperfine Interact.* 2014, 224, 89−97.
9. Iturrondobeitia, A.; Goñi, A.; Orue, I.; Gil de Muro, I.; Lezama, L.; Doeff, M.; Rojo, T. Effect of Carbon coating on the physicochemical and electrochemical properties of Fe_2O_3 nanoparticles for anode application in high performance lithium ion batteries. *Inorg. Chem.*, 2015, 54, 5239-5248.
10. Jin, S.; Deng, H.; Long, D.; Liu, X.; Zhan, L.; Liang, X.; Qiao, W.; Ling, L. Facile synthesis of hierarchically structured Fe_3O_4/carbon micro-flowers and their application to lithium-ion battery anodes. *J. Power Sources* 2011, 196, 3887.
11. López, M. C.; Ortiz, G. F.; Lavela, P.; Alcántara, R.; Tirado, J. L. Improved energy storage solution based on hybrid oxide materials. *ACS Sustainable Chem. Eng.* 2013, 1, 46−56.

SciForum
Mol2Net

Chiral Imines on the Wave: Reactivity of *tert*-Butyl Acrylate and Stereoselectivity Determination Using NMR in Liquid Crystals

Lucie VANDROMME [1] , Li CHEN [1] , Lai WEI [1] , Franck LE BIDEAU [1], André LOUPY [2] , Philippe LESOT [3] , Olivier LAFON [3] , Elise TRAN HUU DAU [4] Pierre CHAMINADE [5] and Françoise DUMAS [1,*]

[1] BioCIS, UMR CNRS 8076, Université Paris Saclay, School of Pharmacy, Université Paris-Sud, 5, rue J.-B. Clément, F-92296 Châtenay-Malabry, France; E-Mails: lucie.vandromme@free.fr (L. V.); li.chen1@u-psud.fr (L. C.); lai.wei@u-psud.fr (L. W.); franck.lebideau@u-psud.fr (F. L. B.); francoise.dumas@u-psud.fr (F.D.);

[2] LRSSS, UMR CNRS 8615, Université Paris Saclay, UFR de Sciences, Université Paris-Sud, ICMMO, Bât. 410, F-91405 Orsay, France; E-Mail: andre.loupy@cegetel.net (A.L.);

[3] Laboratoire de RMN en Milieu Orienté, CNRS UMR 8182, ICMMO, Bât. 410, Université Paris Saclay, UFR de Sciences, 91405 Orsay cedex, France. E-mail: philippe.lesot@u-psud.fr

[4] ICSN, UPR CNRS 2301,1 avenue de la Terrasse, F-91190 Gif sur Yvette, France; E-Mail: elise.tran@icsn.cnrs-gif.fr.

[5] Lip(Sys)2, EA 4041, Université Paris Saclay, School of Pharmacy, Université Paris-Sud, 5, rue J.-B. Clément, F-92296 Châtenay-Malabry, France ; Email : pierre.chaminade@u-psud.fr

* Author to whom correspondence should be addressed; E-Mail: francoise.dumas@u-psud.fr; Tel.: +33-146-835-563.

Published: 4 December 2015

Abstract: In connection with synthetic applications, we have foreseen to study the reactivity of the poorly reactive *tert*-butyl acrylate electrophile with chiral imines of unsymmetrical ketones, using conventional or microwave flash heating under carefully controlled reaction conditions in order to develop a selective and efficient access to the corresponding Michael adducts in which a created stereogenic quaternary carbon center was fully stereocontrolled. Depending on the conditions, either a keto ester or a lactam were obtained. A good correlation was obtained between experiment and theoretical calculations. The stereoselectivity of the process (e>95%) was determined using natural abundance ^{13}C-{^{1}H} NMR in a chiral polypeptide liquid crystal. The scope of the reaction was screened using a set of electrophilic alkenes giving Michael adducts in good yield and similar high enantiocontrol.

Keywords: Aza-ene process, Asymmetric Michael reaction, Enamine activation, Unsymmetrical ketones, *tert*-Butyl acrylate, Quaternary carbon center, chiral polypeptide, liquid crystal, ^{13}C-{^{1}H} NMR, MW assisted synthesis.

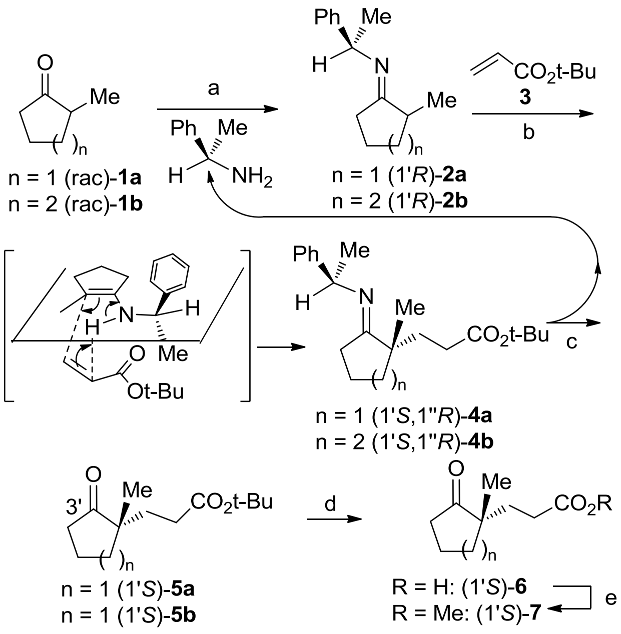
1. Introduction

Although a broad range of methods able to generate new carbon-carbon bonds exists, the establishment of quaternary carbon centres in the proper configuration is among the most restrictive in organic synthesis.[1] Since forty years, the asymmetric Michael addition of chiral imines (AMACI) under neutral conditions has attracted widespread attention as a versatile carbon-carbon bond-forming method, leading to Michael adducts with high levels of regio- and stereocontrol.[2] Such Michael adducts featuring a stereogenic tetrasubstituted carbon center are useful synthons for the asymmetric synthesis of a variety of bioactive compounds.[3] Besides its remarkable efficiency due to a concerted *aza-ene* type mechanism,[4] the reaction tolerates a large variability in both carbonyl compounds and Michael acceptors, with some limitations for hindered systems for which high pressure activation is required.[5] Moreover, in the context of improved reaction efficiency, clean processes and short reaction times are desired. In this respect, µW is an efficient way of promoting organic transformations, mainly in solvent-free systems. Thus, interest in µW assisted organic reactions has recently considerably increased.[6] Nevertheless, little attention has been devoted to the µW effects on selectivity,[7] particularly for asymmetric Michael reactions.[8] Herein, we describe our investigation of the reactivity of chiral imines with hindered tert-butyl acrylate and related electrophiles under microwave activation.

2. Results and Discussion

As a part of our program directed at exploring the scope of the AMACI, and in connection with synthetic applications for which an orthogonal ester protection cleavable in acidic medium was desired in the Michael adduct, we have foreseen to study the reactivity of *tert*-butyl acrylate **3** with chiral imine **2a** derived from 2-methyl-cyclopentanone **1a**. Although considerable attention has recently been focused on organocatalytic asymmetric transformations as efficient and convenient methodologies owing to their environmentally friendly characteristics, none of them address the reactivity of bulky electrophiles in this reaction.[9] Prior to engage in such studies, and because the chiral inductor is available at low price and easily recovered without loss of optical activity at the end of the process, we first analyze the stoichiometric transformation. Due to the steric hindrance of the acceptor, we have turned to use microwave irradiation (µW) under carefully controlled reaction conditions, in order to develop a simple and efficient access to Michael adduct **5a** and the derived keto-acid **6** and compared this method to conventional heating (Δ) (Scheme 1).

In general, the AMACI is carried out from the intermediary chiral imine and the electrophile, in the absence of solvent, at room temperature or using moderate Δ (up to 80 °C) as the regio- and stereocontrol of this reaction are only slightly sensitive to heat.[2] When the reaction was performed at 25 °C for 7 days, the conversion was not complete and 15% yield of Michael adduct **5a** was obtained upon hydrolysis (Table 1, entry 1). Thus, due to its bulkiness, *tert*-butyl acrylate **3** reacts much slower than its methyl counterpart.[10] At 60 °C, all the imine **2a** was consumed in one day; however, side polymerization reactions resulted in a modest 58% yield of adduct **5** (Table 1, entry 2).



Scheme 1 *Reagents and conditions*: (a) 1.01 eq 1-phenylethylamine, cyclohexane, reflux, Dean-Stark, overnight, 89% (b) 2 equiv. **2**, 25-200 °C (see Tables 1 and 2); (c) 20% aq. AcOH, THF, 20 °C (d) HCO$_2$H, 20 °C, 89%; (e) CH$_2$N$_2$, Et$_2$O, 0 to 20 °C, 2 h, 95%.

When the reaction was performed at 100 °C for 4 h, alkylated imine **4a**[11] was obtained as the sole product leading upon hydrolysis to keto ester **5** in 79% yield. On the basis of its ^{13}C NMR spectrum, crude imine **4a** exists as a single stereoisomer (de >95%),[11] as the result of a highly stereo-controlled process, giving rise to Michael adduct **5a** with a >95% enantiomeric excess (ee). The reaction duration was markedly reduced to 30 min at 150 °C and ketoester **5** was obtained in an optimum 92% yield upon hydrolysis. However, the stereoselectivity proved to be lower (ee 80%) than at 100 °C (Table 1, entry 6). Finally, both the efficiency and the stereoselectivity dropped at 200 °C (Table 1, entry 7).

.

Table 1. Effect of temperature on the synthesis of ketoester **4** by condensation of chiral imine **1** with *tert*-butyl acrylate **2**.[a]

Entry	Temperature (°C)	Time	**5** yield %[b]	**5** ee%[d]
1	25	7 d	15	nd[c]
2	60	1 d	58	> 95
3[b]	100	4 h	79	> 95
4	150	5 min	14	nd[c]
5	150	15 min	59	nd[c]
6	150	30 min	92	80
7	200	30 min	57	80

[a]: 2 Equivalents; [b]: Isolated yield of purified keto-ester **5**; [c]: Not determined
[d]: Determined by ^{13}C-{^{1}H} NMR in chiral liquid crystals.

Concerning the stereoselectivity of the process, attempts to measure accurately the ee in adduct **5a** or in the related keto-acid **6** using either chiral HPLC or ^{1}H NMR spectroscopy in the presence of chiral shift reagent [Eu(hfc)$_3$] were unsuccessful. However, screening of the selectivity was made possible on the examination of ^{13}CNMR spectra of crude imine **4a** and gave satisfactory results. In order to ascertain that no epimerization of Michael adduct **5a** occurred during hydrolysis of imine **4a**, we turned our attention to NMR spectroscopy in polypeptide chiral liquid crystals (CLC) that generally provides an efficient alternative to classical methods when these latter failed or gave poor results.[12] We used here ^{13}C-{^{1}H}.[13] Spectral enantio-discriminations using ^{13}C-{^{1}H} NMR in a CLC are based on ^{13}C chemical shift anisotropy (CSA) differences. In practice, when the enantiomers are oriented differently inside the CLC, we can expect to observe two distinct resonances for each non-equivalent carbon atom discriminated. A priori, each carbon atom is a potential spy, thus increasing the possibility to visualize enantiomers.[14] Various carbon sites show enantiodiscrimination, but the best spectral separation was obtained for C3' in compound **5**. Considering the S/N ratio, the error on the ee has been estimated around 5% of the true value. Figure 1 shows the evolution of two ^{13}C-{^{1}H} signals associated to C-3' (Scheme 2) both in racemic (a) and enantio-enriched forms [Table 1, entry 6 (b) and Table 2, entry 4 (c)] oriented in a PBLG/ dichloromethane phase. The differences in peak intensity reveal the evolution of the ee. Although the separation observed at 100 MHz is rather small (< 3 Hz), a suitable evaluation of the ee is possible using deconvolution process. The absence of peak for the minor enantiomer in Fig. 1c indicates that the ee is >95%. Accuracy of the method was ascertained by measuring gradual mixtures of the racemic ketoester **5a** (prepared from racemic 1-phenymethymamine using the same conditions) with the pure enantiomer.

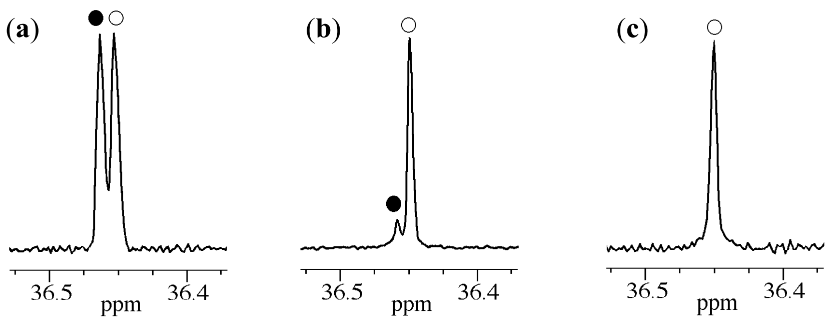
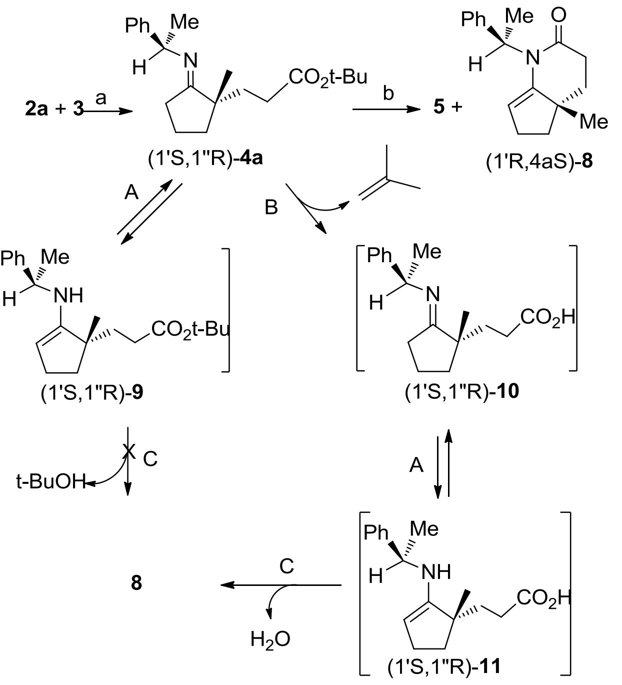


Figure 1. ^{13}C-{^{1}H} signals of carbon atom C-3' in ketoester **5** prepared in racemic form (**a**) and (*S*)-enriched one (**b**, **c**).

Analysis of NMR results obtained for synthesis of adduct **5a** using Δ clearly indicates a sharp decrease in the stereoselectivity (>95 to 80% ee) when the temperature was increased from 100 °C to 200 °C (Table 1, entries 3-4 and 6-7). This phenomenon can tentatively be explained assuming a possible competitive retro-Michael addition leading to a partial racemization of the Michael adduct under rather drastic conditions.[15]

Having secured an access to Michael adduct **5a**, a chemical correlation of its corresponding methyl keto ester **6**[10] was undertaken to ascertain the absolute configuration in **5a** (Scheme 1). Thus, keto acid **6** was easily obtained upon formic acid treatment of ketoester **5a** (Table 1, entry 3) in good yield and directly subjected to diazomethane esterification leading to (-)-(*S*)-**7**. As expected, the sense of induction is in accordance to the empirical rule defined for this reaction,[2,3] with the same sense of asymmetric induction as those obtained using methyl acrylate as the electrophile. This stereoselectivity originates from the *aza-ene* type mechanism in which internal transfer of the proton born by the nitrogen atom of the more substituted tautomeric enamine is concerted with the creation of the C-C bond (Scheme 1).

We then turned our focus to the study of this Michael reaction, carrying out the experiments under μW. As can be seen from Table 2 and Scheme 3, μW affects the reactivity, the stereo- and unexpectedly the chemoselectivity (Table 2). The closed vessel system was chosen in order to contain the toxic and volatile Michael acceptors in the reaction vessel, and to monitor the possible extent of pressure elevation during the microwave irradiation. The possibility of running reactions in an inert gas atmosphere is another distinct advantage with the sealed reaction vessel strategy. Despite sensitivity of chiral imines **2** toward water, this was not necessary in this case. The first observation was the role of stirring upon efficiency of the reaction. This parameter was found to be critical to the success of the reaction (compare entries 1,3,5 with entries 2,4,6 and 7).[16]



Scheme 2 Reaction pathway to lactam **8** under μW. *Reagents and conditions*: (a) 2 equiv. **3**, μW, 200 °C, 30 min; (b) 20% aq. AcOH, THF, 20 °C.

When mixtures of imine **2a** and alkene **3** were submitted to μW for 30 min at 100 °C with an optimal power of 30 W (Table 2, entry 2), the only reaction product was the expected chiral imine **4a**, leading to keto-ester **5a** upon hydrolytic workup. This result is a noteworthy improvement over results obtained from conventional heating at the same temperature for 4 hours (Table 1, entry 3) in terms of yields and reaction time, the enantioselectivity remaining the same.

At 150 °C, within 30 min, whereas yields were nearly comparable (Table 1, entry 6 and Table 2, entry 4), the enantioselectivity was largely improved under μW when compared to Δ (see Figure 1) with similar set of conditions (temperature, pressure, profiles of heating rates).

Table 2. Effect of μW irradiation and stirring upon condensation of imine **2a** (entries 1-7) and **2b** (entries 8-9) to *tert*-butyl acrylate **3**[a].

Entry	P (W)	T (°C)	Stirring	ΔP (bar)	**5a/5b** yield (ee) %	**8**
1	30	100	no	0.1	85	0
2	30	100	yes	0.1	100 (> 95[d])	0
3	80	150	no	0.7	67	0
4	80	150	yes	0.3	89 (> 95[d])	0
5	80	200	no	3.5	11	25
6	80	200	yes	1.4	41	14
7	100	200	yes	13.0	0	53
8	80	100	yes	0.8	92 (> 95[d])	0
9	100	200	yes	14.0	0	68

[a]: 2 Equivalents; [b]: Isolated yield of purified keto-ester **5a** after 30 min µW irradiation and subsequent hydrolysis; [c]: Not determined; [d]: Determined by ^{13}C-{^{1}H} NMR in chiral liquid crystals.

The most intriguing feature concerned the production of lactam **8** under forcing µW conditions (Scheme 3; Table 2, entries 5-7) instead of keto ester **5a**, since this lactam **8** was never observed in the conventional thermal process (Δ). This very important specific µW effect appeared when the reaction was performed at 200 °C. In contrast with Δ where the expected Michael adduct **5a** was obtained (Table 1, entry 7), within 30 min under µW irradiation at 200 °C, quite surprisingly, the lactam **8** was formed as the sole product (Scheme 2; Table 2, entry 7).

This noticeable finding on chemoselectivity can be justified by considering the possibility for µW to favor a very polar mechanism consisting in the nucleophilic addition (Scheme 2, routes C) of the enamines **9** or **11** to the carbonyl group of either a *tert*-butyl ester (**9** → **8)** with the release of *tert*-butanol, or an acid (**11** → **8**) with elimination of a water molecule. Secondary enamines **9** or **11** are in tautomeric equilibrium (Scheme 2, routes A) with imines **4a** or **10** respectively, the latter being issued from the thermolysis of the *tert*-butyl ester group in imine **4** with concomitant generation of isobutene (Scheme 2, route B, **4** → **10**).

In order to elucidate the pathway to lactam **8**, a GC-mass analysis of the headspace of the reaction mixture was undertaken, and compared to those of the starting chiral imine **2**, the acrylate **3** and *tert*-butanol having been separately irradiated under the same conditions (200 °C, 100 W, 30 min in closed reaction vessels). While isobutene was detected in the headspace of the reaction mixture and the *tert*-butyl acrylate **3** sample, it was not present in the *tert*-butanol or starting imine **1** ones.[17] This indicates that, under µW at 200 °C, the formation of lactam **8** occurred via a tandem Michael addition/deprotection/aza-annulation sequence implying the thermolysis of the *tert*-butyl ester group in imine **4** (path B/C).

Dealing with µW effect on enantioselectivity, the superiority of µW reveals the intervention of non-purely thermal µW specific effects. Although the effects observed in microwave-irradiated chemical transformations can in most cases be rationalized by purely bulk thermal phenomena associated with rapid heating to elevated temperatures,[18] we have conducted all experiments in the same conditions (closed vessels, same scale, same magnetic barrel) either in the microwave chamber or by immersion in a preheated oil bath in order to avoid as possible any difference in the temperature profiles. They can be justified by considering the reaction mechanism,[4] expecting µW effects when the polarity of the system increases during the reaction progress. It will be the case when the transition state (TS) of a reaction is more polar than its ground state, thus leading to a decrease in the activation energy.[19] Data obtained for the transition states for the *Re* and *Si*-approaches of imine **2** and *tert*-butyl acrylate **3** are consistent with the previous studies:[4b] *Re* approach: forming CC bond: 1.87 Å and forming CH bond: 2.56 Å; Si approach: forming CC bond: 1.89 Å and forming CH bond: 2.50 Å (Figure 2). As it was shown that this Michael addition proceeds through a concerted *asynchronous* '*aza-ene* like' mechanism,[4] the TS has thus a certain polarity, a situation for which µW effects are expected. This fact explains the intervention of µW effects upon exaltation of reactivity (comparing yields at 100 °C). To support this assumption, taking into account previous related AM1 computational investigations,[20] we calculate the corresponding approaches between the enamine tautomer of imine **2** with *tert*-butyl acrylate **3**. The *Re*-approach (the favored one) leads to a slightly more polar as well as more asynchronous than the *Si*-approach TS (Figure 2).[21] Consequently, this TS will be slightly favored due to a better dipole-dipole stabilization by µW. Therefore, under µW, the selectivity in favor of the *Re*-approach will be even more improved (exp. from 80 to > 95% ee). Reduced reversibility of the Michael reaction under microwave conditions could also contribute to the superior enantioselectivity, both phenomenon leading to an increased stereoselectivity in such conditions.

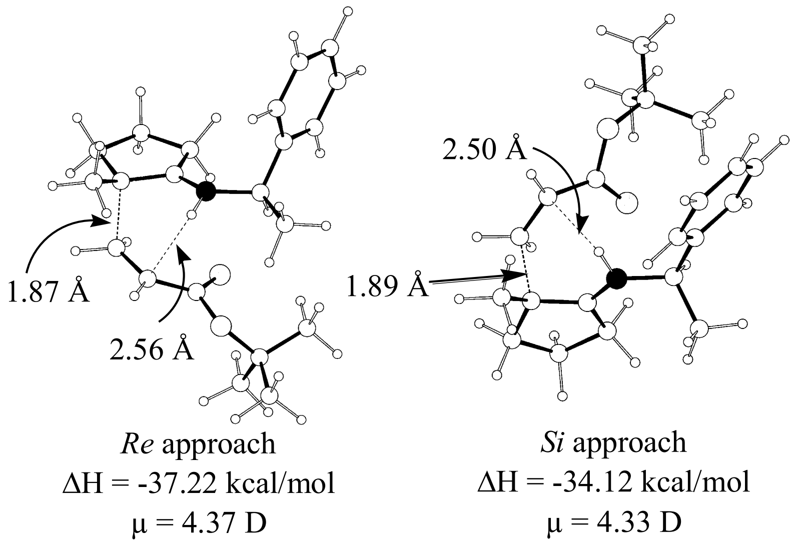
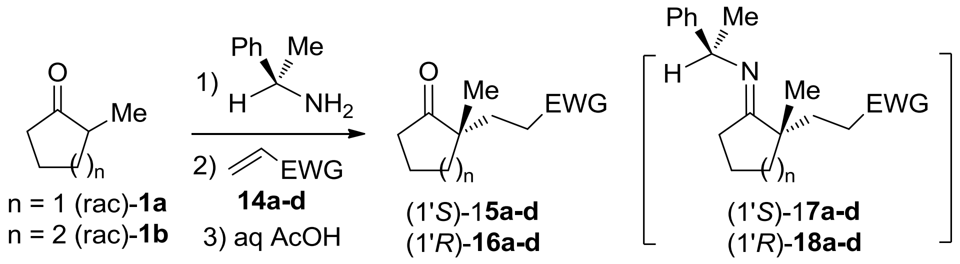
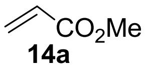
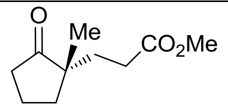
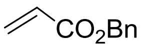
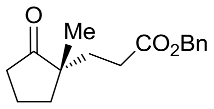
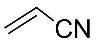
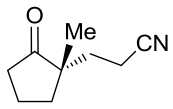
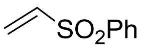
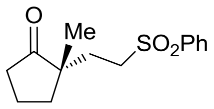
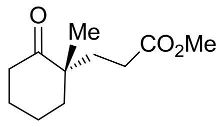
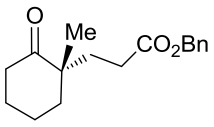
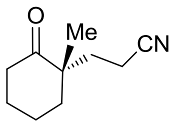


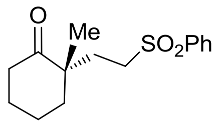
Figure 2. Transition structures at the RHF AM1 level for *Re* (left) and *Si* (right) approaches of *tert*-butyl acrylate **3** to enamine imine **2**, their respective enthalpies of formation and dipole moment.[20]

The same trend was observed with chiral imine **2b** derived from 2-methylcyclohexanone **1b**. Lactame **8b** was obtained when the reaction was performed at 200 °C upon irradiation at 100 W for 30 min (Table 2, entry 9) in a 68% yield while the expected Michael adduct 5b was obtained in the optimized conditions (100 °C, 80 W, Table 2, entry 8) in 92% yield with >95% ee. We next examine the reactivity of these imine in this µW promoted AMACI (Table 3).

Table 3. Effect of temperature on the synthesis of ketoester **4** by condensation of chiral imine **1** with electrophilic alkenes **12a-d**.[a]



Entry	Electrophile	Ketone	Michael adduct	Yield %[b] ee%
1	 **14a**	**1a**	 **7a**	87 [(d)] > 95
2	 **14b**	**1a**	 **15b**	76 [(d)] > 95
3[b]	 **14c**	**1a**	 **15c**	47 [(e)] > 95
4	 **14d**	**1a**	 **15d**	95 [(b)] >95
5	**14a**	**1b**	 **7b**	77 [(d)] > 95
6	**14b**	**1b**	 **16b**	60 [(c)] > 95
7	**14c**	**1c**		74 [(c)] > 95

			16c	
8	**14d**	**1d**	 **16d**	33 (c) > 95

[a]: 2 Equivalents; [b]: Isolated yield of purified keto-esters; (c) 80 W, 150 °C, 30 min.; (d) 80 W, 100 °C, 30 min.; (e) 80 W, 100 °C, 15 min.

The reaction performed well, giving the expected Michael adducts in adduct in 33-95% yield, in a first series of experiments. Ee of the adducts were measured at the level of the crude imines as >95% (no diastereoisomers detected).

With these satisfactory results in hand, we will turn to the study of a catalytic version of this reaction, in the context of a greener generation of quaternary carbon centers in a simple manner. Work is in progress to extend this µW activation mode to engage substituted acceptors in the AMACI.

3. Materials and Methods

3a. Chemistry: General: All reactions not involving aqueous media were carried out under a nitrogen atmosphere in a flame-dried glassware. Commercial reagents were used without further purification. Reactions were followed by ^{1}H NMR in $CDCl_3$ or using thin-layer chromatography, carried out on silica gel plates, which were viewed by UV irradiation at 254 nm and/or by staining with phosphomolybdic acid or *p*-anisaldehyde. Flash column chromatography was performed with 230-400 mesh silica gel. Melting points were recorded on a digital melting point apparatus. IR spectra were recorded with a Fourier transform spectrometer Bruker VECTOR 22. NMR spectra of the crude reaction mixtures were recorded in $CDCl_3$ containing a pinch of sodium carbonate in order to prevent hydrolysis of the imines. Imines proved to be stable for days in such conditions. ^{1}H NMR spectra were recorded at 300 K, at 200 or 400 MHz on a Bruker AC 200 or Bruker Avance 400 spectrometer, with $CHCl_3$ as internal standard (δ_H = 7.26 ppm). ^{13}C NMR spectra were recorded at 300 K, at 50 or 100 MHz, with the central peak of $CHCl_3$ as internal standard (δ_C = 77.0 ppm, central line). Recognition of methyl, methylene, methine and quaternary carbon nuclei in ^{13}C NMR spectra rests on the *J*-modulated spin-echo sequence. 2Dl NMR experiments (COSY, HMQC, HMBC and NOESY) were used for the assignments of signals in the ^{1}H and ^{13}C NMR spectra. For NMR measurements, ^{13}C 1D NMR experiments in polypeptidic oriented solvents were performed on a Bruker DRX-400 equipped with a BBO probe, and hence no additional hardware equipment is basically required. All proton-decoupled ^{13}C NMR experiments were recorded by applying the WALTZ-16 composite pulse sequence to decouple protons and benefit from NOE effect.

For unambiguous assignment of enantiomers in chiral NMR, comparison was made in all cases with the corresponding racemates. Optical rotations were measured at 589 nm in a 1 dm-cell using an Optical Activity Limited AA-10R apparatus and are expressed in g/100mL. Elemental analyses were performed by the Service de microanalyse, BioCIS, Châtenay-Malabry, France, with a Perkin-Elmer 2400 analyzer. A CEM Discover monomode reactor with an accurate control of temperature and pressure by modulation of emitted W was used for the microwave experiments. **Caution:** It is essential that great precaution be taken when carrying out organic reactions in sealed vessels. In particular, safety devices are to be used including appropriate septa as a pressure relief system and an automatic cut off of the microwave irradiation before the pressure limit of the vessels has been reached. See: Raner, K. D.; Strauss, C. R.; Trainor, R. W.; Thorn, S. J. *J. Org. Chem.* **1995**, *60*, 2456 and references cited therein. In this study, the maximum developed pressure was 14 bar at 100 W and 200 °C, far below the pressure limit (20 bar).

Preparation of chiral imines **2** exemplified for **2a**: In a 100 mL round bottom flask equipped with a Dean-Stark apparatus, 21.6 mL (0.2 mol) 2-methylcyclopentanone, 27.3 mL (0.21 mol, 1.05 equiv.) (*R*)-1-phenylethylamine (ee = 99%) are successively added to cyclohexane (50 mL). The resulting mixture was stirred for 18 h under nitrogen at 110 °C (oil bath) with azeotropic removal of water. Cyclohexane was then distilled and fractional distillation of the crude under reduced pressure afforded the desired chiral imine **1** as a colourless oil.

(1’*R*)-(2-Methylcyclopentylidene)-(1’-phenyl-ethyl)-amine (2a). Colorless oil (89%); B.p. = 80 °C (0.01 Torr); IR (neat, ν cm^{-1}): 2958, 1673; ^{1}H NMR (200 MHz, $CDCl_3$) δ ppm: 7.33-7.05 (m, 5H), 4.43-4.28 (m, 1H), 2.45-1.15 (m, 7H), 1.43 and 1.41 (d, *J* = 6.1 Hz, 3H), 1.11 and 1.10 (d, *J* = 6.7 Hz, 3H); ^{13}C NMR (50 MHz, $CDCl_3$) δ ppm: 181.0 and 180.6 (C), 146.2 and 145.9 (C), 128.1 (2CH), 126.5 (CH), 126.3 (2CH), 61.5 and 61.2 (CH), 41.2 and 41.1 (CH), 32.8 and 32.7 (CH_2), 28.7 and 28.5 (CH_2), 24.8 and 24.5 (CH_3), 22.5 and 22.4 (CH_2), 17.5 (CH_3).

General procedure for the asymmetric Michael reactions using microwaves: Mixtures of imines **2** (1 to 6 mmol) and 2 equivalents of electrophilic alkene were placed in sealed,[33] 10 mL heavy-walled pyrex tubes.[34] The tubes were introduced in the cavity of a single-mode[35] device allowing control of irradiation power (up to 300 W), time, temperature and pressure (see article, Tables 1 and 2).[36] The tube was opened after the reaction mixture was rapidly cooled down to room temperature, and excess *tert*-butyl acrylate was removed in vacuo. An aliquot of crude reaction mixture was analyzed by ^{1}H and ^{13}C NMR. For the crude reaction mixtures containing alkylated imine, after being vigorously stirred with 20% aqueous acetic acid (2 mL/mmol) and THF (2 mL/mmol) for 17 hours, the reaction mixture was concentrated, then thoroughly extracted with Et_2O (3 x 10 mL). The combined organic phase was washed successively with saturated $NaHCO_3$ and NaCl solutions, dried ($MgSO_4$), filtered over Celite® and concentrated in vacuo. Chromatographic purification on silica gel (cyclohexane:ethyl acetate, 9:1) afforded keto derivatives **5** and/or lactam **8** as colorless oils.

***tert*-Butyl (1’*S*,1”*R*)-3-[1’-Methyl-2-(1”-phenyl-ethyl-imino)-cyclopentyl]-propionate 4a.** IR (neat, ν cm^{-1}): 2966, 2868, 1726, 1673; ^{1}H NMR (200 MHz, $CDCl_3$) δ ppm: 7.38-7.07 (m, 5H), 4.35 (q, 1H, *J* = 6.6 Hz), 2.37-2.04 (m, 4H), 1.86-1.41 (m, 6H), 1.38 (s, 9H), 1.36 (d, 3H, *J* = 6.6 Hz), 0.98 (s, 3H); ^{13}C NMR (50 MHz, $CDCl_3$) δ ppm: 180.7 (C), 173.5 (C), 146.2 (C), 128.0 (2CH), 126.3 (2CH), 126.1 (CH), 79.7 (C), 60.8 (CH), 45.5 (C), 37.0 (CH_2), 33.4 (CH_2), 30.9

(CH_2), 28.4 (CH_2), 28.0 (3 CH_3), 24.8 (CH_3), 23.9 (CH_3), 20.5 (CH_2).

***tert*-Butyl (1'*S*,1"*R*)-3-[1'-Methyl-2-(1"-phenyl-ethyl-imino)-cyclohexyl]-propionate 4b.** IR (neat, ν cm^{-1}): 3061, 3063, 2966, 2926, 1658, 1493, 1447; ^1H NMR (300 MHz, $CDCl_3$) δ ppm: 7.41-7.12 (m, 5H), 4.69 (q, 3H, *J* = 6.4 Hz), 2.40-2.21 (m, 1H), 1.57-1.30 (m, 8H), 1.17-1.11 (2d, 3H, *J* = 4.3 Hz), 1.01-0.99 (d, 3H, *J* = 4,9 Hz); RMN ^{13}C (75 MHz, $CDCl_3$) δ ppm: 173.1 (C), 146.7 (C), 127.9 (CH), 126.3 (CH), 126.2 (CH), 125.9 (CH), 125.8 (CH), 57.3 (CH), 42.0 (CH_2), 35.6 (CH), 27.5 (CH_2), 25.4 (CH_3), 25.3 (CH_2).

***tert*-Butyl (1'*S*)-3-(1'-Methyl-2'-oxocyclopentyl)-propionate 5a.** IR (neat, ν cm^{-1}): 2968, 2934, 2872, 1727; ^1H NMR (200 MHz, $CDCl_3$) δ ppm: 2.38-2.06 (m, 4H), 2.0-1.54 (m, 6H), 1.36 (s, 9H), 0.98 (s, 3H); ^{13}C NMR (50 MHz, $CDCl_3$) δ ppm: 222.3 (C), 172.6 (C), 80.1 (C), 47.4 (C), 37.3 (CH_2), 35.9 (CH_2), 31.3 (CH_2), 30.5 (CH_2), 27.9 ($3CH_3$), 21.3 (CH_3), 18.4 (CH_2); $[\alpha]_D^{24}$ = -27.8 (c = 4.5, $EtOH_{abs}$); Anal. calcd for $C_{13}H_{22}N$: C, 68.99; H, 9.80. Found: C, 68.84; H, 9.75%.

***tert*-Butyl (1'*S*)-3-(1'-Methyl-2'-oxocyclohexyl)-propionate 5b.** IR (neat, ν cm^{-1}): 2933-2866, 1727, 1704; RMN ^1H (200 MHz, $CDCl_3$) δ ppm: 2.49-2.33 (2H), 2.28-2.12 (1H), 2.10-1.94 (3H), 1.92-1.52 (8H), 1.42 (9H), 1,02 (3H); RMN ^{13}C (75 MHz, $CDCl_3$): δ ppm: 215.7 (C), 173.1 (C), 82.3 (C), 47.9 (C), 39.3 (CH_2), 38.7 (CH_2), 32.4 (CH_2), 30.3 (CH_2), 28.7 ($3CH_2$), 27.4 (CH_2), 22.4 (CH_3), 21.0 (CH_2).

(4a*S*,1'*R*)-4a-Methyl-1-(1'-phenylethyl)-1,3,4,4a, 5,6-hexahydro-[1]pyrindin-2-one 8. IR (neat, ν cm^{-1}): 1665, 1629; ^1H NMR ($CDCl_3$, 200 MHz) δ ppm: 7.24-7.06 (m, 5H), 6.14 (q, 1H, *J* = 7.1 Hz), 4.31 (t, 1H, *J* = 2.5 Hz), 2.62-2.53 (m, 1H), 2.57 (dd, 1H, *J* = 8.3, 4.0 Hz), 2.30 (m, 1H), 2.05 (ddd, 1H, *J* = 15.6, 9.0, 3.1 Hz), 1.76-1.43 (m, 4H), 1.55 (d, 3H, *J* = 7.1 Hz), 1.05 (s, 3H); ^{13}C NMR ($CDCl_3$, 50 MHz) δ ppm: 169.3 (C), 143.7 (C), 141.2 (C), 128.3 (2CH), 126.5 (CH), 126.1 (2CH), 105.7 (CH), 49.9 (CH), 43.5 (C), 38.3 (CH_2), 33.4 (CH_2), 29.9 (CH_2), 28.5 (CH_2), 21.1 (CH_3), 14.8 (CH_3); Anal. calcd for $C_{17}H_{21}NO$: C, 79.96; H, 8.29; N, 5.49. Found: C, 80.04; H, 8.56; N, 5.15%.

3-[1-Methyl-2-(1-phenyl-ethylimino)-cyclopentyl]-propionitrile 17c. IR (neat, ν cm^{-1}): 3027, 2962, 2867, 2245, 1672; ^1H NMR ($CDCl_3$, 300 MHz) δ : 7.35-7.17 (m, 5H), 4.41 (q, 1H, *J* = 6.6 Hz), 2.65-2.27 (m, 4H), 1.94-1.57 (m, 6H), 1.41 (d, 3H, *J* = 6.6, Hz), 1.06 (s, 3H); ^{13}C NMR ($CDCl_3$, 75 MHz) δ ppm: 179.9 (C), 145.8 (C), 128.2 (2CH), 126.4 (CH), 126.3 (2CH), 120.7 (C), 61.2 (CH), 45.4 (C), 37.1 (CH_2), 35.8 (CH_2), 28.5 (CH_2), 24.9 (CH_3), 23.8 (CH_3), 20.5 (CH_2), 12.4 (CH_2).

3-[1-Methyl-2-(1-phenyl-ethylimino)-cyclohexyl]-propionitrile 18c. IR (neat, ν cm^{-1}): 2969, 2931, 2866, 2247, 1705, 1650; ^1H NMR ($CDCl_3$, 300 MHz) δ ppm: 7.39-7.20 (m, 5H), 4.70 (q, 1H, *J* = 6.6 Hz), 2.53-2.14 (m, 5H), 2.13-1.81 (m, 1H), 1.74-1.27 (m, 6H), 1.36 (d, 3H, *J* = 6.6, Hz), 1.04 (s, 3H); ^{13}C NMR ($CDCl_3$, 75 MHz) δ ppm: 172.1 (C), 146.6 (C), 128.2 (2CH), 126.4 (3CH), 121.1 (C), 57.8 (CH), 43.2 (C), 39.0 (CH_2), 35.2 (CH_2), 27.1 (CH_2), 25.8 (CH_3), 24.9 (CH_2), 23.9 (CH_3), 21.2 (CH_2), 12.4 (CH_2).

Methyl (1'*S*)-3-(1'-Methyl-2'-oxocyclohexyl)-propionate 7b. IR (neat, ν cm^{-1}): 2960, 1731, 1437; ^1H NMR (200 MHz, $CDCl_3$) δ ppm: 3.33 (s, 3H), 2.74-2.18 (m, 4H), 1.98-1.62 (m, 6H), 1.01 (s, 3H); ^{13}C NMR (50 MHz, $CDCl_3$) δ ppm: 222.3 (C), 173.8 (C), 51.5 (CH_3), 47.4 (C), 37.3 (CH_2), 35.9 (CH_2), 32.0 (CH_2), 29.2 (CH_2), 21.2 (CH_3), 18.4 (CH_2); Anal. for $C_{10}H_{16}O_3$, calcd. C, 65.19; H, 8.75; found C, 65.08; H, 8.79.; **12a** $[\alpha]_D^{20}$ -34.5 (c = 2; $EtOH_{abs}$); lit.[38] *ent*-**12a** $[\alpha]_D^{20}$ +35.7 (c = 2, $EtOH_{abs}$).

3-(1-Methyl-2-oxo-cyclopentyl)-propionitrile 15c. IR (neat, ν cm^{-1}): 2965, 2927, 2872, 2247, 1731; ^{1}H NMR (300 MHz, $CDCl_3$) δ ppm: 2.46-2.14 (m, 4H), 1.97-1.69 (m, 6H), 1.02 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ ppm: 221.2 (C), 119.7 (C), 47.2 (C), 37.2 (CH_2), 35.7 (CH_2), 32.0 (CH_2), 20.9 (CH_3), 18.4 (CH_2), 12.4 (CH_2); $[\alpha]_D^{20}$ = -34.6° (c = 0.003, $EtOH_{abs.}$); Anal. Calcd for $C_9H_{13}NO$: C, 71.49; H, 8.67; N, 9.26; O, 10.58. Found: C, 71.06; H, 8.12; N, 9.07.

3-(1-Methyl-2-oxo-cyclohexyl)-propionitrile 16c. IR (neat, ν cm^{-1}): 2937, 2867, 2246, 1700, 1451; ^{1}H NMR (300 MHz, $CDCl_3$) δ ppm: 2.48-2.37 (m, 1H), 2.33-2.26 (m, 3H), 1.97-1.64 (m, 8H), 1.12 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ ppm: 214.3 (C), 120.0 (C), 47.6 (C), 38.5 ($2CH_2$), 33.6 (CH_2), 27.1 (CH_2), 22.2 (CH_3), 20.8 (CH_2), 12.3 (CH_2); Anal. for $C_{10}H_{15}NO$, calcd. C, 72.69; H, 9.15; found C, 72.23; H, 8.49; MS (ESI): 166 (M+1) **16c** $[\alpha]_D^{24}$ 9.7 (c = 0.02, $EtOH_{abs}$).

[2-(2-Benzenesulfonyl-ethyl)-2-methyl-cyclopentylidene]-(1-phenyl-ethyl)-amine 17d IR (neat, ν cm^{-1}): 2962, 1671, 1447, 1305, 1145; ^{1}H NMR (300 MHz, $CDCl_3$) δ ppm: 7.70-7.45 (m, 10H), 4.34 (q, 1H, *J* = 6.6 Hz), 3.37 (m_c, 2H), 2.24 (bt, 2H, *J* = 5.9 Hz), 1.80-1.40 (m, 6H), 1.31 (d, 3H, *J* = 6.6 Hz), 1.00 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ ppm: 179.6 (C), 146.0 (C), 138.3 (CH), 133.6 (CH), 129.2 (2CH), 129.1 (2CH), 128.1 (CH), 127.7 (CH), 126.2 (CH), 67.8 (CH_2), 61.1 (CH), 52.2 (C), 37.6 (CH_2), 28.3 (CH_2), 25.5 (CH_2), 25.0 (CH_3), 23.8 (CH_3), 20.4 (CH_2).

[2-(2-Benzenesulfonyl-ethyl)-2-methyl-cyclohexylidene]-(1-phenyl-ethyl)-amine 18d IR (neat, ν cm^{-1}): 2928, 1707, 1650; ^{1}H NMR (300 MHz, $CDCl_3$) δ ppm: 8.00-7.87 (m, 4H), 7.66-7.49 (m, 6H), 4.61 (q, 1H, *J* = 6.6 Hz), 3.35-3.20 (m, 2H), 2.42-2.29 (m, 2H), 2.18-2.00 (m, 2H), 1.90-1.38 (m, 6H), 1.24 (d, 3H, *J* = 6.6 Hz), 1.01 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ ppm: 172.2 (C), 146.6 (C), 139.3 (C), 133.4 (CH), 129.1 (CH), 128.2 (CH), 126.4 (CH), 57.7 (CH), 52.2 (CH_2), 43.0 (C), 39.3 (CH_2), 31.9 (CH_2), 27.0 (CH_2), 25.7 (CH_3), 24.7 (CH_2), 24.2 (CH_3), 21.2 (CH_2).

2-(2-Benzenesulfonyl-ethyl)-2-methyl-cyclopentanone 15d Mp = 69 °C; IR (neat, ν cm^{-1}): 2965, 2870, 1729; ^{1}H NMR (300 MHz, $CDCl_3$) δ ppm: 7.87 (bd, 2H, *J* = 8.3 Hz), 7.68-7.45 (m, 3H), 3.14 (ddd, 1H, *J* = 18.4, 12.0, 4.7 Hz), 2.96 (ddd, 1H, *J* = 18.4, 12.2, 4.9 Hz), 2.35-2.07 (m, 2H), 1.91-1.72 (m, 6H), 0.94 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ ppm: 221.4 (C), 138.7 (C), 133.7 (CH), 129.2 (2CH), 127.9 (2CH), 51.8 (CH_2), 46.7 (C), 37.2 (CH_2), 36.2 (CH_2), 28.9 (CH_2), 21.0 (CH_3), 18.4 (CH_2); $[\alpha]_D^{23}$ = -19.4° (c = 0.0075, $EtOH_{abs.}$); MS (APCI): m/z 267 (100%) $[M + H]^+$; Anal. Calcd for $C_{14}H_{18}O_3S$: C, 63.13; H, 6.81; O, 18.02; S, 12.04. Found: C, 62.65; H, 6.80.

2-(2-Benzenesulfonyl-ethyl)-2-methyl-cyclohexanone 16d Mp = 74-77 °C; IR (neat, ν cm^{-1}): 2934, 2868, 1700; ^{1}H NMR (300 MHz, $CDCl_3$) δ ppm: 7.86 (bd, 2H, *J* = 8.3 Hz), 7.65-7.49 (m, 3H), 3.06 (ddd, 1H, *J* = 13.7, 11.9, 5.0 Hz), 2.98 (ddd, 1H, *J* = 13.7, 11.7, 5.0 Hz), 2.39-2.13 (m, 2H), 2.00-1.88 (m, 1H), 1.79-1.55 (m, 7H), 1.01 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ ppm: 214.3 (C), 138.9 (C), 133.6 (CH), 129.2 (2CH), 127.9 (2CH), 51.8 (CH_2), 47.4 (C), 38.9 (CH_2), 38.4 (CH_2), 30.2 (CH_2), 27.1 (CH_2), 22.3 (CH_3), 20.8 (CH_2); $[\alpha]_D^{24}$ = +2.9° (c = 0.01, $EtOH_{abs.}$); MS (APCI): m/z 281 (100%) $[M + H]^+$; Anal. Calcd for $C_{15}H_{20}O_3S$: C, 64.26; H, 7.19; O, 17.12; S, 11.44. Found: C, 64.40; H, 7.03.

General procedure for the synthesis of ketoacid 6

A mixture of adduct **6a** (452 mg, 2 mmol) and formic acid (2 mL) was stirred at 20 °C for 2 h. Formic acid was distilled, the crude was taken up in Et_2O (10 mL), washed with saturated $NaHCO_3$

solution (2 x 10 mL). The aqueous layer was acidified at 0 °C (6*N* HCl) and thoroughly extracted (4 x 10 mL Et_2O). The combined organic phase was dried ($MgSO_4$) and filtered (Celite®) and the crude concentrated in vacuo to give keto acid **6a** (302 mg, 89%) as a colorless oil. This material was used without further purification in the next step.

(1'*S*)-3-(1'-Methyl-2-oxocyclopentyl)-propionic acid 6a. IR (neat, ν cm^{-1}): 3512, 3090,2963, 2873, 2663, 1729, 1706; ^1H NMR (200 MHz, $CDCl_3$) δ ppm: 11.0 (bs, 1H), 1.67-1.33 (m, 4H), 1.19-0.81 (m, 6H), 0.20 (s, 3H); ^{13}C NMR (50 MHz, $CDCl_3$) δ ppm: 222.8 (C), 178.3 (C), 47.0 (C), 36.8 (CH_2), 35.4 (CH_2), 30.6 (CH_2), 28.7 (CH_2), 20.7 (CH_3), 18.0 (CH_2); Anal. for $C_9H_{14}O_3$, calcd. C, 63.51; H, 8.29; found C, 63.28; H, 8.39; $[\alpha]^{25}_D$ -40.6 (c = 1.6, $EtOH_{abs}$).

(1'*S*)-3-(1'-Methyl-2-oxocyclohexyl)-propionic acid 6b. IR (neat, ν cm^{-1}): 3502, 2935, 2866, 1700; ^1H NMR (300 MHz, $CDCl_3$) δ ppm: 9.4 (bs, 1H), 2.38-2.29 (m, 3H), 2.23-2.12(m, 1H), 2.03-1.93 (m, 1H), 1.84.68 (m, 6H), 1.62-158 (m, 1H), 1.05 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ ppm: 215.5 (C), 179.3 (C), 47.7 (C), 39.0 (CH_2), 38.5 (CH_2), 32.2 (CH_2), 28.9 (CH_2), 27.3 (CH_2), 22.3 (CH_3), 20.8 (CH_2); MS (ESI) 207 (M+23) 391 (2M+23); $[\alpha]^{28}_D$ -85 (c = 0.13, $EtOH_{abs}$).

Chemical correlation to Methyl (1'*S*)-3-(1-Methyl-2-oxo-cyclopentyl)-propionate 7a: A 0.5 M solution of diazomethane in Et_2O (5 mL) was added to a solution of the acid **6** (97 mg, 0.57 mmol) in dry Et_2O (20 mL) at 0 °C. The resulting mixture was stirred at room temperature for 2 hours and the excess of diazomethane was destroyed with acetic acid. The reaction mixture was washed with brine (2×5 mL), dried ($MgSO_4$), and concentrated in vacuo. The residue was purified by column chromatography (cyclohexane/AcOEt, 9:1) to give methyl ester **7a** as a colourless oil (100 mg, 95%); $[\alpha]_D^{20}$ -32.5 (c = 2; $EtOH_{abs}$); lit.[10] *ent*-**12a** $[\alpha]_D^{20}$ +35.7 (c = 2, $EtOH_{abs}$).

3b. Enantiomeric excess determination: For ^{13}C-{^1H} 1D NMR experiments in polypeptidic oriented solvents, the sample preparation consisted of directly weighting 25-30 mg of solute, around 140 mg of poly-γ-benzyl-L-glutamate (PBLG, DP= 463, commercially available) and adding about 350 mg of dichloromethane into a 5 mm NMR tube. Under these conditions, the total volume of the sample is optimal compared to the length of the coil of a 5 mm diameter probe-head. Compared to previous work, we have used a larger amount of PBLG than usual (100 mg) in order to obtain a clean liquid crystalline phase. This is a consequence of the relatively low degree of polymerization (DP= 463, i.e. MW ≈101000 $g.mol^{-1}$). The exact composition of each NMR sample is given in Table 3. To avoid the evaporation of dichloromethane during long NMR experimental time, we have sealed the samples. Note here that a mixture of protonated and deuterated dichloromethane was used. This solution allows to easily shim the magnet on the proton FID as well as to minimize the digitization problems associated with the dynamic range of the Analogue-to-Digital Converter (ADC) induced by the difference of ^{13}C signal intensity between dichloromethane and solute. In other hands, the shape and the line width of ^{13}C resonances of dichloromethane provide a serious control of the magnet stability as well as possible time-evolution of the sample homogeneity during the experiments. The sample is then centrifuged during few seconds, then inverted and centrifuged again. This process is repeated until an optically homogeneous birefringent phase is obtained.

Table 3. Composition of liquid-crystalline representative NMR samples investigated

Sample	Solute	Solute / mg[a]	Co-solvent / mg[a]	Polymer % in weight
1	***(rac)*-6**	27	75/275	27.1
2	**(*S*)-6**	28	75/277	26.9
3	**(*S*)-6**	28	75/277	27.1

Conditions: Polymeric solvent: PBLG; Degree of polymerization of polypeptide used: 463; Co-solvent: CH_2Cl_2 / CD_2Cl_2; *Polymer*/mg: 140; [a]The accuracy on the weighing is 1 mg.

The NMR tube was not spun in the magnet and its temperature was regulated carefully at 299 K using the standard variable temperature control unit (BVT 3000). ^{13}C spectra were recorded by adding 1000 to 3000 scans. Gaussian filtering was applied to improve the spectral separation of resonances. The area measurement was performed using a curve fitting algorithm based on complex least squares treatment of the ^{13}C NMR signals with and without filtering. Note that the experiments and the area measurements were repeated several times to estimate accurately the error on the enantiomeric excess of the mixture.

3c. GC-MS mechanistic studies: An HP 5989A GC-MS system (Hewlett-Packard, Palo Alto, CA, USA) was used. The chromatographic separation was performed with an Omega delta-3 capillary column (length: 25 m; I.D. 0.20 mm; film thickness: 0.2 µm) (Macherey-Nagel, Düren, Germany). In view of comparison, samples consisting of either the asymmetric Michael addition reaction mixture [1 equiv. chiral imine (**2a**) and 1.1 equiv. *tert*-butyl acrylate (**3**)] or each of the individual components (*tert*-butyl acrylate (**3**), chiral imine (**2a**), 2-methyl cyclopentanone **1a**, 1-phenylethylamine, 2-methyl-2-propanol) were separately irradiated at 100 W and 200 °C for 30 min in 10 mL teflon sealed glass vials and cooled to r.t. prior to GC-mass analysis. The teflon sealed glass vials filled with the reaction medium were maintained at 70°C during 15 minutes prior analysis and immediately processed. The head space (5 µL) was sampled using an airtight syringe and injected in splitless mode. Helium pressure was 50 kPa. The injector temperature was 250 °C and the initial oven temperature was 35 °C. This temperature was maintained for 1 min, the temperature was then programmed as follows: 4 °C/min up to 50 °C then 6 °C/min up to 100 °C followed by a 5 min hold. The transfer line temperature was set to 280 °C. Analysis was performed by electronic impact ionisation. The ion source and quadrupole temperature was set to 200 °C and 100 °C respectively. The electron energy was 70 eV. Acquisition was performed in scan mode over the range of 20 to 150 at a scan rate of 0.9 scan/sec (4 samplings per scan). Analysis of the results obtained for the head space of the asymmetric Michael reaction showed that the peak observed at 1.605 min correspond to 2-methyl-2-butene (MW = 56 g.mol^{-1}, identical fragmentation and comparable abundances)[17] A similar peak was not detected from the head space of the other samples, except for the *tert*-butyl acrylate (MW = 128 g.mol^{-1}) one. However, ions at m/z 55, 57 and 59 are not present in the same ratio for the 2-methyl-2-butene mass spectrum. Moreover, analysis of the *tert*-butanol (MW = 74 g.mol^{-1}) sample indicates that in the reaction conditions, *tert* butanol did not led to 2-methyl-2-butene. This set of results gave evidence that 2-methyl-2-butene and not *tert*-butanol was released during the lactamization process of Michael adduct (**5**) (see article, Scheme 3).

3d. Therometical calculations: Geometries for the reactants were optimized by means of gradient technique at RHF AM1 level[20] by using the semi-empirical molecular orbital program MOPAC.[21] All the RHF AM1 transition structures were located using the procedures implemented in MOPAC (Version 5.0). All variables were optimized by minimizing the sum of the squared scalar gradients (NLLSQ and SIGMA).[22,23] Force calculations were carried out to ensure that the transition structures located had one imaginary frequency. Final values of the gradient norms were <1 kcal/Å and each transition structure had one negative eigenvalue in the Hessian matrix as required.

.

4. Conclusions

In conclusion, we have demonstrated that the reaction of hindered *tert*-butyl acrylate **2** in the AMACI was efficiently promoted under µW activation, compared to conventional heating. As expected, regardless the activation mode, the control of the stereochemistry of the newly created quaternary carbon center in such Michael adducts is always dictated by the configuration of the chiral inductor. The stereoselectivity of the process was determined using natural abundance ^{13}C-{^{1}H} NMR in a chiral polypeptide liquid crystal. Moreover, the temperature profiles achieved under microwave irradiation are not accessible in conventional heating and can allow a differentiation in the reaction pathways. A direct and stereoselective access to lactams **8** was thus achieved only under µW, although in moderate yield. A highly stereoselective process (ee > 95 %) was obtained either at 100 °C for 4 h (Δ) or for 30 min (µW, 100 W). A good correlation was obtained between experiment and theoretical calculations. Both the more polar and asynchronous transition state led to the expected Michael adduct (*S*)-**4**, and are favored under µW activation, allowing the reaction to proceed efficiently in minutes. Finally, Michael adducts from methyl acrylate, benzyl acrylate, vinylsulfone and acrylonitrile are regio- and stereoselectively obtained in high yield and short time using the microwave process.

Acknowledgments

We thank Dr Kathalyn L. Serre, Hanaa Tagzhouty and Kristine Aerola (BioCIS) for preliminary experiments, Sophie Mairesse-Lebrun for performing the microanalysis, CNRS and MENRS for financial support and CEM for technical support to A. L.

Author Contributions

Françoise Dumas ensure the conception and design of the chemistry and Philippe Lesot the the enantiomeric purity determination using natural abundance ^{13}C-{^{1}H} NMR. Lucie Vandromme, Li Chen and Lai Wei performed the chemical experiments and analyzed the data, Franck Le Bideau, and Françoise Dumas wrote the manuscript, André Loupy helped and advised us for the microwave chemistry, Olivier Lafon and Philippe Lesot were in charge of the NMR determination of selectivity in chiral liquid phase, Elise Tran Huu Dau performed the theoretical calculations and Pierre Chaminade the CPV analysis.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Denissova, I.; Barriault, L. Stereoselective formation of quaternary carbon centers and related fucntions. Tetrahedron report number 661, *Tetrahedron* **2003**, *59*, 10105-10146. (b) *Quaternary*

Stereocenters: Challenge and Solutions for Organic Synthesis (Eds.: J. Christoffers, A. Baro), Wiley-VCH, Weinheim, Germany, **2005**.

2. Reviews: (a) d'Angelo, J.; Desmaële, D.; Dumas, F.; Guingant, A. The asymmetric Michael addition reactions using chiral imines. *Tetrahedron: Asymmetry* **1992**, *3*, 459-505. (b) d'Angelo, J.; Cavé, C.; Desmaële, D.; Dumas, F. The Asymmetric Michael Addition Reactions Using Chiral Imines: Application to the synthesis of Compounds of Biological Interest. in *Trends in Organic Chemistry* Pandalai S. G. Ed.; Transworld Research Network, Trivandrum, India **1993**, volume 4, pp 555-616.
3. See for example: (a) Pizzonero, M.; Dumas, F.; d'Angelo, J. Enantioselective Synthesis of (*R*)-1-Azaspiro[4.4]nonane-2,6-dione Ethylene Ketal, Key Chiral Intermediate in the Elaboration of (-)-Cephalotaxine. *Heterocycles* **2005**, *66,* 31-37. (b) Kousara, M.; Ferry, A.; Le Bideau, F.; Serré, K. L.; Chataigner, I.; Morvan, I.; Dubois, J.; Chéron, M.; Dumas, F. First enantioselective total synthesis and configurational assignments of suberosenone and suberosanone as potential antitumor agents**.** *Chem. Commun.*, **2015**, *51*, 3458-3461. (c) Ito, F.; Ohbatake, Y.; Aoyama, S.; Ikeda, T.; Arima, S.; Yamada, Y.; Ikeda, H.; Nagamitsu, T.: Total Synthesis of (+)-Clavulatriene A. *Synthesis* **2015**, *47*, 1348-1355.
4. Sevin, A.; Tortajada, M., Pfau, M. Toward a transition-state model in the asymmetric alkylation of chiral ketone secondary enamines by electron-deficient alkenes. A theoretical MO study. *J. Org. Chem.* **1986**, *51*, 2671-2675. (b) Lucero, M. J.; Houk, K. N. Conformational Transmission of Chirality: The Origin of 1,4-Asymmetric Induction in Michael Reactions of Chiral Imines. *J. Am. Chem. Soc.* **1997**, *119*, 826-827. (b) Tran Huu Dau, M. E.; Riche, C.; Dumas, F.; d'Angelo, J. The origin of the stereoselectivity in the asymmetric Michael reaction using chiral imines/secondary enamines under neutral conditions: a computational investigation. *Tetrahedron: Asymmetry* **1998**, *9*, 1059-1064 and quoted references.
5. See interalia: (a) Camara, C.; Joseph, D.; Dumas, F.; d'Angelo, J.; Chiaroni, A. High pressure activation in the asymmetric Michael addition of chiral imines to alkyl and aryl crotonates *Tetrahedron Lett.* **2002**, *43,* 1445-1448. (b) Camara, C.; Keller, L.; Jean-Charles, K.; Joseph, D.; Dumas, F. A comparative study of high pressure versus other activation modes in the asymmetric Michael reaction of chiral imines. *Int. J. High Press. Res.* **2004**, *24*, 149-162.
6. (a) Perreux, L.; Loupy, A. Tetrahedron Report number 588, A tentative rationalization of microwave effects in organic synthesis according to the reaction medium, and mechanistic considerations. *Tetrahedron*, **2001**, *57*, 9199-9223. (b) Kappe, C. O. Controlled microwave heating in modern organic synthesis. *Angew. Chem. Int. Ed.* **2004**, *43*, 6250-6284. (c) De la Hoz, A.; Diaz-Ortiz, A.; Moreno, A. Microwaves in organic synthesis. Thermal and non-thermal microwave effects. *Chem. Soc. Rev.* **2005**, *34*, 164-178. (d) Microwave Heating as a Tool for Sustainable Chemistry; Leadbeater, N. E., Ed.; CRC Press: Boca Raton, FL, USA, 2011. (e) Kappe, C. O.; Stadler, A.; Dallinger, D. Microwaves in Organic and Medicinal Chemistry, 2nd ed.; Wiley-VCH: Weinheim, Germany, 2012. (f) Microwaves in Organic Synthesis, 3rd ed.; De La Hoz, A., Loupy, A., Eds.; Wiley-VCH: Weinheim, Germany, 2013.
7. (a) Loupy, A.; Perreux, P.; Liagre, M.; Burle, K.; Moneuse, M. Reactivity and selectivity under microwaves in organic chemistry. Relation with medium effects and reaction mechanisms. *Pure*

Appl. Chem., **2001**, *73*, 161-166. (b) De la Hoz, A.; Diaz-Ortiz, A.; Moreno, A. Selectivity in organic synthesis under microwave irradiation. *Current Org. Chem.* **2004**, *8*, 903-918. (c) Strauss, C. R.; Rooney, D. W, Accounting for clean, fast and high yielding reactions under microwave conditions. *Green Chem.*, **2010**, *12*, 1340-1344. (d) See for example: Yus, M.; Foubelo, F.; Jesús García-Muñoz, M. Stereoselective Aza-Henry Reaction of Chiral tert-Butanesulfinyl Imines with Methyl or Ethyl 4-Nitrobutanoate: Easy Access to Enantioenriched 6-Substituted Piperidine-2,5-diones. *Heterocycles* **2015**, *90*, 1419-1432.

8. (a) Langa, F.; De la Cruz, P.; De la Hoz, A.; Diaz-Ortiz, A.; Diez-Barra, E. Microwave irradiation: more than just a method for accelerating reactions. *Contemp. Org. Synth.* **1997**, *4*, 373-386. (b) Manhas, M. S.; Banik, B. K.; Mathur, A.; Vincent, J. E.; Bose, A. K. Vinyl-β-lactams as Efficient Synthons. Eco-friendly Approaches via Microwave Assisted Reactions. *Tetrahedron* **2000**, *56*, 5587-5601. (c) Camara, C.; Keller, L.; Dumas, F. Microwave activation of an asymmetric Michael reaction: unexpected behavior of chiral α-alkoxy imines. *Tetrahedron: Asymmetry* **2003**, *14,* 3263-3266. (d) Narasimhan, S.; Velmathi, S. Effect of Microwaves in the Chiral Switching Asymmetric Michael Reaction *Molecules*, **2003**, *8*, 256-262. (e) Escalante, J.; Díaz-Coutiño, F. D. Synthesis of γ-Nitro Aliphatic Methyl Esters Via Michael Additions Promoted by Microwave Irradiation. *Molecules* **2009***, 14,* 1595-1604. (f) Worzakowska, M. Thermal properties of neryl long-chain esters obtained under microwave irradiation. *J. Thermal Anal. Calor.* **2015**, 120, 1715-1722.
9. (a) Kang, J. Y.; Carter, R. G. Primary Amine, Thiourea-Based Dual Catalysis Motif for Synthesis of Stereogenic, All-Carbon Quaternary Center-Containing Cycloalkanones *Org. Lett.* **2012**, *14*, 3178-3181. (b) Horinouchi, R.; Kamei, K.; Watanabe, R.; Hieda, N.; Tatsumi, N.; Nakano, K.; Ichikawa, Y.; Kotsuki, H. Enantioselective Synthesis of Quaternary Carbon Stereogenic Centers through the Primary Amine-Catalyzed Michael Addition Reaction of α-Substituted Cyclic Ketones at High Pressure. *Eur. J. Org. Chem.* 2015, 4457-4463.
10. Pfau, M.; Revial, G.; Guingant, A.; d'Angelo, J. Enantioselective synthesis of quaternary carbon centers through Michael-type alkylation of chiral imines. *J. Am. Chem. Soc.* **1985**, *107*, 273-274.
11. 1-Phenylethylamine (99% ee) was used to direct the regio and stereoselectivity of the Michael addition and additionnaly served as a chirality marker in the product **4**, allowing determination of the diastereoselectivity of the reaction using NMR, and by consequence of the enantioselectivity of the Michael process.
12. Sarfati, M.; Lesot, P.; Merlet, D.; Courtieu, J. Theoretical and experimental aspects of enantiomeric differentiation using natural abundance multinuclear NMR spectroscopy in chiral polypeptide liquid crystals. *Chem. Commun.* **2000**, 2069-2081.
13. Meddour, A.; Berdagué, P.; Hedli, A.; Courtieu J.; Lesot, P. Proton-Decoupled Carbon-13 NMR Spectroscopy in a Lyotropic Chiral Nematic Solvent as an Analytical Tool for the Measurement of the Enantiomeric Excess. *J. Am. Chem. Soc*. **1997**, *119*, 4502-4508.
14. (a) Lesot, P.; Sarfati, M.; Courtieu, J.; Natural abundance deuterium NMR spectroscopy in polypeptide liquid crystals as a new and incisive means for the enantiodifferentiation of chiral hydrocarbons. *Chemistry* **2003**, *14*, 1724-1745. (b) Luy, B. Disinction of enantiomers by NMR spectroscopy using chiral orienting media, *J. Indian Inst. Sci.,* **2010**, *90,* 119-132. (c) Lesot, P.;

Aroulanda, C.; Zimmermann, H.; Luz, Z. Enantiotopic discrimination in the NMR spectrum of prochiral solutes in chiral liquid crystals. *Chem. Soc. Rev.*, **2015**, *44*, 2330-2375.

15. See in relation (a) Tan, K.; Alvarez, R.; Nour, M.; Cavé, C.; Chiaroni, A.; Riche, C.; d'Angelo, J. Racemization processes at a quaternary carbon center in the context of the asymmetric Michael reaction. *Tetrahedron Lett.* **2001**, *42*, 5021-5023. (b) Pizzonero, M.; Hendra, F.; Delarue-Cochin, S.; Tran Huu Dau, M.-E.; Dumas, F.; Cavé, C.; Nour, M.; d'Angelo, J. The asymmetric Michael-type alkylation of chiral β-enamino esters: critical role of a benzyl ester group in the racemization of adducts. *Tetrahedron: Asymmetry* **2005**, *16*, 3853-3857.
16. Design and evaluation of improved magnetic stir bars for single-mode microwave reactors Obermayer, D.; Damm, M.; Kappe, C. O. *Org. Biomol. Chem.* **2013**, *11*, 4949-4956.
17. Derivative spectra were compared with those in the NBS75K database (provided by Hewlett Packard with the GC/MS control and data processing software).
18. Kappe, C. O.; Pieber, B.; Dallinge, D. Microwave Effects in Organic Synthesis: Myth or Reality? Ang. Chem. Int. Ed. **2013**, *52*, 1088-1094.
19. Chemical quantum calculations related to experimental results have shown that µW effects are increasing with the asynchronous character of a mechanism. See for example: (a) Diaz-Ortiz, A.; Carrillo, J. R.; Cossio, F. P.; Gomez-Escalonilla, M. J.; De La Hoz, A.; Moreno, A.; Prieto, P. Synthesis of Pyrazolo[3,4-*b*]pyridines by Cycloaddition Reactions under Microwave Irradiation. *Tetrahedron* **2000**, *56*, 1569-1577. (b) Loupy, A.; Maurel, F.; Sabatie-Gogova, A. Improvements in Diels–Alder cycloadditions with some acetylenic compounds under solvent-free microwave-assisted conditions: experimental results and theoretical approaches. *Tetrahedron* **2004**, 60, 1683-1691. (c) Langa, F.; de la Cruz, P.; de la Hoz, A.; Espildora, E.; Cossio, P.; Lecea, B. Modification of Regioselectivity in Cycloadditions to C_{70} under Microwave Irradiation. *J. Org Chem*. **2000**, *65*, 2499-2507.
20. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. P. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.
21. Stewart, J. P. P. Quantum Chemistry Program Exchange, University of Indiana, Bloomington, U.S.A.; Program 455.
22. Bertels, R. H.; Report CNA-44, **1972**, University of Texas, Center for numerical analysis.
23. McIver, J. W.; Komornicki, A. Structure of transition states in organic reactions. General theory and an application to the cyclobutene-butadiene isomerization using a semiempirical molecular orbital method. *J. Am. Chem. Soc.* **1972**, *94*, 2625-2633.

SciForum
Mol2Net

Chiral Imines on the Wave: Reactivity of *tert*-Butyl Acrylate and Stereoselectivity Determination Using NMR in Liquid Crystals

Lucie VANDROMME [1] , Li CHEN [1] , Lai WEI [1] , Franck LE BIDEAU [1], André LOUPY [2] , Philippe LESOT [3] , Olivier LAFON [3] , Elise TRAN HUU DAU [4] Pierre CHAMINADE [5] and Françoise DUMAS [1,*]

[1] BioCIS, UMR CNRS 8076, Université Paris Saclay, School of Pharmacy, Université Paris-Sud, 5, rue J.-B. Clément, F-92296 Châtenay-Malabry, France; E-Mails: lucie.vandromme@free.fr (L. V.); li.chen1@u-psud.fr (L. C.); lai.wei@u-psud.fr (L. W.); franck.lebideau@u-psud.fr (F. L. B.); francoise.dumas@u-psud.fr (F.D.);

[2] LRSSS, UMR CNRS 8615, Université Paris Saclay, UFR de Sciences, Université Paris-Sud, ICMMO, Bât. 410, F-91405 Orsay, France; E-Mail: andre.loupy@cegetel.net (A.L.);

[3] Laboratoire de RMN en Milieu Orienté, CNRS UMR 8182, ICMMO, Bât. 410, Université Paris Saclay, UFR de Sciences, 91405 Orsay cedex, France. E-mail: philippe.lesot@u-psud.fr

[4] ICSN, UPR CNRS 2301,1 avenue de la Terrasse, F-91190 Gif sur Yvette, France; E-Mail: elise.tran@icsn.cnrs-gif.fr.

[5] Lip(Sys)2, EA 4041, Université Paris Saclay, School of Pharmacy, Université Paris-Sud, 5, rue J.-B. Clément, F-92296 Châtenay-Malabry, France ; Email : pierre.chaminade@u-psud.fr

* Author to whom correspondence should be addressed; E-Mail: francoise.dumas@u-psud.fr; Tel.: +33-146-835-563.

Published: 4 December 2015

Abstract: In connection with synthetic applications, we have foreseen to study the reactivity of the poorly reactive *tert*-butyl acrylate electrophile with chiral imines of unsymmetrical ketones, using conventional or microwave flash heating under carefully controlled reaction conditions in order to develop a selective and efficient access to the corresponding Michael adducts in which a created stereogenic quaternary carbon center was fully stereocontrolled. Depending on the conditions, either a keto ester or a lactam were obtained. A good correlation was obtained between experiment and theoretical calculations. The stereoselectivity of the process (e>95%) was determined using natural abundance ^{13}C-{^{1}H} NMR in a chiral polypeptide liquid crystal. The scope of the reaction was screened using a set of electrophilic alkenes giving Michael adducts in good yield and similar high enantiocontrol.

Keywords: Aza-ene process, Asymmetric Michael reaction, Enamine activation, Unsymmetrical ketones, *tert*-Butyl acrylate, Quaternary carbon center, chiral polypeptide, liquid crystal, ^{13}C-{^{1}H} NMR, MW assisted synthesis.

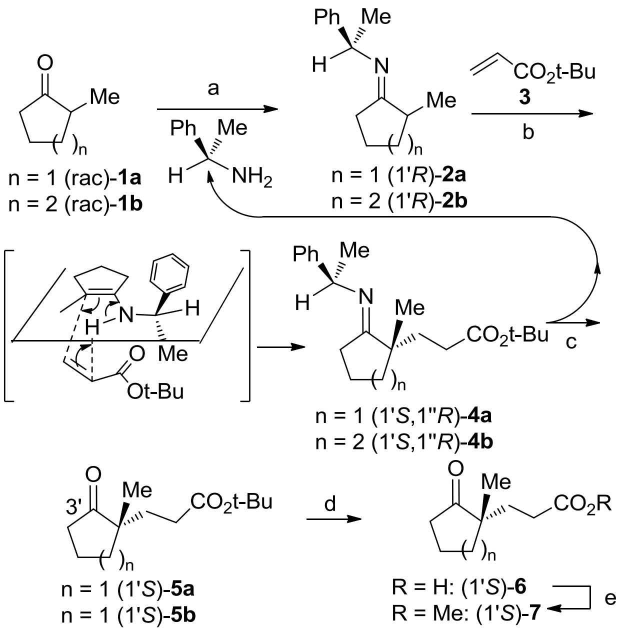
1. Introduction

Although a broad range of methods able to generate new carbon-carbon bonds exists, the establishment of quaternary carbon centres in the proper configuration is among the most restrictive in organic synthesis.[1] Since forty years, the asymmetric Michael addition of chiral imines (AMACI) under neutral conditions has attracted widespread attention as a versatile carbon-carbon bond-forming method, leading to Michael adducts with high levels of regio- and stereocontrol.[2] Such Michael adducts featuring a stereogenic tetrasubstituted carbon center are useful synthons for the asymmetric synthesis of a variety of bioactive compounds.[3] Besides its remarkable efficiency due to a concerted *aza-ene* type mechanism,[4] the reaction tolerates a large variability in both carbonyl compounds and Michael acceptors, with some limitations for hindered systems for which high pressure activation is required.[5] Moreover, in the context of improved reaction efficiency, clean processes and short reaction times are desired. In this respect, µW is an efficient way of promoting organic transformations, mainly in solvent-free systems. Thus, interest in µW assisted organic reactions has recently considerably increased.[6] Nevertheless, little attention has been devoted to the µW effects on selectivity,[7] particularly for asymmetric Michael reactions.[8] Herein, we describe our investigation of the reactivity of chiral imines with hindered tert-butyl acrylate and related electrophiles under microwave activation.

2. Results and Discussion

As a part of our program directed at exploring the scope of the AMACI, and in connection with synthetic applications for which an orthogonal ester protection cleavable in acidic medium was desired in the Michael adduct, we have foreseen to study the reactivity of *tert*-butyl acrylate **3** with chiral imine **2a** derived from 2-methyl-cyclopentanone **1a**. Although considerable attention has recently been focused on organocatalytic asymmetric transformations as efficient and convenient methodologies owing to their environmentally friendly characteristics, none of them address the reactivity of bulky electrophiles in this reaction.[9] Prior to engage in such studies, and because the chiral inductor is available at low price and easily recovered without loss of optical activity at the end of the process, we first analyze the stoichiometric transformation. Due to the steric hindrance of the acceptor, we have turned to use microwave irradiation (µW) under carefully controlled reaction conditions, in order to develop a simple and efficient access to Michael adduct **5a** and the derived keto-acid **6** and compared this method to conventional heating (Δ) (Scheme 1).

In general, the AMACI is carried out from the intermediary chiral imine and the electrophile, in the absence of solvent, at room temperature or using moderate Δ (up to 80 °C) as the regio- and stereocontrol of this reaction are only slightly sensitive to heat.[2] When the reaction was performed at 25 °C for 7 days, the conversion was not complete and 15% yield of Michael adduct **5a** was obtained upon hydrolysis (Table 1, entry 1). Thus, due to its bulkiness, *tert*-butyl acrylate **3** reacts much slower than its methyl counterpart.[10] At 60 °C, all the imine **2a** was consumed in one day; however, side polymerization reactions resulted in a modest 58% yield of adduct **5** (Table 1, entry 2).



Scheme 1 *Reagents and conditions*: (a) 1.01 eq 1-phenylethylamine, cyclohexane, reflux, Dean-Stark, overnight, 89% (b) 2 equiv. **2**, 25-200 °C (see Tables 1 and 2); (c) 20% aq. AcOH, THF, 20 °C (d) HCO_2H, 20 °C, 89%; (e) CH_2N_2, Et_2O, 0 to 20 °C, 2 h, 95%.

When the reaction was performed at 100 °C for 4 h, alkylated imine **4a**[11] was obtained as the sole product leading upon hydrolysis to keto ester **5** in 79% yield. On the basis of its ^{13}C NMR spectrum, crude imine **4a** exists as a single stereoisomer (de >95%),[11] as the result of a highly stereo-controlled process, giving rise to Michael adduct **5a** with a >95% enantiomeric excess (ee). The reaction duration was markedly reduced to 30 min at 150 °C and ketoester **5** was obtained in an optimum 92% yield upon hydrolysis. However, the stereoselectivity proved to be lower (ee 80%) than at 100 °C (Table 1, entry 6). Finally, both the efficiency and the stereoselectivity dropped at 200 °C (Table 1, entry 7).

.

Table 1. Effect of temperature on the synthesis of ketoester **4** by condensation of chiral imine **1** with *tert*-butyl acrylate **2**.^{a}

Entry	Temperature (°C)	Time	**5** yield %^{b}	**5** ee%^{d}
1	25	7 d	15	nd^{c}
2	60	1 d	58	> 95
3^{b}	100	4 h	79	> 95
4	150	5 min	14	nd^{c}
5	150	15 min	59	nd^{c}
6	150	30 min	92	80
7	200	30 min	57	80

^{a}: 2 Equivalents; ^{b}: Isolated yield of purified keto-ester **5**; ^{c}: Not determined
^{d}: Determined by ^{13}C-{^{1}H} NMR in chiral liquid crystals.

Concerning the stereoselectivity of the process, attempts to measure accurately the ee in adduct **5a** or in the related keto-acid **6** using either chiral HPLC or ^{1}H NMR spectroscopy in the presence of chiral shift reagent [Eu(hfc)$_3$] were unsuccessful. However, screening of the selectivity was made possible on the examination of ^{13}CNMR spectra of crude imine **4a** and gave satisfactory results. In order to ascertain that no epimerization of Michael adduct **5a** occurred during hydrolysis of imine **4a**, we turned our attention to NMR spectroscopy in polypeptide chiral liquid crystals (CLC) that generally provides an efficient alternative to classical methods when these latter failed or gave poor results.[12] We used here ^{13}C-{^{1}H}.[13] Spectral enantio-discriminations using ^{13}C-{^{1}H} NMR in a CLC are based on ^{13}C chemical shift anisotropy (CSA) differences. In practice, when the enantiomers are oriented differently inside the CLC, we can expect to observe two distinct resonances for each non-equivalent carbon atom discriminated. A priori, each carbon atom is a potential spy, thus increasing the possibility to visualize enantiomers.[14] Various carbon sites show enantiodiscrimination, but the best spectral separation was obtained for C3' in compound **5**. Considering the S/N ratio, the error on the ee has been estimated around 5% of the true value. Figure 1 shows the evolution of two ^{13}C-{^{1}H} signals associated to C-3' (Scheme 2) both in racemic (a) and enantio-enriched forms [Table 1, entry 6 (b) and Table 2, entry 4 (c)] oriented in a PBLG/ dichloromethane phase. The differences in peak intensity reveal the evolution of the ee. Although the separation observed at 100 MHz is rather small (< 3 Hz), a suitable evaluation of the ee is possible using deconvolution process. The absence of peak for the minor enantiomer in Fig. 1c indicates that the ee is >95%. Accuracy of the method was ascertained by measuring gradual mixtures of the racemic ketoester **5a** (prepared from racemic 1-phenymethymamine using the same conditions) with the pure enantiomer.

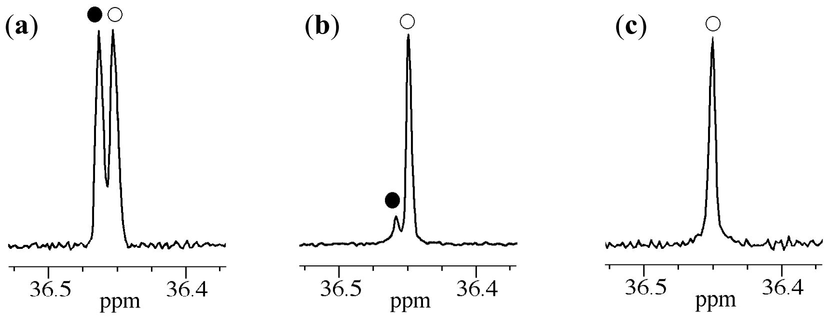
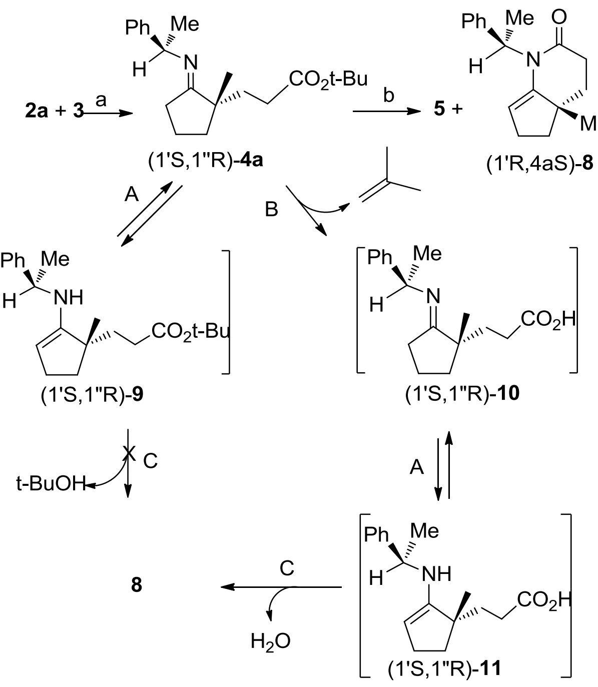


Figure 1. ^{13}C-{^{1}H} signals of carbon atom C-3' in ketoester **5** prepared in racemic form (**a**) and (*S*)-enriched one (**b**, **c**).

Analysis of NMR results obtained for synthesis of adduct **5a** using Δ clearly indicates a sharp decrease in the stereoselectivity (>95 to 80% ee) when the temperature was increased from 100 °C to 200 °C (Table 1, entries 3-4 and 6-7). This phenomenon can tentatively be explained assuming a possible competitive retro-Michael addition leading to a partial racemization of the Michael adduct under rather drastic conditions.[15]

Having secured an access to Michael adduct **5a**, a chemical correlation of its corresponding methyl keto ester **6**[10] was undertaken to ascertain the absolute configuration in **5a** (Scheme 1). Thus, keto acid **6** was easily obtained upon formic acid treatment of ketoester **5a** (Table 1, entry 3) in good yield and directly subjected to diazomethane esterification leading to (-)-(*S*)-**7**. As expected, the sense of induction is in accordance to the empirical rule defined for this reaction,[2,3] with the same sense of asymmetric induction as those obtained using methyl acrylate as the electrophile. This stereoselectivity originates from the *aza-ene* type mechanism in which internal transfer of the proton born by the nitrogen atom of the more substituted tautomeric enamine is concerted with the creation of the C-C bond (Scheme 1).

We then turned our focus to the study of this Michael reaction, carrying out the experiments under µW. As can be seen from Table 2 and Scheme 3, µW affects the reactivity, the stereo- and unexpectedly the chemoselectivity (Table 2). The closed vessel system was chosen in order to contain the toxic and volatile Michael acceptors in the reaction vessel, and to monitor the possible extent of pressure elevation during the microwave irradiation. The possibility of running reactions in an inert gas atmosphere is another distinct advantage with the sealed reaction vessel strategy. Despite sensitivity of chiral imines **2** toward water, this was not necessary in this case. The first observation was the role of stirring upon efficiency of the reaction. This parameter was found to be critical to the success of the reaction (compare entries 1,3,5 with entries 2,4,6 and 7).[16]



Scheme 2 Reaction pathway to lactam **8** under µW. *Reagents and conditions*: (a) 2 equiv. **3**, µW, 200 °C, 30 min; (b) 20% aq. AcOH, THF, 20 °C.

When mixtures of imine **2a** and alkene **3** were submitted to µW for 30 min at 100 °C with an optimal power of 30 W (Table 2, entry 2), the only reaction product was the expected chiral imine **4a**, leading to keto-ester **5a** upon hydrolytic workup. This result is a noteworthy improvement over results obtained from conventional heating at the same temperature for 4 hours (Table 1, entry 3) in terms of yields and reaction time, the enantioselectivity remaining the same.

At 150 °C, within 30 min, whereas yields were nearly comparable (Table 1, entry 6 and Table 2, entry 4), the enantioselectivity was largely improved under µW when compared to Δ (see Figure 1) with similar set of conditions (temperature, pressure, profiles of heating rates).

Table 2. Effect of µW irradiation and stirring upon condensation of imine **2a** (entries 1-7) and **2b** (entries 8-9) to *tert*-butyl acrylate **3**[a].

Entry	P	T (°C)	Stirring	ΔP	**5a/5b**	**8**
	(W)			(bar)	yield (ee) %	
1	30	100	no	0.1	85	0
2	30	100	yes	0.1	100 (> 95^{d})	0
3	80	150	no	0.7	67	0
4	80	150	yes	0.3	89 (> 95^{d})	0
5	80	200	no	3.5	11	25
6	80	200	yes	1.4	41	14
7	100	200	yes	13.0	0	53
8	80	100	yes	0.8	92 (> 95^{d})	0
9	100	200	yes	14.0	0	68

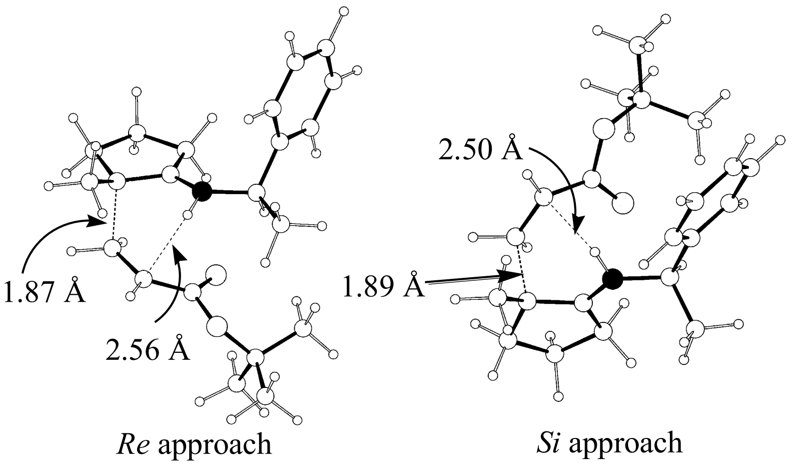
[a]: 2 Equivalents; [b]: Isolated yield of purified keto-ester **5a** after 30 min μW irradiation and subsequent hydrolysis; [c]: Not determined; [d]: Determined by ^{13}C-{^{1}H} NMR in chiral liquid crystals.

The most intriguing feature concerned the production of lactam **8** under forcing μW conditions (Scheme 3; Table 2, entries 5-7) instead of keto ester **5a**, since this lactam **8** was never observed in the conventional thermal process (Δ). This very important specific μW effect appeared when the reaction was performed at 200 °C. In contrast with Δ where the expected Michael adduct **5a** was obtained (Table 1, entry 7), within 30 min under μW irradiation at 200 °C, quite surprisingly, the lactam **8** was formed as the sole product (Scheme 2; Table 2, entry 7).

This noticeable finding on chemoselectivity can be justified by considering the possibility for μW to favor a very polar mechanism consisting in the nucleophilic addition (Scheme 2, routes C) of the enamines **9** or **11** to the carbonyl group of either a *tert*-butyl ester (**9** → **8)** with the release of *tert*-butanol, or an acid (**11** → **8**) with elimination of a water molecule. Secondary enamines **9** or **11** are in tautomeric equilibrium (Scheme 2, routes A) with imines **4a** or **10** respectively, the latter being issued from the thermolysis of the *tert*-butyl ester group in imine **4** with concomitant generation of isobutene (Scheme 2, route B, **4** → **10**).

In order to elucidate the pathway to lactam **8**, a GC-mass analysis of the headspace of the reaction mixture was undertaken, and compared to those of the starting chiral imine **2**, the acrylate **3** and *tert*-butanol having been separately irradiated under the same conditions (200 °C, 100 W, 30 min in closed reaction vessels). While isobutene was detected in the headspace of the reaction mixture and the *tert*-butyl acrylate **3** sample, it was not present in the *tert*-butanol or starting imine **1** ones.[17] This indicates that, under μW at 200 °C, the formation of lactam **8** occurred via a tandem Michael addition/deprotection/aza-annulation sequence implying the thermolysis of the *tert*-butyl ester group in imine **4** (path B/C).

Dealing with μW effect on enantioselectivity, the superiority of μW reveals the intervention of non-purely thermal μW specific effects. Although the effects observed in microwave-irradiated chemical transformations can in most cases be rationalized by purely bulk thermal phenomena associated with rapid heating to elevated temperatures,[18] we have conducted all experiments in the same conditions (closed vessels, same scale, same magnetic barrel) either in the microwave chamber or by immersion in a preheated oil bath in order to avoid as possible any difference in the temperature profiles. They can be justified by considering the reaction mechanism,[4] expecting µW effects when the polarity of the system increases during the reaction progress. It will be the case when the transition state (TS) of a reaction is more polar than its ground state, thus leading to a decrease in the activation energy.[19] Data obtained for the transition states for the *Re* and *Si*-approaches of imine **2** and *tert*-butyl acrylate **3** are consistent with the previous studies:[4b] *Re* approach: forming CC bond: 1.87 Å and forming CH bond: 2.56 Å; Si approach: forming CC bond: 1.89 Å and forming CH bond: 2.50 Å (Figure 2). As it was shown that this Michael addition proceeds through a concerted *asynchronous* '*aza-ene* like' mechanism,[4] the TS has thus a certain polarity, a situation for which µW effects are expected. This fact explains the intervention of µW effects upon exaltation of reactivity (comparing yields at 100 °C). To support this assumption, taking into account previous related AM1 computational investigations,[20] we calculate the corresponding approaches between the enamine tautomer of imine **2** with *tert*-butyl acrylate **3**. The *Re*-approach (the favored one) leads to a slightly more polar as well as more asynchronous than the *Si*-approach TS (Figure 2).[21] Consequently, this TS will be slightly favored due to a better dipole-dipole stabilization by µW. Therefore, under µW, the selectivity in favor of the *Re*-approach will be even more improved (exp. from 80 to > 95% ee). Reduced reversibility of the Michael reaction under microwave conditions could also contribute to the superior enantioselectivity, both phenomenon leading to an increased stereoselectivity in such conditions.



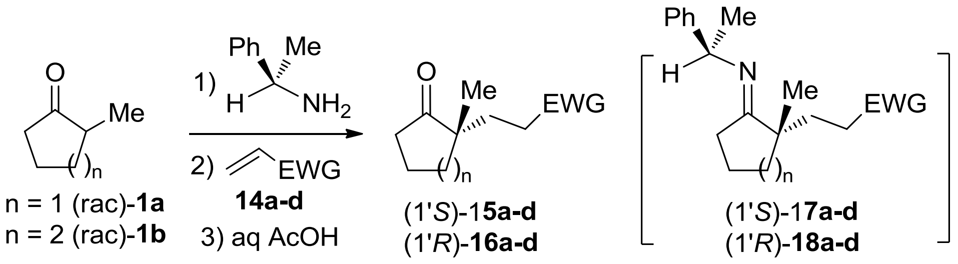
Re approach
ΔH = -37.22 kcal/mol
µ = 4.37 D

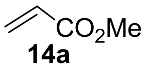
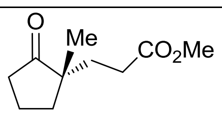
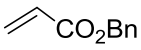
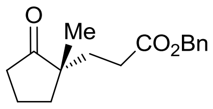
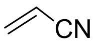
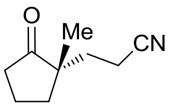
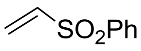
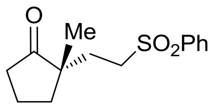
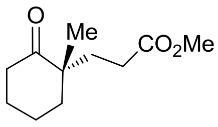
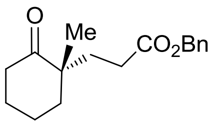
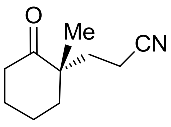
Si approach
ΔH = -34.12 kcal/mol
µ = 4.33 D

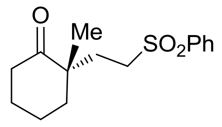
Figure 2. Transition structures at the RHF AM1 level for *Re* (left) and *Si* (right) approaches of *tert*-butyl acrylate **3** to enamine imine **2**, their respective enthalpies of formation and dipole moment.[20]

The same trend was observed with chiral imine **2b** derived from 2-methylcyclohexanone **1b**. Lactame **8b** was obtained when the reaction was performed at 200 °C upon irradiation at 100 W for 30 min (Table 2, entry 9) in a 68% yield while the expected Michael adduct 5b was obtained in the optimized conditions (100 °C, 80 W, Table 2, entry 8) in 92% yield with >95% ee. We next examine the reactivity of these imine in this µW promoted AMACI (Table 3).

Table 3. Effect of temperature on the synthesis of ketoester **4** by condensation of chiral imine **1** with electrophilic alkenes **12a-d**.[a]



Entry	Electrophile	Ketone	Michael adduct	Yield %[b] ee%
1	 **14a**	**1a**	 **7a**	87 [(d)] > 95
2	 **14b**	**1a**	 **15b**	76 [(d)] > 95
3[b]	 **14c**	**1a**	 **15c**	47 [(e)] > 95
4	 **14d**	**1a**	 **15d**	95 [(b)] >95
5	**14a**	**1b**	 **7b**	77 [(d)] > 95
6	**14b**	**1b**	 **16b**	60 [(c)] > 95
7	**14c**	**1c**		74 [(c)] > 95

			16c	
8	**14d**	**1d**	 **16d**	33 (c) > 95

[a]: 2 Equivalents; [b]: Isolated yield of purified keto-esters; (c) 80 W, 150 °C, 30 min.; (d) 80 W, 100 °C, 30 min.; (e) 80 W, 100 °C, 15 min.

The reaction performed well, giving the expected Michael adducts in adduct in 33-95% yield, in a first series of experiments. Ee of the adducts were measured at the level of the crude imines as >95% (no diastereoisomers detected).

With these satisfactory results in hand, we will turn to the study of a catalytic version of this reaction, in the context of a greener generation of quaternary carbon centers in a simple manner. Work is in progress to extend this µW activation mode to engage substituted acceptors in the AMACI.

3. Materials and Methods

3a. Chemistry: General: All reactions not involving aqueous media were carried out under a nitrogen atmosphere in a flame-dried glassware. Commercial reagents were used without further purification. Reactions were followed by ^{1}H NMR in $CDCl_3$ or using thin-layer chromatography, carried out on silica gel plates, which were viewed by UV irradiation at 254 nm and/or by staining with phosphomolybdic acid or *p*-anisaldehyde. Flash column chromatography was performed with 230-400 mesh silica gel. Melting points were recorded on a digital melting point apparatus. IR spectra were recorded with a Fourier transform spectrometer Bruker VECTOR 22. NMR spectra of the crude reaction mixtures were recorded in $CDCl_3$ containing a pinch of sodium carbonate in order to prevent hydrolysis of the imines. Imines proved to be stable for days in such conditions. ^{1}H NMR spectra were recorded at 300 K, at 200 or 400 MHz on a Bruker AC 200 or Bruker Avance 400 spectrometer, with $CHCl_3$ as internal standard (δ_H = 7.26 ppm). ^{13}C NMR spectra were recorded at 300 K, at 50 or 100 MHz, with the central peak of $CHCl_3$ as internal standard (δ_C = 77.0 ppm, central line). Recognition of methyl, methylene, methine and quaternary carbon nuclei in ^{13}C NMR spectra rests on the *J*-modulated spin-echo sequence. 2Dl NMR experiments (COSY, HMQC, HMBC and NOESY) were used for the assignments of signals in the ^{1}H and ^{13}C NMR spectra. For NMR measurements, ^{13}C 1D NMR experiments in polypeptidic oriented solvents were performed on a Bruker DRX-400 equipped with a BBO probe, and hence no additional hardware equipment is basically required. All proton-decoupled ^{13}C NMR experiments were recorded by applying the WALTZ-16 composite pulse sequence to decouple protons and benefit from NOE effect.

For unambiguous assignment of enantiomers in chiral NMR, comparison was made in all cases with the corresponding racemates. Optical rotations were measured at 589 nm in a 1 dm-cell using an Optical Activity Limited AA-10R apparatus and are expressed in g/100mL. Elemental analyses were performed by the Service de microanalyse, BioCIS, Châtenay-Malabry, France, with a Perkin-Elmer 2400 analyzer. A CEM Discover monomode reactor with an accurate control of temperature and pressure by modulation of emitted W was used for the microwave experiments. **Caution:** It is essential that great precaution be taken when carrying out organic reactions in sealed vessels. In particular, safety devices are to be used including appropriate septa as a pressure relief system and an automatic cut off of the microwave irradiation before the pressure limit of the vessels has been reached. See: Raner, K. D.; Strauss, C. R.; Trainor, R. W.; Thorn, S. J. *J. Org. Chem.* **1995**, *60*, 2456 and references cited therein. In this study, the maximum developed pressure was 14 bar at 100 W and 200 °C, far below the pressure limit (20 bar).

Preparation of chiral imines **2** exemplified for **2a**: In a 100 mL round bottom flask equipped with a Dean-Stark apparatus, 21.6 mL (0.2 mol) 2-methylcyclopentanone, 27.3 mL (0.21 mol, 1.05 equiv.) (*R*)-1-phenylethylamine (ee = 99%) are successively added to cyclohexane (50 mL). The resulting mixture was stirred for 18 h under nitrogen at 110 °C (oil bath) with azeotropic removal of water. Cyclohexane was then distilled and fractional distillation of the crude under reduced pressure afforded the desired chiral imine **1** as a colourless oil.

(1'*R*)-(2-Methylcyclopentylidene)-(1'-phenyl-ethyl)-amine (2a). Colorless oil (89%); B.p. = 80 °C (0.01 Torr); IR (neat, ν cm^{-1}): 2958, 1673; ^{1}H NMR (200 MHz, $CDCl_3$) δ ppm: 7.33-7.05 (m, 5H), 4.43-4.28 (m, 1H), 2.45-1.15 (m, 7H), 1.43 and 1.41 (d, *J* = 6.1 Hz, 3H), 1.11 and 1.10 (d, *J* = 6.7 Hz, 3H); ^{13}C NMR (50 MHz, $CDCl_3$) δ ppm: 181.0 and 180.6 (C), 146.2 and 145.9 (C), 128.1 (2CH), 126.5 (CH), 126.3 (2CH), 61.5 and 61.2 (CH), 41.2 and 41.1 (CH), 32.8 and 32.7 (CH_2), 28.7 and 28.5 (CH_2), 24.8 and 24.5 (CH_3), 22.5 and 22.4 (CH_2), 17.5 (CH_3).

General procedure for the asymmetric Michael reactions using microwaves: Mixtures of imines **2** (1 to 6 mmol) and 2 equivalents of electrophilic alkene were placed in sealed,[33] 10 mL heavy-walled pyrex tubes.[34] The tubes were introduced in the cavity of a single-mode[35] device allowing control of irradiation power (up to 300 W), time, temperature and pressure (see article, Tables 1 and 2).[36] The tube was opened after the reaction mixture was rapidly cooled down to room temperature, and excess *tert*-butyl acrylate was removed in vacuo. An aliquot of crude reaction mixture was analyzed by ^{1}H and ^{13}C NMR. For the crude reaction mixtures containing alkylated imine, after being vigorously stirred with 20% aqueous acetic acid (2 mL/mmol) and THF (2 mL/mmol) for 17 hours, the reaction mixture was concentrated, then thoroughly extracted with Et_2O (3 x 10 mL). The combined organic phase was washed successively with saturated $NaHCO_3$ and NaCl solutions, dried ($MgSO_4$), filtered over Celite® and concentrated in vacuo. Chromatographic purification on silica gel (cyclohexane:ethyl acetate, 9:1) afforded keto derivatives **5** and/or lactam **8** as colorless oils.

***tert*-Butyl (1'*S*,1''*R*)-3-[1'-Methyl-2-(1''-phenyl-ethyl-imino)-cyclopentyl]-propionate 4a.** IR (neat, ν cm^{-1}): 2966, 2868, 1726, 1673; ^{1}H NMR (200 MHz, $CDCl_3$) δ ppm: 7.38-7.07 (m, 5H), 4.35 (q, 1H, *J* = 6.6 Hz), 2.37-2.04 (m, 4H), 1.86-1.41 (m, 6H), 1.38 (s, 9H), 1.36 (d, 3H, *J* = 6.6 Hz), 0.98 (s, 3H); ^{13}C NMR (50 MHz, $CDCl_3$) δ ppm: 180.7 (C), 173.5 (C), 146.2 (C), 128.0 (2CH), 126.3 (2CH), 126.1 (CH), 79.7 (C), 60.8 (CH), 45.5 (C), 37.0 (CH_2), 33.4 (CH_2), 30.9

(CH_2), 28.4 (CH_2), 28.0 (3 CH_3), 24.8 (CH_3), 23.9 (CH_3), 20.5 (CH_2).

***tert*-Butyl (1’*S*,1”*R*)-3-[1’-Methyl-2-(1”-phenyl-ethyl-imino)-cyclohexyl]-propionate 4b.** IR (neat, ν cm^{-1}): 3061, 3063, 2966, 2926, 1658, 1493, 1447; ^{1}H NMR (300 MHz, $CDCl_3$) δ ppm: 7.41-7.12 (m, 5H), 4.69 (q, 3H, *J* = 6.4 Hz), 2.40-2.21 (m, 1H), 1.57-1.30 (m, 8H), 1.17-1.11 (2d, 3H, *J* = 4.3 Hz), 1.01-0.99 (d, 3H, *J* = 4,9 Hz); RMN ^{13}C (75 MHz, $CDCl_3$) δ ppm: 173.1 (C), 146.7 (C), 127.9 (CH), 126.3 (CH), 126.2 (CH), 125.9 (CH), 125.8 (CH), 57.3 (CH), 42.0 (CH_2), 35.6 (CH), 27.5 (CH_2), 25.4 (CH_3), 25.3 (CH_2).

***tert*-Butyl (1’*S*)-3-(1’-Methyl-2’-oxocyclopentyl)-propionate 5a.** IR (neat, ν cm^{-1}): 2968, 2934, 2872, 1727; ^{1}H NMR (200 MHz, $CDCl_3$) δ ppm: 2.38-2.06 (m, 4H), 2.0-1.54 (m, 6H), 1.36 (s, 9H), 0.98 (s, 3H); ^{13}C NMR (50 MHz, $CDCl_3$) δ ppm: 222.3 (C), 172.6 (C), 80.1 (C), 47.4 (C), 37.3 (CH_2), 35.9 (CH_2), 31.3 (CH_2), 30.5 (CH_2), 27.9 ($3CH_3$), 21.3 (CH_3), 18.4 (CH_2); $[\alpha]_D^{24}$ = -27.8 (c = 4.5, $EtOH_{abs}$); Anal. calcd for $C_{13}H_{22}N$: C, 68.99; H, 9.80. Found: C, 68.84; H, 9.75%.

***tert*-Butyl (1’*S*)-3-(1’-Methyl-2’-oxocyclohexyl)-propionate 5b.** IR (neat, ν cm^{-1}): 2933-2866, 1727, 1704; RMN ^{1}H (200 MHz, $CDCl_3$) δ ppm: 2.49-2.33 (2H), 2.28-2.12 (1H), 2.10-1.94 (3H), 1.92-1.52 (8H), 1.42 (9H), 1,02 (3H); RMN ^{13}C (75 MHz, $CDCl_3$): δ ppm: 215.7 (C), 173.1 (C), 82.3 (C), 47.9 (C), 39.3 (CH_2), 38.7 (CH_2), 32.4 (CH_2), 30.3 (CH_2), 28.7 ($3CH_2$), 27.4 (CH_2), 22.4 (CH_3), 21.0 (CH_2).

(4a*S*,1’*R*)-4a-Methyl-1-(1’-phenylethyl)-1,3,4,4a, 5,6-hexahydro-[1]pyrindin-2-one 8. IR (neat, ν cm^{-1}): 1665, 1629; ^{1}H NMR ($CDCl_3$, 200 MHz) δ ppm: 7.24-7.06 (m, 5H), 6.14 (q, 1H, *J* = 7.1 Hz), 4.31 (t, 1H, *J* = 2.5 Hz), 2.62-2.53 (m, 1H), 2.57 (dd, 1H, *J* = 8.3, 4.0 Hz), 2.30 (m, 1H), 2.05 (ddd, 1H, *J* = 15.6, 9.0, 3.1 Hz), 1.76-1.43 (m, 4H), 1.55 (d, 3H, *J* = 7.1 Hz), 1.05 (s, 3H); ^{13}C NMR ($CDCl_3$, 50 MHz) δ ppm: 169.3 (C), 143.7 (C), 141.2 (C), 128.3 (2CH), 126.5 (CH), 126.1 (2CH), 105.7 (CH), 49.9 (CH), 43.5 (C), 38.3 (CH_2), 33.4 (CH_2), 29.9 (CH_2), 28.5 (CH_2), 21.1 (CH_3), 14.8 (CH_3); Anal. calcd for $C_{17}H_{21}NO$: C, 79.96; H, 8.29; N, 5.49. Found: C, 80.04; H, 8.56; N, 5.15%.

3-[1-Methyl-2-(1-phenyl-ethylimino)-cyclopentyl]-propionitrile 17c. IR (neat, ν cm^{-1}): 3027, 2962, 2867, 2245, 1672; ^{1}H NMR ($CDCl_3$, 300 MHz) δ : 7.35-7.17 (m, 5H), 4.41 (q, 1H, *J* = 6.6 Hz), 2.65-2.27 (m, 4H), 1.94-1.57 (m, 6H), 1.41 (d, 3H, *J* = 6.6, Hz), 1.06 (s, 3H); ^{13}C NMR ($CDCl_3$, 75 MHz) δ ppm: 179.9 (C), 145.8 (C), 128.2 (2CH), 126.4 (CH), 126.3 (2CH), 120.7 (C), 61.2 (CH), 45.4 (C), 37.1 (CH_2), 35.8 (CH_2), 28.5 (CH_2), 24.9 (CH_3), 23.8 (CH_3), 20.5 (CH_2), 12.4 (CH_2).

3-[1-Methyl-2-(1-phenyl-ethylimino)-cyclohexyl]-propionitrile 18c. IR (neat, ν cm^{-1}): 2969, 2931, 2866, 2247, 1705, 1650; ^{1}H NMR ($CDCl_3$, 300 MHz) δ ppm: 7.39-7.20 (m, 5H), 4.70 (q, 1H, *J* = 6.6 Hz), 2.53-2.14 (m, 5H), 2.13-1.81 (m, 1H), 1.74-1.27 (m, 6H), 1.36 (d, 3H, *J* = 6.6, Hz), 1.04 (s, 3H); ^{13}C NMR ($CDCl_3$, 75 MHz) δ ppm: 172.1 (C), 146.6 (C), 128.2 (2CH), 126.4 (3CH), 121.1 (C), 57.8 (CH), 43.2 (C), 39.0 (CH_2), 35.2 (CH_2), 27.1 (CH_2), 25.8 (CH_3), 24.9 (CH_2), 23.9 (CH_3), 21.2 (CH_2), 12.4 (CH_2).

Methyl (1’*S*)-3-(1’-Methyl-2’-oxocyclohexyl)-propionate 7b. IR (neat, ν cm^{-1}): 2960, 1731, 1437; ^{1}H NMR (200 MHz, $CDCl_3$) δ ppm: 3.33 (s, 3H), 2.74-2.18 (m, 4H), 1.98-1.62 (m, 6H), 1.01 (s, 3H); ^{13}C NMR (50 MHz, $CDCl_3$) δ ppm: 222.3 (C), 173.8 (C), 51.5 (CH_3), 47.4 (C), 37.3 (CH_2), 35.9 (CH_2), 32.0 (CH_2), 29.2 (CH_2), 21.2 (CH_3), 18.4 (CH_2); Anal. for $C_{10}H_{16}O_3$, calcd. C, 65.19; H, 8.75; found C, 65.08; H, 8.79.; **12a** $[\alpha]_D^{20}$ -34.5 (c = 2; $EtOH_{abs}$); lit.[38] *ent*-**12a** $[\alpha]_D^{20}$ +35.7 (c = 2, $EtOH_{abs}$).

3-(1-Methyl-2-oxo-cyclopentyl)-propionitrile 15c. IR (neat, ν cm^{-1}): 2965, 2927, 2872, 2247, 1731; ^{1}H NMR (300 MHz, $CDCl_3$) δ ppm: 2.46-2.14 (m, 4H), 1.97-1.69 (m, 6H), 1.02 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ ppm: 221.2 (C), 119.7 (C), 47.2 (C), 37.2 (CH_2), 35.7 (CH_2), 32.0 (CH_2), 20.9 (CH_3), 18.4 (CH_2), 12.4 (CH_2); $[\alpha]_D^{20}$ = -34.6° (c = 0.003, $EtOH_{abs.}$); Anal. Calcd for $C_9H_{13}NO$: C, 71.49; H, 8.67; N, 9.26; O, 10.58. Found: C, 71.06; H, 8.12; N, 9.07.

3-(1-Methyl-2-oxo-cyclohexyl)-propionitrile 16c. IR (neat, ν cm^{-1}): 2937, 2867, 2246, 1700, 1451; ^{1}H NMR (300 MHz, $CDCl_3$) δ ppm: 2.48-2.37 (m, 1H), 2.33-2.26 (m, 3H), 1.97-1.64 (m, 8H), 1.12 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ ppm: 214.3 (C), 120.0 (C), 47.6 (C), 38.5 (2CH_2), 33.6 (CH_2), 27.1 (CH_2), 22.2 (CH_3), 20.8 (CH_2), 12.3 (CH_2); Anal. for $C_{10}H_{15}NO$, calcd. C, 72.69; H, 9.15; found C, 72.23; H, 8.49; MS (ESI): 166 (M+1) **16c** $[\alpha]_D^{24}$ 9.7 (c = 0.02, $EtOH_{abs}$).

[2-(2-Benzenesulfonyl-ethyl)-2-methyl-cyclopentylidene]-(1-phenyl-ethyl)-amine 17d IR (neat, ν cm^{-1}): 2962, 1671, 1447, 1305, 1145; ^{1}H NMR (300 MHz, $CDCl_3$) δ ppm: 7.70-7.45 (m, 10H), 4.34 (q, 1H, *J* = 6.6 Hz), 3.37 (m_c, 2H), 2.24 (bt, 2H, *J* = 5.9 Hz), 1.80-1.40 (m, 6H), 1.31 (d, 3H, *J* = 6.6 Hz), 1.00 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ ppm: 179.6 (C), 146.0 (C), 138.3 (CH), 133.6 (CH), 129.2 (2CH), 129.1 (2CH), 128.1 (CH), 127.7 (CH), 126.2 (CH), 67.8 (CH_2), 61.1 (CH), 52.2 (C), 37.6 (CH_2), 28.3 (CH_2), 25.5 (CH_2), 25.0 (CH_3), 23.8 (CH_3), 20.4 (CH_2).

[2-(2-Benzenesulfonyl-ethyl)-2-methyl-cyclohexylidene]-(1-phenyl-ethyl)-amine 18d IR (neat, ν cm^{-1}): 2928, 1707, 1650; ^{1}H NMR (300 MHz, $CDCl_3$) δ ppm: 8.00-7.87 (m, 4H), 7.66-7.49 (m, 6H), 4.61 (q, 1H, *J* = 6.6 Hz), 3.35-3.20 (m, 2H), 2.42-2.29 (m, 2H), 2.18-2.00 (m, 2H), 1.90-1.38 (m, 6H), 1.24 (d, 3H, *J* = 6.6 Hz), 1.01 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ ppm: 172.2 (C), 146.6 (C), 139.3 (C), 133.4 (CH), 129.1 (CH), 128.2 (CH), 126.4 (CH), 57.7 (CH), 52.2 (CH_2), 43.0 (C), 39.3 (CH_2), 31.9 (CH_2), 27.0 (CH_2), 25.7 (CH_3), 24.7 (CH_2), 24.2 (CH_3), 21.2 (CH_2).

2-(2-Benzenesulfonyl-ethyl)-2-methyl-cyclopentanone 15d Mp = 69 °C; IR (neat, ν cm^{-1}): 2965, 2870, 1729; ^{1}H NMR (300 MHz, $CDCl_3$) δ ppm: 7.87 (bd, 2H, *J* = 8.3 Hz), 7.68-7.45 (m, 3H), 3.14 (ddd, 1H, *J* = 18.4, 12.0, 4.7 Hz), 2.96 (ddd, 1H, *J* = 18.4, 12.2, 4.9 Hz), 2.35-2.07 (m, 2H), 1.91-1.72 (m, 6H), 0.94 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ ppm: 221.4 (C), 138.7 (C), 133.7 (CH), 129.2 (2CH), 127.9 (2CH), 51.8 (CH_2), 46.7 (C), 37.2 (CH_2), 36.2 (CH_2), 28.9 (CH_2), 21.0 (CH_3), 18.4 (CH_2); $[\alpha]_D^{23}$ = -19.4° (c = 0.0075, $EtOH_{abs.}$); MS (APCI): m/z 267 (100%) $[M + H]^+$; Anal. Calcd for $C_{14}H_{18}O_3S$: C, 63.13; H, 6.81; O, 18.02; S, 12.04. Found: C, 62.65; H, 6.80.

2-(2-Benzenesulfonyl-ethyl)-2-methyl-cyclohexanone 16d Mp = 74-77 °C; IR (neat, ν cm^{-1}): 2934, 2868, 1700; ^{1}H NMR (300 MHz, $CDCl_3$) δ ppm: 7.86 (bd, 2H, *J* = 8.3 Hz), 7.65-7.49 (m, 3H), 3.06 (ddd, 1H, *J* = 13.7, 11.9, 5.0 Hz), 2.98 (ddd, 1H, *J* = 13.7, 11.7, 5.0 Hz), 2.39-2.13 (m, 2H), 2.00-1.88 (m, 1H), 1.79-1.55 (m, 7H), 1.01 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ ppm: 214.3 (C), 138.9 (C), 133.6 (CH), 129.2 (2CH), 127.9 (2CH), 51.8 (CH_2), 47.4 (C), 38.9 (CH_2), 38.4 (CH_2), 30.2 (CH_2), 27.1 (CH_2), 22.3 (CH_3), 20.8 (CH_2); $[\alpha]_D^{24}$ = +2.9° (c = 0.01, $EtOH_{abs.}$); MS (APCI): m/z 281 (100%) $[M + H]^+$; Anal. Calcd for $C_{15}H_{20}O_3S$: C, 64.26; H, 7.19; O, 17.12; S, 11.44. Found: C, 64.40; H, 7.03.

General procedure for the synthesis of ketoacid 6

A mixture of adduct **6a** (452 mg, 2 mmol) and formic acid (2 mL) was stirred at 20 °C for 2 h. Formic acid was distilled, the crude was taken up in Et_2O (10 mL), washed with saturated $NaHCO_3$

solution (2 x 10 mL). The aqueous layer was acidified at 0 °C (6*N* HCl) and thoroughly extracted (4 x 10 mL Et_2O). The combined organic phase was dried ($MgSO_4$) and filtered (Celite®) and the crude concentrated in vacuo to give keto acid **6a** (302 mg, 89%) as a colorless oil. This material was used without further purification in the next step.

(1'*S*)-3-(1'-Methyl-2-oxocyclopentyl)-propionic acid 6a. IR (neat, ν cm^{-1}): 3512, 3090,2963, 2873, 2663, 1729, 1706; ^1H NMR (200 MHz, $CDCl_3$) δ ppm: 11.0 (bs, 1H), 1.67-1.33 (m, 4H), 1.19-0.81 (m, 6H), 0.20 (s, 3H); ^{13}C NMR (50 MHz, $CDCl_3$) δ ppm: 222.8 (C), 178.3 (C), 47.0 (C), 36.8 (CH_2), 35.4 (CH_2), 30.6 (CH_2), 28.7 (CH_2), 20.7 (CH_3), 18.0 (CH_2); Anal. for $C_9H_{14}O_3$, calcd. C, 63.51; H, 8.29; found C, 63.28; H, 8.39; $[\alpha]^{25}_D$ -40.6 (c = 1.6, $EtOH_{abs}$).

(1'*S*)-3-(1'-Methyl-2-oxocyclohexyl)-propionic acid 6b. IR (neat, ν cm^{-1}): 3502, 2935, 2866, 1700; ^1H NMR (300 MHz, $CDCl_3$) δ ppm: 9.4 (bs, 1H), 2.38-2.29 (m, 3H), 2.23-2.12(m, 1H), 2.03-1.93 (m, 1H), 1.84.68 (m, 6H), 1.62-158 (m, 1H), 1.05 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ ppm: 215.5 (C), 179.3 (C), 47.7 (C), 39.0 (CH_2), 38.5 (CH_2), 32.2 (CH_2), 28.9 (CH_2), 27.3 (CH_2), 22.3 (CH_3), 20.8 (CH_2); MS (ESI) 207 (M+23) 391 (2M+23); $[\alpha]^{28}_D$ -85 (c = 0.13, $EtOH_{abs}$).

Chemical correlation to Methyl (1'*S*)-3-(1-Methyl-2-oxo-cyclopentyl)-propionate 7a: A 0.5 M solution of diazomethane in Et_2O (5 mL) was added to a solution of the acid **6** (97 mg, 0.57 mmol) in dry Et_2O (20 mL) at 0 °C. The resulting mixture was stirred at room temperature for 2 hours and the excess of diazomethane was destroyed with acetic acid. The reaction mixture was washed with brine (2×5 mL), dried ($MgSO_4$), and concentrated in vacuo. The residue was purified by column chromatography (cyclohexane/AcOEt, 9:1) to give methyl ester **7a** as a colourless oil (100 mg, 95%); $[\alpha]_D^{20}$ -32.5 (c = 2; $EtOH_{abs}$); lit.[10] *ent*-**12a** $[\alpha]_D^{20}$ +35.7 (c = 2, $EtOH_{abs}$).

3b. Enantiomeric excess determination: For ^{13}C-{^1H} 1D NMR experiments in polypeptidic oriented solvents, the sample preparation consisted of directly weighting 25-30 mg of solute, around 140 mg of poly-γ-benzyl-L-glutamate (PBLG, DP= 463, commercially available) and adding about 350 mg of dichloromethane into a 5 mm NMR tube. Under these conditions, the total volume of the sample is optimal compared to the length of the coil of a 5 mm diameter probe-head. Compared to previous work, we have used a larger amount of PBLG than usual (100 mg) in order to obtain a clean liquid crystalline phase. This is a consequence of the relatively low degree of polymerization (DP= 463, i.e. MW ≈101000 $g.mol^{-1}$). The exact composition of each NMR sample is given in Table 3. To avoid the evaporation of dichloromethane during long NMR experimental time, we have sealed the samples. Note here that a mixture of protonated and deuterated dichloromethane was used. This solution allows to easily shim the magnet on the proton FID as well as to minimize the digitization problems associated with the dynamic range of the Analogue-to-Digital Converter (ADC) induced by the difference of ^{13}C signal intensity between dichloromethane and solute. In other hands, the shape and the line width of ^{13}C resonances of dichloromethane provide a serious control of the magnet stability as well as possible time-evolution of the sample homogeneity during the experiments. The sample is then centrifuged during few seconds, then inverted and centrifuged again. This process is repeated until an optically homogeneous birefringent phase is obtained.

Table 3. Composition of liquid-crystalline representative NMR samples investigated

Sample	Solute	Solute / mg[a]	Co-solvent / mg[a]	Polymer % in weight
1	***(rac)*-6**	27	75/275	27.1
2	**(*S*)-6**	28	75/277	26.9
3	**(*S*)-6**	28	75/277	27.1

Conditions: Polymeric solvent: PBLG; Degree of polymerization of polypeptide used: 463; Co-solvent: CH_2Cl_2 / CD_2Cl_2; *Polymer*/mg: 140; [a]The accuracy on the weighing is 1 mg.

The NMR tube was not spun in the magnet and its temperature was regulated carefully at 299 K using the standard variable temperature control unit (BVT 3000). ^{13}C spectra were recorded by adding 1000 to 3000 scans. Gaussian filtering was applied to improve the spectral separation of resonances. The area measurement was performed using a curve fitting algorithm based on complex least squares treatment of the ^{13}C NMR signals with and without filtering. Note that the experiments and the area measurements were repeated several times to estimate accurately the error on the enantiomeric excess of the mixture.

3c. GC-MS mechanistic studies: An HP 5989A GC-MS system (Hewlett-Packard, Palo Alto, CA, USA) was used. The chromatographic separation was performed with an Omega delta-3 capillary column (length: 25 m; I.D. 0.20 mm; film thickness: 0.2 µm) (Macherey-Nagel, Düren, Germany). In view of comparison, samples consisting of either the asymmetric Michael addition reaction mixture [1 equiv. chiral imine (**2a**) and 1.1 equiv. *tert*-butyl acrylate (**3**)] or each of the individual components (*tert*-butyl acrylate (**3**), chiral imine (**2a**), 2-methyl cyclopentanone **1a**, 1-phenylethylamine, 2-methyl-2-propanol) were separately irradiated at 100 W and 200 °C for 30 min in 10 mL teflon sealed glass vials and cooled to r.t. prior to GC-mass analysis. The teflon sealed glass vials filled with the reaction medium were maintained at 70°C during 15 minutes prior analysis and immediately processed. The head space (5 µL) was sampled using an airtight syringe and injected in splitless mode. Helium pressure was 50 kPa. The injector temperature was 250 °C and the initial oven temperature was 35 °C. This temperature was maintained for 1 min, the temperature was then programmed as follows: 4 °C/min up to 50 °C then 6 °C/min up to 100 °C followed by a 5 min hold. The transfer line temperature was set to 280 °C. Analysis was performed by electronic impact ionisation. The ion source and quadrupole temperature was set to 200 °C and 100 °C respectively. The electron energy was 70 eV. Acquisition was performed in scan mode over the range of 20 to 150 at a scan rate of 0.9 scan/sec (4 samplings per scan). Analysis of the results obtained for the head space of the asymmetric Michael reaction showed that the peak observed at 1.605 min correspond to 2-methyl-2-butene (MW = 56 g.mol^{-1}, identical fragmentation and comparable abundances)[17] A similar peak was not detected from the head space of the other samples, except for the *tert*-butyl acrylate (MW = 128 g.mol^{-1}) one. However, ions at m/z 55, 57 and 59 are not present in the same ratio for the 2-methyl-2-butene mass spectrum. Moreover, analysis of the *tert*-butanol (MW = 74 g.mol^{-1}) sample indicates that in the reaction conditions, *tert* butanol did not led to 2-methyl-2-butene. This set of results gave evidence that 2-methyl-2-butene and not *tert*-butanol was released during the lactamization process of Michael adduct (**5**) (see article, Scheme 3).

3d. Theroretical calculations: Geometries for the reactants were optimized by means of gradient technique at RHF AM1 level[20] by using the semi-empirical molecular orbital program MOPAC.[21] All the RHF AM1 transition structures were located using the procedures implemented in MOPAC (Version 5.0). All variables were optimized by minimizing the sum of the squared scalar gradients (NLLSQ and SIGMA).[22,23] Force calculations were carried out to ensure that the transition structures located had one imaginary frequency. Final values of the gradient norms were <1 kcal/Å and each transition structure had one negative eigenvalue in the Hessian matrix as required.

.

4. Conclusions

In conclusion, we have demonstrated that the reaction of hindered *tert*-butyl acrylate **2** in the AMACI was efficiently promoted under µW activation, compared to conventional heating. As expected, regardless the activation mode, the control of the stereochemistry of the newly created quaternary carbon center in such Michael adducts is always dictated by the configuration of the chiral inductor. The stereoselectivity of the process was determined using natural abundance ^{13}C-{^{1}H} NMR in a chiral polypeptide liquid crystal. Moreover, the temperature profiles achieved under microwave irradiation are not accessible in conventional heating and can allow a differentiation in the reaction pathways. A direct and stereoselective access to lactams **8** was thus achieved only under µW, although in moderate yield. A highly stereoselective process (ee > 95 %) was obtained either at 100 °C for 4 h (Δ) or for 30 min (µW, 100 W). A good correlation was obtained between experiment and theoretical calculations. Both the more polar and asynchronous transition state led to the expected Michael adduct (*S*)-**4**, and are favored under µW activation, allowing the reaction to proceed efficiently in minutes. Finally, Michael adducts from methyl acrylate, benzyl acrylate, vinylsulfone and acrylonitrile are regio- and stereoselectively obtained in high yield and short time using the microwave process.

Acknowledgments

We thank Dr Kathalyn L. Serre, Hanaa Tagzhouty and Kristine Aerola (BioCIS) for preliminary experiments, Sophie Mairesse-Lebrun for performing the microanalysis, CNRS and MENRS for financial support and CEM for technical support to A. L.

Author Contributions

Françoise Dumas ensure the conception and design of the chemistry and Philippe Lesot the the enantiomeric purity determination using natural abundance ^{13}C-{^{1}H} NMR. Lucie Vandromme, Li Chen and Lai Wei performed the chemical experiments and analyzed the data, Franck Le Bideau, and Françoise Dumas wrote the manuscript, André Loupy helped and advised us for the microwave chemistry, Olivier Lafon and Philippe Lesot were in charge of the NMR determination of selectivity in chiral liquid phase, Elise Tran Huu Dau performed the theoretical calculations and Pierre Chaminade the CPV analysis.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Denissova, I.; Barriault, L. Stereoselective formation of quaternary carbon centers and related fucntions. Tetrahedron report number 661, *Tetrahedron* **2003**, *59*, 10105-10146. (b) *Quaternary*

Stereocenters: Challenge and Solutions for Organic Synthesis (Eds.: J. Christoffers, A. Baro), Wiley-VCH, Weinheim, Germany, **2005**.

2. Reviews: (a) d'Angelo, J.; Desmaële, D.; Dumas, F.; Guingant, A. The asymmetric Michael addition reactions using chiral imines. *Tetrahedron: Asymmetry* **1992**, *3*, 459-505. (b) d'Angelo, J.; Cavé, C.; Desmaële, D.; Dumas, F. The Asymmetric Michael Addition Reactions Using Chiral Imines: Application to the synthesis of Compounds of Biological Interest. in *Trends in Organic Chemistry* Pandalai S. G. Ed.; Transworld Research Network, Trivandrum, India **1993**, volume 4, pp 555-616.
3. See for example: (a) Pizzonero, M.; Dumas, F.; d'Angelo, J. Enantioselective Synthesis of (*R*)-1-Azaspiro[4.4]nonane-2,6-dione Ethylene Ketal, Key Chiral Intermediate in the Elaboration of (-)-Cephalotaxine. *Heterocycles* **2005**, *66,* 31-37. (b) Kousara, M.; Ferry, A.; Le Bideau, F.; Serré, K. L.; Chataigner, I.; Morvan, I.; Dubois, J.; Chéron, M.; Dumas, F. First enantioselective total synthesis and configurational assignments of suberosenone and suberosanone as potential antitumor agents**.** *Chem. Commun.*, **2015**, *51*, 3458-3461. (c) Ito, F.; Ohbatake, Y.; Aoyama, S.; Ikeda, T.; Arima, S.; Yamada, Y.; Ikeda, H.; Nagamitsu, T.: Total Synthesis of (+)-Clavulatriene A. *Synthesis* **2015**, *47*, 1348-1355.
4. Sevin, A.; Tortajada, M., Pfau, M. Toward a transition-state model in the asymmetric alkylation of chiral ketone secondary enamines by electron-deficient alkenes. A theoretical MO study. *J. Org. Chem.* **1986**, *51*, 2671-2675. (b) Lucero, M. J.; Houk, K. N. Conformational Transmission of Chirality: The Origin of 1,4-Asymmetric Induction in Michael Reactions of Chiral Imines. *J. Am. Chem. Soc.* **1997**, *119*, 826-827. (b) Tran Huu Dau, M. E.; Riche, C.; Dumas, F.; d'Angelo, J. The origin of the stereoselectivity in the asymmetric Michael reaction using chiral imines/secondary enamines under neutral conditions: a computational investigation. *Tetrahedron: Asymmetry* **1998**, *9*, 1059-1064 and quoted references.
5. See interalia: (a) Camara, C.; Joseph, D.; Dumas, F.; d'Angelo, J.; Chiaroni, A. High pressure activation in the asymmetric Michael addition of chiral imines to alkyl and aryl crotonates *Tetrahedron Lett.* **2002**, *43,* 1445-1448. (b) Camara, C.; Keller, L.; Jean-Charles, K.; Joseph, D.; Dumas, F. A comparative study of high pressure versus other activation modes in the asymmetric Michael reaction of chiral imines. *Int. J. High Press. Res.* **2004**, *24*, 149-162.
6. (a) Perreux, L.; Loupy, A. Tetrahedron Report number 588, A tentative rationalization of microwave effects in organic synthesis according to the reaction medium, and mechanistic considerations. *Tetrahedron*, **2001**, *57*, 9199-9223. (b) Kappe, C. O. Controlled microwave heating in modern organic synthesis. *Angew. Chem. Int. Ed.* **2004**, *43*, 6250-6284. (c) De la Hoz, A.; Diaz-Ortiz, A.; Moreno, A. Microwaves in organic synthesis. Thermal and non-thermal microwave effects. *Chem. Soc. Rev.* **2005**, *34*, 164-178. (d) Microwave Heating as a Tool for Sustainable Chemistry; Leadbeater, N. E., Ed.; CRC Press: Boca Raton, FL, USA, 2011. (e) Kappe, C. O.; Stadler, A.; Dallinger, D. Microwaves in Organic and Medicinal Chemistry, 2nd ed.; Wiley-VCH: Weinheim, Germany, 2012. (f) Microwaves in Organic Synthesis, 3rd ed.; De La Hoz, A., Loupy, A., Eds.; Wiley-VCH: Weinheim, Germany, 2013.
7. (a) Loupy, A.; Perreux, P.; Liagre, M.; Burle, K.; Moneuse, M. Reactivity and selectivity under microwaves in organic chemistry. Relation with medium effects and reaction mechanisms. *Pure*

Appl. Chem., **2001**, *73*, 161-166. (b) De la Hoz, A.; Diaz-Ortiz, A.; Moreno, A. Selectivity in organic synthesis under microwave irradiation. *Current Org. Chem.* **2004**, *8*, 903-918. (c) Strauss, C. R.; Rooney, D. W, Accounting for clean, fast and high yielding reactions under microwave conditions. *Green Chem.*, **2010**, *12*, 1340-1344. (d) See for example: Yus, M.; Foubelo, F.; Jesús García-Muñoz, M. Stereoselective Aza-Henry Reaction of Chiral tert-Butanesulfinyl Imines with Methyl or Ethyl 4-Nitrobutanoate: Easy Access to Enantioenriched 6-Substituted Piperidine-2,5-diones. *Heterocycles* **2015**, *90*, 1419-1432.

8. (a) Langa, F.; De la Cruz, P.; De la Hoz, A.; Diaz-Ortiz, A.; Diez-Barra, E. Microwave irradiation: more than just a method for accelerating reactions. *Contemp. Org. Synth.* **1997**, *4*, 373-386. (b) Manhas, M. S.; Banik, B. K.; Mathur, A.; Vincent, J. E.; Bose, A. K. Vinyl-β-lactams as Efficient Synthons. Eco-friendly Approaches via Microwave Assisted Reactions. *Tetrahedron* **2000**, *56*, 5587-5601. (c) Camara, C.; Keller, L.; Dumas, F. Microwave activation of an asymmetric Michael reaction: unexpected behavior of chiral α-alkoxy imines. *Tetrahedron: Asymmetry* **2003**, *14,* 3263-3266. (d) Narasimhan, S.; Velmathi, S. Effect of Microwaves in the Chiral Switching Asymmetric Michael Reaction *Molecules*, **2003**, *8*, 256-262. (e) Escalante, J.; Díaz-Coutiño, F. D. Synthesis of γ-Nitro Aliphatic Methyl Esters Via Michael Additions Promoted by Microwave Irradiation. *Molecules* **2009***, 14,* 1595-1604. (f) Worzakowska, M. Thermal properties of neryl long-chain esters obtained under microwave irradiation. *J. Thermal Anal. Calor.* **2015**, 120, 1715-1722.
9. (a) Kang, J. Y.; Carter, R. G. Primary Amine, Thiourea-Based Dual Catalysis Motif for Synthesis of Stereogenic, All-Carbon Quaternary Center-Containing Cycloalkanones *Org. Lett.* **2012**, *14*, 3178-3181. (b) Horinouchi, R.; Kamei, K.; Watanabe, R.; Hieda, N.; Tatsumi, N.; Nakano, K.; Ichikawa, Y.; Kotsuki, H. Enantioselective Synthesis of Quaternary Carbon Stereogenic Centers through the Primary Amine-Catalyzed Michael Addition Reaction of α-Substituted Cyclic Ketones at High Pressure. *Eur. J. Org. Chem.* 2015, 4457-4463.
10. Pfau, M.; Revial, G.; Guingant, A.; d'Angelo, J. Enantioselective synthesis of quaternary carbon centers through Michael-type alkylation of chiral imines. *J. Am. Chem. Soc.* **1985**, *107*, 273-274.
11. 1-Phenylethylamine (99% ee) was used to direct the regio and stereoselectivity of the Michael addition and additionnaly served as a chirality marker in the product **4**, allowing determination of the diastereoselectivity of the reaction using NMR, and by consequence of the enantioselectivity of the Michael process.
12. Sarfati, M.; Lesot, P.; Merlet, D.; Courtieu, J. Theoretical and experimental aspects of enantiomeric differentiation using natural abundance multinuclear NMR spectroscopy in chiral polypeptide liquid crystals. *Chem. Commun.* **2000**, 2069-2081.
13. Meddour, A.; Berdagué, P.; Hedli, A.; Courtieu J.; Lesot, P. Proton-Decoupled Carbon-13 NMR Spectroscopy in a Lyotropic Chiral Nematic Solvent as an Analytical Tool for the Measurement of the Enantiomeric Excess. *J. Am. Chem. Soc.* **1997**, *119*, 4502-4508.
14. (a) Lesot, P.; Sarfati, M.; Courtieu, J.; Natural abundance deuterium NMR spectroscopy in polypeptide liquid crystals as a new and incisive means for the enantiodifferentiation of chiral hydrocarbons. *Chemistry* **2003**, *14*, 1724-1745. (b) Luy, B. Disinction of enantiomers by NMR spectroscopy using chiral orienting media, *J. Indian Inst. Sci.,* **2010**, *90,* 119-132. (c) Lesot, P.;

Aroulanda, C.; Zimmermann, H.; Luz, Z. Enantiotopic discrimination in the NMR spectrum of prochiral solutes in chiral liquid crystals. *Chem. Soc. Rev.*, **2015**, *44*, 2330-2375.

15. See in relation (a) Tan, K.; Alvarez, R.; Nour, M.; Cavé, C.; Chiaroni, A.; Riche, C.; d'Angelo, J. Racemization processes at a quaternary carbon center in the context of the asymmetric Michael reaction. *Tetrahedron Lett.* **2001**, *42*, 5021-5023. (b) Pizzonero, M.; Hendra, F.; Delarue-Cochin, S.; Tran Huu Dau, M.-E.; Dumas, F.; Cavé, C.; Nour, M.; d'Angelo, J. The asymmetric Michael-type alkylation of chiral β-enamino esters: critical role of a benzyl ester group in the racemization of adducts. *Tetrahedron: Asymmetry* **2005**, *16*, 3853-3857.
16. Design and evaluation of improved magnetic stir bars for single-mode microwave reactors Obermayer, D.; Damm, M.; Kappe, C. O. *Org. Biomol. Chem.* **2013**, *11*, 4949-4956.
17. Derivative spectra were compared with those in the NBS75K database (provided by Hewlett Packard with the GC/MS control and data processing software).
18. Kappe, C. O.; Pieber, B.; Dallinge, D. Microwave Effects in Organic Synthesis: Myth or Reality? Ang. Chem. Int. Ed. **2013**, *52*, 1088-1094.
19. Chemical quantum calculations related to experimental results have shown that µW effects are increasing with the asynchronous character of a mechanism. See for example: (a) Diaz-Ortiz, A.; Carrillo, J. R.; Cossio, F. P.; Gomez-Escalonilla, M. J.; De La Hoz, A.; Moreno, A.; Prieto, P. Synthesis of Pyrazolo[3,4-*b*]pyridines by Cycloaddition Reactions under Microwave Irradiation. *Tetrahedron* **2000**, *56*, 1569-1577. (b) Loupy, A.; Maurel, F.; Sabatie-Gogova, A. Improvements in Diels–Alder cycloadditions with some acetylenic compounds under solvent-free microwave-assisted conditions: experimental results and theoretical approaches. *Tetrahedron* **2004**, 60, 1683-1691. (c) Langa, F.; de la Cruz, P.; de la Hoz, A.; Espildora, E.; Cossio, P.; Lecea, B. Modification of Regioselectivity in Cycloadditions to C_{70} under Microwave Irradiation. *J. Org Chem*. **2000**, *65*, 2499-2507.
20. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. P. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.
21. Stewart, J. P. P. Quantum Chemistry Program Exchange, University of Indiana, Bloomington, U.S.A.; Program 455.
22. Bertels, R. H.; Report CNA-44, **1972**, University of Texas, Center for numerical analysis.
23. McIver, J. W.; Komornicki, A. Structure of transition states in organic reactions. General theory and an application to the cyclobutene-butadiene isomerization using a semiempirical molecular orbital method. *J. Am. Chem. Soc.* **1972**, *94*, 2625-2633.

SciForum
Mol2Net

AlzPred-SVV: Free Web Tool for Alzheimer Prediction using Spectroscopy Voxel Volume

Virginia Mato Abad [1,*], Cristian R. Munteanu [2] , Carlos Fernández-Lozano [2] and Alejandro Pazos [2]

[1] Rey Juan Carlos University, Móstoles, Spain; E-Mail: virginia.mato@urjc.es

[2] RNASA-IMEDIR Group, Faculty of Computer Science, University of A Coruna, Spain; E-Mails: crm.publish@gmail.com (C.M.), carlos.fernandez@udc.es (C.F.); apazos@udc.es (A.P.)

* E-Mail: virginia.mato@urjc.es; Tel.: +34 914888522.

Published: 4 December 2015

Abstract: Neuroimaging data from magnetic resonance techniques are widely used as non-invasive biomarkers for the evaluation and early diagnosis of Alzheimer's Disease (AD). Alzheimer Prediction by Spectroscopy Voxel Volume is a free Web tool to predict the AD diagnosis (AlzPred-SVV: http://bio-aims.udc.es/AlzPredSVV.php). The inputs are two variables related to a voxel acquired in the left hippocampus: The total volume and the volume of CSF contained in the voxel. The classification method is based on Machine Learning techniques. The tool is based on an HTML/PHP user interface with a Python/Java implementation of the model.

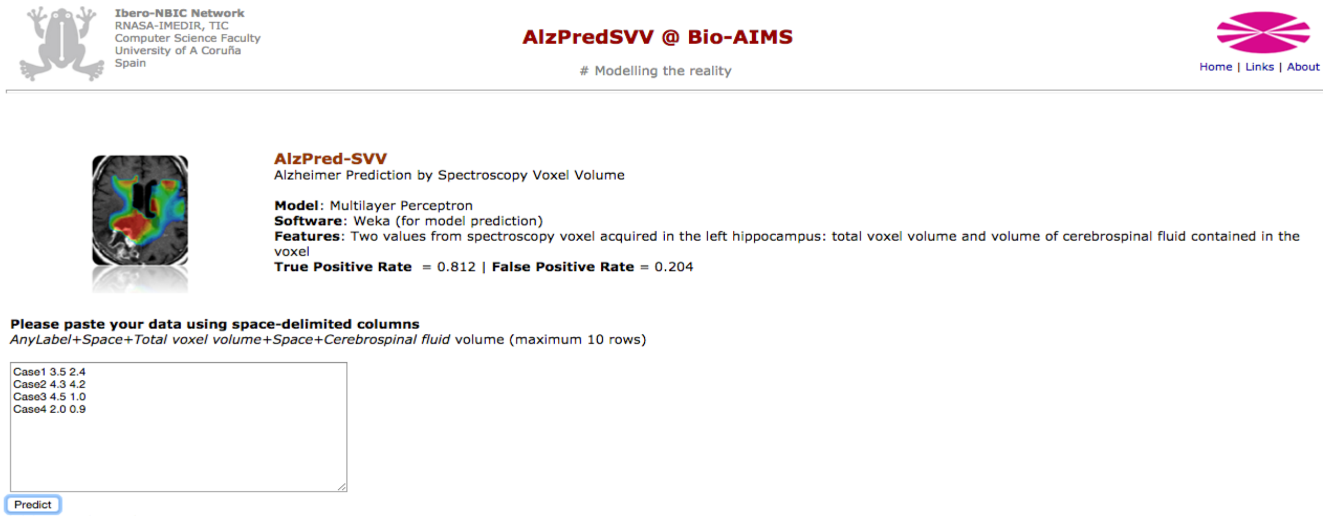
Keywords: Magnetic Resonance Spectroscopy; Alzheimer's Disease; Machine-learning; Single-Layer Perceptron;

1. Introduction

Several magnetic resonance techniques have been proposed as non-invasive imaging bio-markers for the evaluation of disease progression and early diagnosis of AD [1] and mild cognitive impairment (MCI), a transitional state between healthy ageing and AD [2]. The analysis of these biomarkers allows the study of differences between groups but they are not applicable on a single-subject level and do not improve the clinical diagnosis potential. Machine-learning techniques have been identified as promising tools in neuroimaging data for individual class prediction [3]. Magnetic resonance spectroscopy (^{1}H-MRS) is a useful technique in the study of the AD [4-5]. We found that just the volumes of grey and white matter (GM,WM) and cerebro-spinal fluid (CSF) within the spectroscopic voxel provide a high correlation with the diagnosed groups showing a strong potential for classify healthy controls, MCI and AD subjects [6].

2. Results and Discussion

Alzheimer Prediction by Spectroscopy Voxel Volume (AlzPred-SVV) is a free Web tool (bio-aims.udc.es/AlzPredSVV.php) to predict the AD diagnosis, based on only 2 variables related to the spectroscopic voxel in the left hippocampus: The total voxel volume and volume of CSF contained in the voxel (Figure 1). This tool is on the free portal Bio-AIMS [7] that offers models based on Artificial Intelligence, Computational Biology and Bioinformatics. The website provides the values of predicted class and error prediction achieved by the model. Inputs should be written using the format <*Label Total_vol CSF_vol*> up to a maximum of 10 rows. Figure 1 shows an example for 4 inputs, labelled as Case1 to Case4 varying the voxel volumes and the CSF proportions. Results are displayed in Figure 2, showing the prediction for each case: 3 inputs are classified as AD with different error predictions and the other one as healthy control.



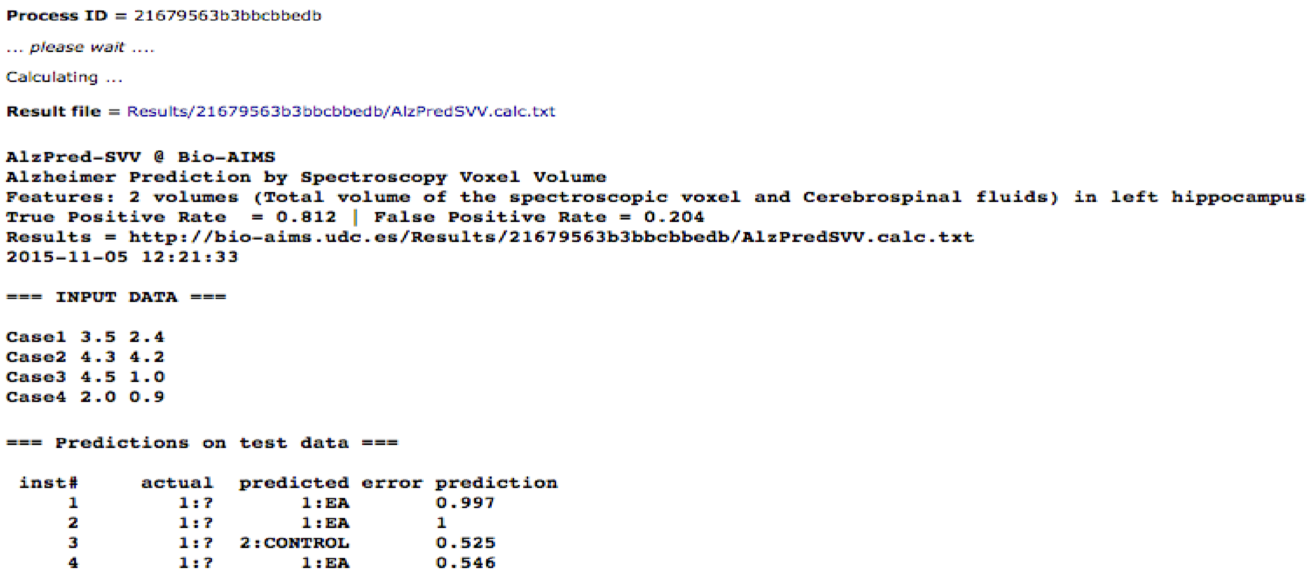
AlzPred-SVV
Alzheimer Prediction by Spectroscopy Voxel Volume

Model: Multilayer Perceptron
Software: Weka (for model prediction)
Features: Two values from spectroscopy voxel acquired in the left hippocampus: total voxel volume and volume of cerebrospinal fluid contained in the voxel
True Positive Rate = 0.812 | **False Positive Rate** = 0.204

Please paste your data using space-delimited columns
AnyLabel+Space+Total voxel volume+Space+Cerebrospinal fluid volume (maximum 10 rows)

```
Case1 3.5 2.4
Case2 4.3 4.2
Case3 4.5 1.0
Case4 2.0 0.9
```

Figure 1. AlzPred-SVV website



```
Process ID = 21679563b3bbcbbedb

... please wait ....

Calculating ...

Result file = Results/21679563b3bbcbbedb/AlzPredSVV.calc.txt

AlzPred-SVV @ Bio-AIMS
Alzheimer Prediction by Spectroscopic Voxel Volume
Features: 2 volumes (Total volume of the spectroscopic voxel and Cerebrospinal fluids) in left hippocampus
True Positive Rate  = 0.812 | False Positive Rate = 0.204
Results = http://bio-aims.udc.es/Results/21679563b3bbcbbedb/AlzPredSVV.calc.txt
2015-11-05 12:21:33

=== INPUT DATA ===

Case1 3.5 2.4
Case2 4.3 4.2
Case3 4.5 1.0
Case4 2.0 0.9

=== Predictions on test data ===

 inst#     actual  predicted error prediction
     1        1:?       1:EA       0.997
     2        1:?       1:EA       1
     3        1:?  2:CONTROL       0.525
     4        1:?       1:EA       0.546
```

Figure 2. Output results from AlzPred-SVV

3. Materials and Methods

A gender-matched cohort of 260 subjects was used to test and evaluate the effectiveness of machine-learning schemes for single-subject level classification of individuals affected by different stages of dementia based on ^{1}H-MRS data [6]. The collection of Weka algorithms was used for this purpose. The study found that the best classifier is a single-layer perceptron with only 2 spectroscopic voxel volumes in the left hippocampus (AUROC:0.86; True positives rate: 0.81; False positives rate:0.20). This model was implemented in AlzPred-SVV. The tool is based on an HTML/PHP user interface with a Python/Java implementation of the model.

4. Conclusions

MR modalities produce extremely high-dimensional raw data that can contain inherent patterns related to AD and machine-learning methods provide tools to observe inherent disease-related patterns in the data. This fact is presented in this work, where just the proportion of CSF within the spectroscopic voxel can discriminate AD from MCI patients and from healthy controls. AlzPred-SVV is an easy-to-use web application that can be useful for both clinicians and patients.

Acknowledgments

This work is supported by the "Collaborative Project on Medical Informatics (CIMED)" PI13/00280 funded by the Carlos III Health Institute from the Spanish National plan for Scientific and Technical Research and Innovation 2013-2016 and the European Regional Development Fund (ERDF). The authors thank the DEMCAM project team, a multicentre study with patients from different centres for dementia in the Autonomous Region of Madrid.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Li T.Q.; Wahlund L.O. The search for neuroimaging biomarkers of Alzheimer’s disease with advanced MRI techniques. *Acta Radiology* **2011,** 211–222.
2. Grundman M.; et al. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Archives of Neurology* **2004** 61(1), 59-66.
3. Falahati F.; Westman E.; Simmons A. Multivariate data analysis and machine learning in alzheimer's disease with a focus on structural magnetic resonance imaging. *Journal of Alzheimer's Disease* **2014** 41(3), 685-708.
4. Frederick B.D.; et al. In vivo proton magnetic resonance spectroscopy of the temporal lobe in alzheimer's disease. *Progress in neuro-psychopharmacology & biological psychiatry* **2004**, 28(8), 1313-1322.
5. Ross A.J.; Sachdev P.S. Magnetic resonance spectroscopy in cognitive research. *Brain research. Brain research reviews*, **2004** 44(2-3), 83-102.
6. Munteanu R.C.; et al. Classification of mild cognitive impairment and Alzheimer’s disease with machine-learning techniques using 1H Magnetic Resonance Spectroscopy data. *Expert Systems with Applications.* **2015** 42, 6205-6214.

7. Bio-AIMS. Artificial Intelligence Model Server in Biosciences. Available online: http://bio-aims.udc.es (accessed on 05 November 2015).

SciForum
Mol2Net

Synthesis and Characterization of Carbon Nanotube/Hydroxyapatite/Clay Based Hybrid Antimicrobial Biomaterial for Potential Tissue Engineering Application

Subrata Kar[1,2], Papiya Nandy[2], Ruma Basu[2,3] and Sukhen Das[1,2,4,*]

[1]Physics Department, Jadavpur University, Kolkata-700 032, India.
[2]Centre for Interdisciplinary Research and Education, Kolkata 700 068, India
[3]Physics Department, Jogamaya Devi College, Kolkata 700 026
[4]Indian Institute of Engineering, Science and Technology, Howrah, India
* Author to whom correspondence should be addressed

Published: 4 December 2015

Abstract: Inorganic ceramic materials have recently been enjoying a great deal of attention for uses as biomaterial over the traditionally used polymeric material. The ability of silica based ceramics to release Si-containing ionic products makes them osteoconductive and their special surface composition has drawn considerable attention for its use as osteogenic proliferation, differentiation and gene expression of tissue cells. Here Pristine Multiwalled carbon nanotube (MWCNT) was functionalized by acid oxidation and used as template for nucleation of hydroxyapatite crystal. The CNT-HAP thus formed was further reinforced with exfoliated Montmoriollite(MMT) clay to increase the mechanical property of the nanocomposite. The synthesized composite viz CNT-HAP (HC) and CNT(HAP)-MMT(CHC) were characterized by fourier transform infrared spectroscopy(FTIR), X-ray diffraction(XRD), Scanning electron microscopy and transmission electron microscopy(TEM) and thermogravimetric analysis(TGA) . The biomedical application of CHC, blood compatibility was studied in terms of hemolysis assay and platelet adhesion assay. Finally, in vitro dissolution assay and the antibacterial activity of the material were further explored for establishing its use as an antibacterial biomaterial. The nanocomposite thus designed exhibited a promising candidate as an antimicrobial biomaterial for tissue Engineering application..

Keywords: Clay; Carbon nanotube; Hydroxyapatite; Nanocomposites; Hemocompatibility; Biomineralization; Antimicrobial biomaterial

1. Introduction

Clays and clay minerals are widely utilized in our society are in geology, agriculture, construction, environmental applications and in Traditional applications that include ceramics, paper, paint, plastics, chemical carriers, decolorization, and catalysis etc(1). Biomedical application of clay dates back to prehistoric era where clay served in wounds and skin irritations caused by *Homo erectus* and *H. neanderthalensis*. Clays were ingested to treat stomach and intestinal problems. Clay application finds its application as active ingredients in antacids, antidiarrhoerics, tropical applications such as cosmetic creams, powders and in pharmaceutics as excipients(2). Precisely the rich electrochemistry of clay mineral attributes to its uses in human health and disease. Montmorillonite is a layered aluminosilicate, belongs to smectite group of clay which consists of a one edge shared octahedral sheet of aluminum hydroxide fused in between two silica tetrahedral(3). The susceptibility of the charged smectite particles for swelling and delamination, results in rich selection of potential interactions between organic molecules and the clay particles. Inorganic ceramic materials have been enjoying a great deal of attention for uses as biomaterials over the traditionally used polymeric material. In the late 1960's Ceramics material replaced the earlier used metals for implants(2). Clay particles have also been implemented in dental adhesives for improving bond strength. MMT particles have also been used as scaffold in bone therapies. The ability of silica based ceramics to release Si-containing ionic products makes them osteoconductive and their special surface composition has drawn considerable attention for its use as osteogenic proliferation, differentiation and gene expression of tissue cells(4). Hydroxyapatite (HAp, $Ca_{10}(PO_4)_6(OH)_2$) ceramics has served as an excellent biomaterial due to their close chemical composition with the bone and in biomedical field as drug delivery vehicle. The ability of HAP to form chemical bond with host tissues offers a great advantage over allografts and metal implants. The poor mechanical property has lead to the discovery of HAP based composites with improved mechanical properties(5).

Carbon nanotube (CNT) is an allotrope of carbon received a significant curiosity pertaining to its excellent physical, chemical and mechanical properties(6). However there has been commendable use of CNT in biology and medicine in terms of pharmacy, medicine, drug delivery vehicle, biosensor and tissue engineering applications. However the darker side of CNT regarding its application goes to the toxicity of CNT depends on several intrinsic and environmental factors such as surface charge and modification, length, agglomeration etc. The conversion of the pristine CNT into its soluble form renders biocompatibility and reduces toxicity for biomedical applications(7,8). Toxicological reports of single walled CNT(SWCNT) and multiwalled CNT(MWCNT) suggested lower toxicity of the later, thus making it a more attractive material for biomedical application.

Till date there has been no attempt in putting together CNT, HAP with MMT clay in a single composite system. In our work, CNT functionalized by acid oxidation served as template for nucleation of hydroxyapatite. The CNT-HAP thus formed was further reinforced with exfoliated MMT clay to further increase the mechanical property of the material. The synthesized composite viz CNT-HAP (HC) and CNT(HAP)-MMT(CHC) were characterized by fourier transform infrared spectroscopy(FTIR), X-ray diffraction(XRD), Scanning electron microscopy and transmission electron microscopy(TEM).

To evaluate the biomedical application of CHC, blood compatibility was studied in terms of hemolysis assay and platelet adhesion assay. In vitro dissolution assay was further studied for exploring its use as potential bone tissue engineering application.

2. Results and Discussion

2.1. XRD analysis

The X ray diffraction pattern of HAP(*f*-MWCNT) and MMT-CNT(HAP) abbreviated as HC as CHC respectively is shown in the figure 1(a,b). The diffraction peaks of HC represent the prominent reflections of hydroxyapatite. The peak at 25.86^0 represents the hexagonal graphite plane was consistent with the HAP plane(9). The diffraction pattern of CHC shows the peak at 7.11^0 representing 001 plane of MMT clay. The peak of the HAP and CNT in the diffraction pattern is subdued due to large intensity of the diffraction 001 plane of clay.

FESEM

The scanning electron micrographs shows the entangled reticulation of MWCNT with HAP clusters for HC(Figure 3a). The FESEM image of CHC shows the distribution of the HAP and reticulated CNT in clay particles (Figure 3b). The TEM images show all three components ie MMT clay, HAP and MWCNT distributed in the matrix(Figure 3c).

.

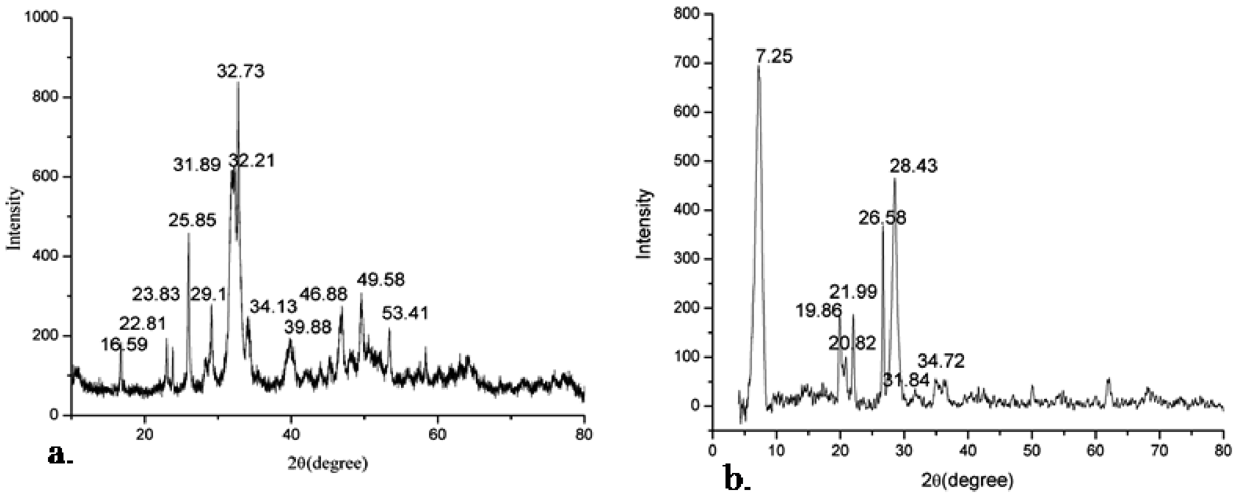


Figure 1: XRD of (a) HC and (b)CHC

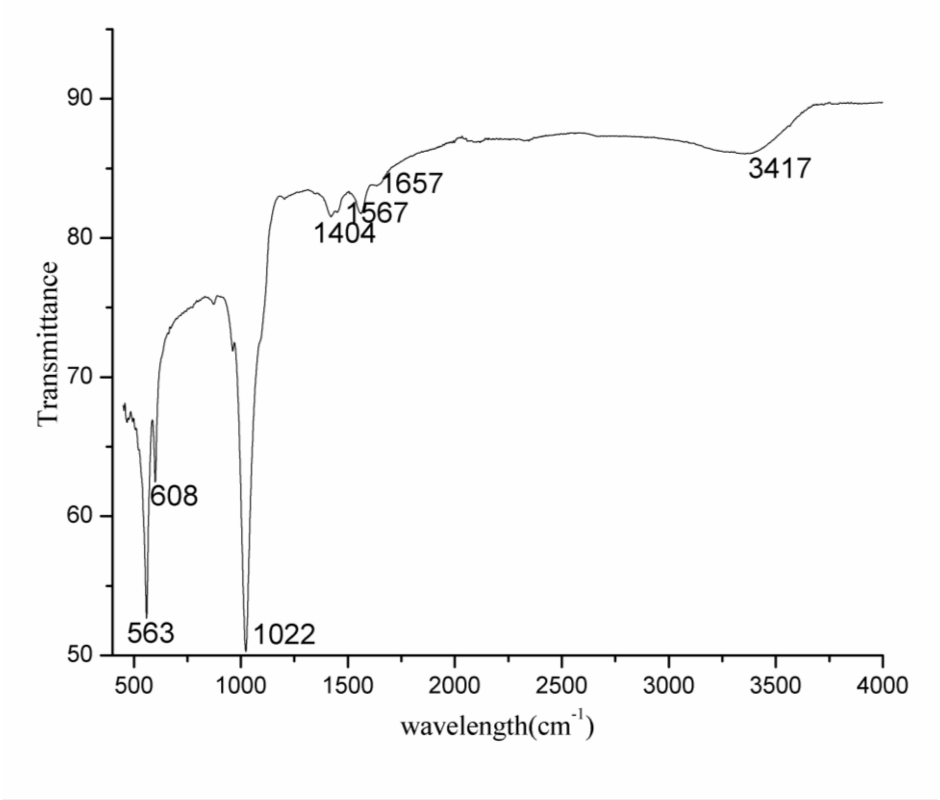


Figure 2: FTIR of HC.

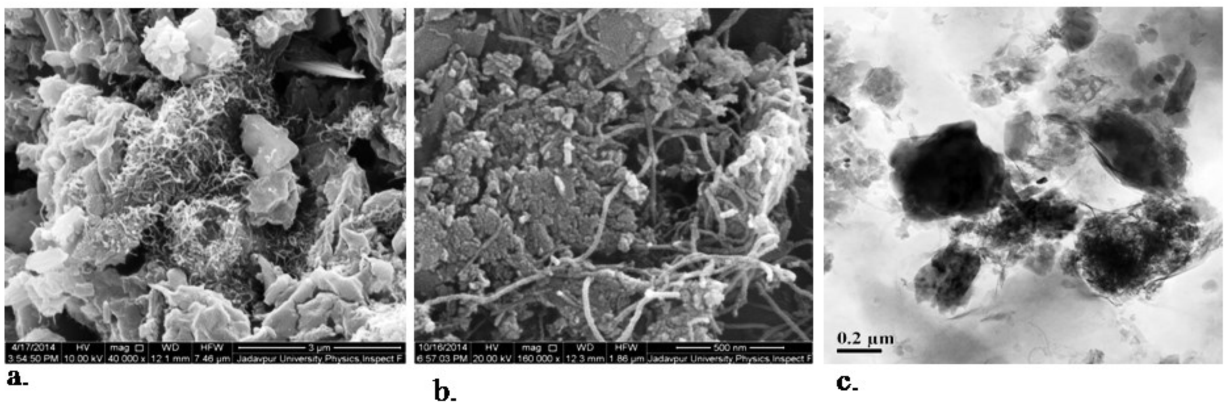


Figure 3: FESEM of (a) HC and (b) CHC. HRTEM of CHC nanocomposite(c)

Table 1. Percentage release of hemoglobin from red blood cells after 60 min incubation with different concentrations of CHC at 37^0 C.

CHC (mg/ml)	**.1**	**.5**	**1**	**3**	**6**	**8**	**10**	**15**	**20**
% Hemolysis	.92	1.76	4.37	4.43	4.51	4.58	4.72	5	9.16

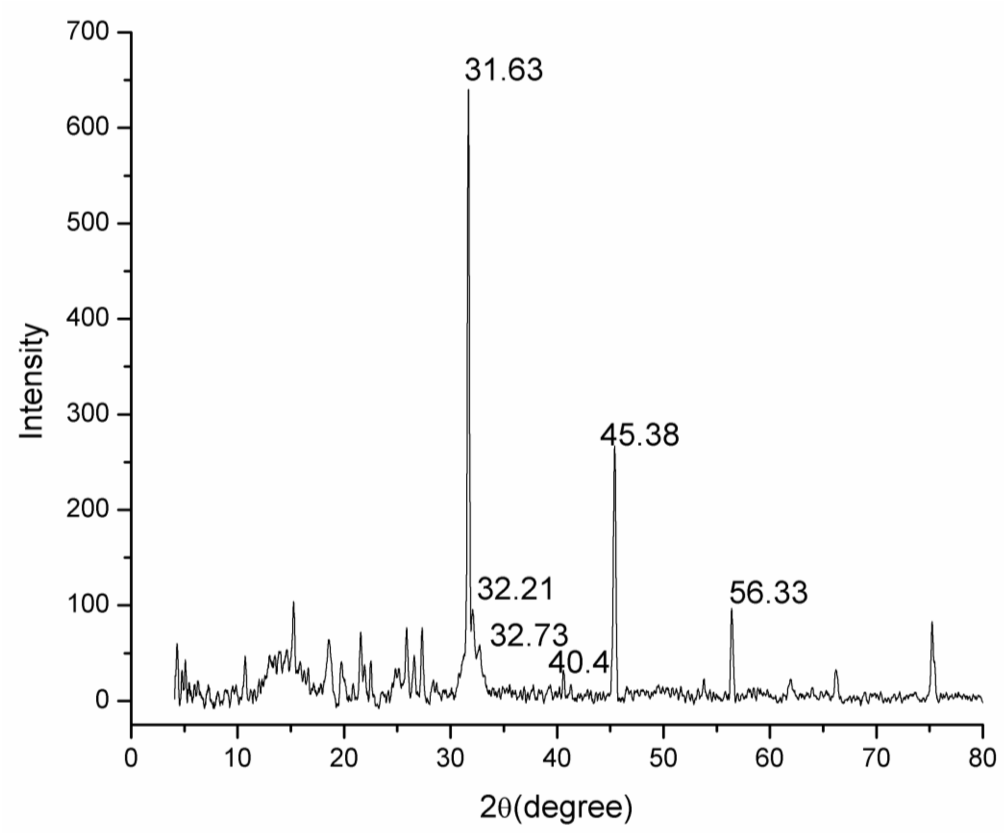


Figure: XRD Pattern of CHC nanocomposite after 30 days of immersion in SBF.

2.2.

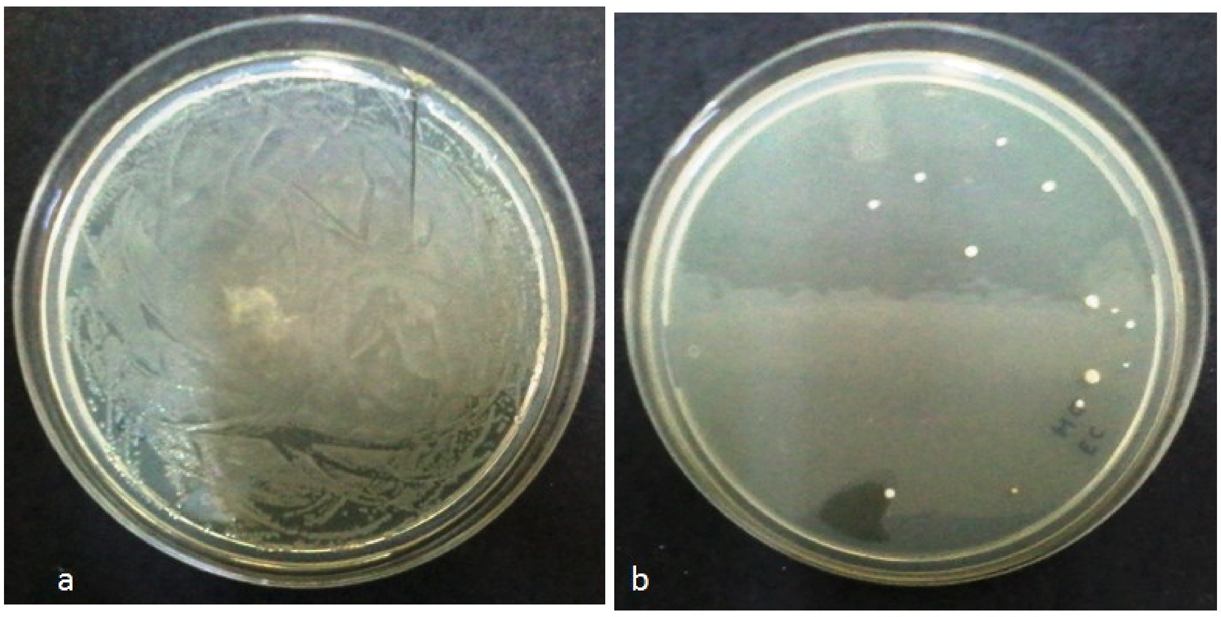


Figure 5: Nutrient Agar plates of E.coli after 24 h (a)control (b)CHC treated.

2.3.FTIR

The FTIR pattern of the HC is represented in Figure 2. The FTIR spectrum of the MWCNT–HAP sample is shown in Fig. 1.The characteristic absorption peaks of the phosphate groups for HAP are observed at 565 cm^{-1}, 611 cm^{-1} and 1033 cm^{-1}, which are attributed to the P-O bond of PO4 $^{3-}$ stretching vibration and the corresponding deformation vibration(9). Small peak appearing at 1400 cm^{-1} is possibly associated with O–H bending deformation in carboxylic acid groups (10). The peak at 1565 cm^{-1} is related to the carboxylate anion stretch mode(11). The peak at 3280–3675 cm^{-1} and 1640–1660 cm^{-1} represent stretch and bend the as-received tubes result from O–H which are assigned to –O–H groups of adsorbed water or covalently bonded functional groups(12).

Determination of Hemolytic activity

The hemolytic activity of the CHC composite in different concentration range is shown in table 1. The nanocomposite shows hemocompatible within 15 mg/ml concentration range showing hemolysis less than 5%(13). Carbon naotube and montmoriollite clay composites have previously been reported as hemocompatible(14,15)

.

.

The platelet adhesion assay was performed using three concentration of CHC, i.e 3,6,15 mg/ml. The value of the platelet adhesion percentage was 12.03, 15.04 and 26.3% respectively, indicating the CHC nanocomposites were nonthombogenic within the concentration range.

2.4. Phase analysis

XRD patterns of the CHC nanocomposite after soaking in SBF is shown in figure 4. The crystallinity of HA is observed to increase significantly after 4 weeks of incubation(16) . The result indicates the formation of apatite on the nanocomposite surface suggesting the bioactivity of the synthesized biomaterials(17).

Antibacterial activity

The nanocomposites show an excellent antibacterial activity against E. coli DH5α (MTCC 1652). The antibacterial nature of the material was mainly due to the presence of MWCNT and MMT clay. The mechanism of antibacterial activity of CNT is associated with their diameter-dependent piercing and length-dependent wrapping on the lysis of microbial walls and membranes, inducing release of intracellular components DNA and RNA and allowing a loss of bacterial membrane potential, demonstrating complete destruction of bacteria(18). The presence of MMT clay further escalates the antibacterial effect of the MWCNT by means of direct interaction of clay with MWCNT leading to antibacterial action by contact inhibition(19).

3. Experimental

3.1. Materials

Montmorillonite clay (MMT) (Na^{+} exchanged) (Nanocor Inc., USA), MWCNTs Arry International Germany, The purity of MWCNT was 60% and its diameter was 30 nm, Calcium chloride ($CaCl_2$), Phosphoric acid (H_3PO_4), Sodium chloride(NaCl), Sodium bicarbonate($NaHCO_3$), Potassium chloride (KCl), Dipotassium hydrogen phosphate trihydrate($K_2HPO_4.3H_20$), Magnesium chloride($MgCl_2.6H_2O$), Hydrochloric acid, Nitric acid (69%), sulphuric acid (98%) Sodium sulphate(Na_2SO_4), Tris, Sodium citrate dehydrate, citric acid anhydrous, D- glucose from merck, India.

3.2. Formation of MMT-CNT(HAP)

3.2.1. Functionalization of carbon nanotubes with carboxylic groups

Functionalization of MWCNT by oxidation was done according to the method followed by Yi et al, 2006 with brief modifications. Briefly 100 mg of MWCNT was refluxed for 10h in a mixture acid solution of Sulphuric acid and Nitric acid (3:1 v/v). The system was allowed to cool down to room temperature and sonicated in a bath sonicator for 3h. The dispersion was washed several times with ethanol and water until PH~7 is reached. Finally f-MWCNT was dried in oven at 70^0C.

3.2.2 Synthesis of HAP(*f*-MWCNT)

CNT sol was prepared by dispersing a small amount of prepared MWCNT in deionised water followed by sonication for 2h. To the sol, 0.5M of calcium chloride and phosphoric acid were separately added very slowly maintaining the Ca/p ratio to 1.67. The mixture was stirred for 1h. The pH of the system was maintained to ~9 by 1N sodium hydroxide. After ageing for 24h the product was dried in oven at 80^0C(20).

3.2.3. Formation of MMT clay-HAP(*f*-MWCNT)

Na-MMT clay were dispersed and stirred overnight in deionised water for allowing the clay layers to swell. To the dispersion HAP(*f*-MWCNT) was added and allowed to mix in a magnetic stirrer followed by sonication in bath sonicator. The composite were then oven dried and finally crushed to fine powder for characterization and further studies.

3.2.4. X-ray diffraction (XRD)

X-ray diffraction (XRD) Powder X-ray diffraction (XRD) patterns were recorded using a Bruker AXS (Model D8, WI, USA) setup with CuKα radiation (1.5409 Å) and scan speed of 5 min^{-1} and scanning range from 5^0 to 80^0 (2θ).

3.2.5. Fourier transform infrared spectroscopy (FTIR)

Fourier transform infrared spectroscopy was performed by FTIR- 8400S model Shimadzu, Tokyo. Samples were prepared by KBr disk method, in which 0.2 g of KBr (spectroscopy grade) was thoroughly mixed with sintered sample powder (1% by weight of KBr) and then made into disks by uniaxial pressing. Scanning range was set from 400 to 2000 cm^{-1} under Happ–Genzel configuration.

3.2.6. Electron microscopic study

Morphological characteristics of the samples were observed by scanning electron microscope (SEM) model FEI Quanta 250 (USA). A minute quantity of the sample was directly place on carbon coated grid, sputter coated with gold and then observed SEM.

3.2.7. Transmission electron microscopy

The particle size of synthesized nanocomposites was observed by JEM- 2100 HRTEM model. A minute quantity of the sample was dispersed in water by sonication and observed under the microscope.

3.2.8. Hemocompatibility

Platelet adhesion analysis

The assay was done according to the method followed by (21)Sun et al, 2014. Whole mice blood (citrated) was centrifuged at 1500 rpm for 15 min to obtain platelet-rich plasma (PRP) supernatant. Samples were equilibrated with normal saline at 37^0 C for 2 h. 1 or 1.5 mL fresh PRP at a density of 5 X 10^5 cells/mL (N1) was added in a 2-mL tube), incubated at 37^0 C for 3 h, then removed. The numbers of platelets in PRP not adhered to the specimens were recorded using a cell counter (N2). The platelet adhesion ratio was calculated using the following equation:

$$\text{Platelet adhesion (\%)} = \frac{N1-N2}{N1} \times 100$$

Hemolysis analysis

Test specimens in 2 mL centrifuge tubes were equilibrated in saline (0.9% (w/v) NaCl) at 37^{O} C for 30 min. Whole blood from healthy mice was collected into sterile sodium citrate buffer, and diluted (0.2 mL in 10 mL saline).Equal amount of the diluted blood was added to the test specimen. Distilled water and physiological saline were used as negative control (N) and positive control (P), respectively. Samples were placed in a static incubator at 37^0 C for a further 60 min. After hemolysis, samples were centrifuged at 2500 rpm for 5 min(21)(Sun et al, 2014). The absorbance of the supernatant was measured at 545 nm. The hemolysis percentage was calculated according to the following equation:

$$Hemolysis\ (\%) = \frac{A2-A1}{A3-A1} \times 100$$

where A1, A2, and A3 are the absorbance of the negative control, sample, and positive control, respectively.

3.2.9. Biomineralization

The test samples bioceramic was immersed into a 1.5-times concentrated simulated body fluid (SBF) at 37^0 C up to 7 days, and the SBF solution was change every 24h. The material was removed from the SBF solution after 7 day incubation, gently rinsed with water, and air dried at room temperature. The phases present in the coating were determined by XRD analysis(16,22).

3.2.10. Antibacterial activity

The antibacterial activity of the synthesized nanocomposites was performed against *E.coli* cells. Briefly overnight grown bacterial cell in nutrient broth representing 10^7 CFU/ml was washed thrice with PBS 7.4 at 5000rpm for 10 mins. This was followed by incubating 10^5 cells in a fresh nutrient broth with 10 mg of CHC and HC and incubated at 37^0 C. Untreated control served as control. After overnight incubation the culture was plated in nutrient agar plate(1.8%) and further incubated overnight.

4. Conclusion

In summary, the synthesized nanocomposite MMT clay, HAP and oxidized MWCNT show a good hemocompatibility, showing minimum hemolysis and platelet adhesion which increases in a concentration dependent manner. Further the material was found to be bioactive, showing apatite formation after immersion in stimulating biological fluid(SBF) after 30 days of immersion. The material exhibited an excellent antibacterial activity. Thus the synthesized nanocomposite may act as a good candidate as an antibacterial biomaterial for biomedical application.

Acknowlwdgement

We are grateful to the Department of Science and Technology, Govt. of India for financial assistance.

Conflicts of Interest

The authors declare no conflict of interest.

.

References and Notes

1. Murray, H. H. (2000). Traditional and new applications for kaolin, smectite, and palygorskite: a general overview. Applied clay science, 17(5), 207-221.Jonathan I. Dawson and Richard O. C. Oreffo, Adv. Mater. 2013, 25, 4069–4086.
2. Dawson, J. I., & Oreffo, R. O. (2013). Clay: new opportunities for tissue regeneration and biomaterial design. *Advanced Materials*, *25*(30), 4069-4086.
3. Katti, K. S., Katti, D. R., & Dash, R. (2008). Synthesis and characterization of a novel chitosan/montmorillonite/hydroxyapatite nanocomposite for bone tissue engineering. *Biomedical Materials*, *3*(3), 034122.
4. Mieszawska, A. J., Fourligas, N., Georgakoudi, I., Ouhib, N. M., Belton, D. J., Perry, C. C., & Kaplan, D. L. (2010). Osteoinductive silk–silica composite biomaterials for bone regeneration. *Biomaterials*, *31*(34), 8902-8910.
5. Liu, D. M., Troczynski, T., & Tseng, W. J. (2001). Water-based sol–gel synthesis of hydroxyapatite: process development. *Biomaterials*, *22*(13), 1721-1730.
6. Hirata, E., Uo, M., Takita, H., Akasaka, T., Watari, F., & Yokoyama, A. (2011). Multiwalled carbon nanotube-coating of 3D collagen scaffolds for bone tissue engineering. *Carbon*, *49*(10), 3284-3291.
7. Firme, C. P., & Bandaru, P. R. (2010). Toxicity issues in the application of carbon nanotubes to biological systems. *Nanomedicine: Nanotechnology, Biology and Medicine*, *6*(2), 245-256.
8. Liu, Y., Zhao, Y., Sun, B., & Chen, C. (2012). Understanding the toxicity of carbon nanotubes. *Accounts of Chemical Research*, *46*(3), 702-713.
9. Liu, Z., Chen, L., Zhang, Z., Li, Y., Dong, Y., & Sun, Y. (2013). Synthesis of multi-walled carbon nanotube–hydroxyapatite composites and its application in the sorption of Co (II) from aqueous solutions. *Journal of Molecular Liquids*, *179*, 46-53.
10. Goyanes, S., Rubiolo, G. R., Salazar, A., Jimeno, A., Corcuera, M. A., & Mondragon, I. (2007). Carboxylation treatment of multiwalled carbon nanotubes monitored by infrared and ultraviolet spectroscopies and scanning probe microscopy. *Diamond and related materials*, *16*(2), 412-417.
11. Abuilaiwi, F. A., Laoui, T., Al-Harthi, M., & Atieh, M. A. (2010). Modification and functionalization of multiwalled carbon nanotube (MWCNT) via fischer esterification. *Arabian Journal for Science and Engineering*, *35*(1C), 37-48.
12. Osswald, S., Havel, M., & Gogotsi, Y. (2007). Monitoring oxidation of multiwalled carbon nanotubes by Raman spectroscopy. *Journal of Raman Spectroscopy*, *38*(6), 728-736.
13. Wang C, Wang S, Li K, Ju Y, Li J, et al. (2014) Preparation of Laponite Bioceramics for Potential Bone Tissue Engineering Applications. PLoS ONE 9(6):e99585.
14. Zhou, N., Fang, S., Xu, D., Zhang, J., Mo, H., & Shen, J. (2009). Montmorillonite–phosphatidyl choline/PDMS films: a novel antithrombogenic material. *Applied Clay Science*, *46*(4), 401-403.
15. Murugesan, S., Park, T. J., Yang, H., Mousa, S., & Linhardt, R. J. (2006). Blood compatible carbon nanotubes-nano-based neoproteoglycans. *Langmuir*,*22*(8), 3461-3463.

16. Gu, Y. W., Khor, K. A., & Cheang, P. (2003). In vitro studies of plasma-sprayed hydroxyapatite/Ti-6Al-4V composite coatings in simulated body fluid (SBF). *Biomaterials*, *24*(9), 1603-1611.
17. Zadpoor, A. A. (2014). Relationship between in vitro apatite-forming ability measured using simulated body fluid and in vivo bioactivity of biomaterials.*Materials Science and Engineering: C*, *35*, 134-143.
18. Chen, H., Wang, B., Gao, D., Guan, M., Zheng, L., Ouyang, H., ... & Feng, W. (2013). Broad-Spectrum Antibacterial Activity of Carbon Nanotubes to Human Gut Bacteria. *Small*, *9*(16), 2735-2746.
19. Bagchi, B., Kar, S., Dey, S. K., Bhandary, S., Roy, D., Mukhopadhyay, T. K., ... & Nandy, P. (2013). In situ synthesis and antibacterial activity of copper nanoparticle loaded natural montmorillonite clay based on contact inhibition and ion release. *Colloids and Surfaces B: Biointerfaces.*
20. Liao, S., Xu, G., Wang, W., Watari, F., Cui, F., Ramakrishna, S., & Chan, C. K. (2007). Self-assembly of nano-hydroxyapatite on multi-walled carbon nanotubes. *Acta Biomaterialia*, *3*(5), 669-675.
21. Sun, D., Hao, Y., Yang, G., & Wang, J. (2015). Hemocompatibility and cytocompatibility of the hirudin-modified silk fibroin. *Journal of Biomedical Materials Research Part B: Applied Biomaterials*, *103*(3), 556-562.
22. Wang, X., Zhou, N., Yuan, J., Wang, W., Tang, Y., Lu, C., ... & Shen, J. (2012). Antibacterial and anticoagulation properties of carboxylated graphene oxide–lanthanum complexes. *Journal of Materials Chemistry*, *22*(4), 1673-1678.

SciForum
Mol2Net

Using the RRegrs R package for Automating Predictive Modelling

Georgia Tsiliki [1,*], Cristian R Munteanu [2], Jose A Seoane [3], Carlos Fernandez-Lozano [2], Haralambos Sarimveis [1] and Egon L Willighagen [4]

[1] School of Chemical Engineering, National Technical University of Athens, 15780, Greece; E-Mails: gtsiliki@central.ntua.gr (G.T.); hsarimv@central.ntua.gr (H.S.)

[2] RNASA-IMEDIR Group, Computer Science Faculty, University of A Coruna, 15071 A Coruña, Spain; E-Mails: crm.publish@gmail.com (CR.M.); carlos.fernandez@udc.es (C.F.-L.)

[3] Stanford Cancer Institute, Stanford University, C. J. Huang Building, 780 Welch Road, Palo Alto, CA 94304, USA; E-Mail: seoane@stanford.edu

[4] Department of Bioinformatics‐BiGCaT, NUTRIM, Maastricht University, P.O. Box 616, UNS50 Box 19, 6200 MD Maastricht, The Netherlands.; E-Mail: egon.willighagen@gmail.com

* Author to whom correspondence should be addressed; E-Mails: gtsiliki@central.ntua.gr; Tel.: +30-210-7723-236; Fax: +30-210-7723-138.

Published: 4 December 2015

Abstract: Cheminformatics and bioinformatics are extensively using predictive modelling and exhibit a need for standardization of methodologies such as data splitting, cross-validation methods, best model criteria and Y-randomization. RRegrs is a new R package, available at https://www.github.com/enanomapper/RRegrs (0.05 release), which suggests an integrated framework to assist model selection and speed up the process of predictive model development. The tool proposes a fully validated scheme by employing repeated 10-fold and leave-one-out cross-validation for ten linear and non-linear regression methods. Standardized reports are produced to compare the output of modelling algorithms and assess cross-validation results for selected models. Here, we demonstrate RRegrs capabilities in terms of performance using five well-established data sets.

Keywords: Multiple regression; QSAR; cross-validation; model selection

1. Introduction

RRegrs introduces an integrated framework for producing reliable and fully validated regression models in an automated way [1]. In its current release 0.05 (DOI:

10.5281/zenodo.32580), ten simple and complex regression methods are implemented, particularly: Multiple Linear regression (LM), Generalized Linear Model with Stepwise Feature Selection (GLM), Partial Least Squares regression (PLS), Lasso regression, Elastic Net regression (ENET), Support vector machine using radial functions (SVRM), Neural Networks regression (NN), Random Forest (RF), Random Forest-Recursive Feature Elimination (RF-RFE) and Support Vector Machines Recursive Feature Elimination (SVM-RFE). The methodology was implemented as an open source R package, available at https://github.com/enanomapper/RRegrs, by reusing and extending on the caret R package [2].

A single RRegrs function call is needed to run the entire workflow and obtain the produced validated models in a reproducible format. RRegrs suggests an easy way to explore the models' search space of linear and non-linear models with special parameters specifications and cross-validation (CV) schemes. Furthermore, model outputs are easily accessible and readable, organized by methods, centralized and averaged by multiple reproducible data set splits. Summary files are also produced helping the user to easily access all methodologies results, which can then be prioritized based on various statistics. A main feature of the package is its exhaustive validation scheme which introduces multiple random data splits. For each algorithm and data split, the model is produced based on training and validation sets, however, the test set is used to select the final best model. Parallel processing is enabled for accelerating the process.

2. Results and Discussion

Although the primary applications of RRegrs are aimed at finding Quantitative Structure – Activity Relationships (QSAR) models [3] under the settings of cheminformatics and nanotoxicology, here we demonstrate its efficiency for five standard data sets from UC Irvine Machine Learning Repository [4], using RRegrs current release 0.05. The five data sets considered, which are derived from diverse disciplines such as environmental economics and medical research, are the Housing [5], Computer Hardware, Wine Quality [6], Automobile [7] and Parkinsons Telemonitoring [8] data sets.

In Table 1 we present two statistic values for the five data sets, namely the R^2_{Test} and $\text{RMSE}_{\text{Test}}$ values, averaged over 10 different data splits and employing 10-fold repeated CV and 10 Y-randomizations. For all data sets, advanced methods such as RF-RFE and RF give the highest R^2_{Test} values. PLS is providing the poorest results in terms of both R^2_{Test} and $\text{RMSE}_{\text{Test}}$ values, whereas LM, GLM and LASSO are performing better in all cases but the Parkinson Telemonitoring data set. Very low $\text{RMSE}_{\text{Test}}$ values are observed, for instance SVRM method exhibits low $\text{RMSE}_{\text{Test}}$, although the corresponding R^2_{Test} values are generally lower compared to alternative methods.

Table 1. Averaged R^2_{Test} and $RMSE_{Test}$ values for the five data sets.

RRegrs	**Housing**		**Computer h/w**		**Red wine**		**Automobile**		**Parkinson t/m**	
method	R^2_{Test}	$RMSE_{Test}$	R^2_{Test}	$RMSE_{Test}$	R^2_{Test}	$RMSE_{Test}$	R^2_{Test}	$RMSE_{Test}$	R^2_{Test}	$RMSE_{Test}$
LM	0.707	0.111	0.822	0.056	0.355	0.131	0.824	0.085	0.154	0.217
GLM	0.709	0.111	0.825	0.056	0.353	0.131	0.824	0.085	0.153	0.217
PLS	0.660	0.120	0.793	0.064	0.331	0.133	0.784	0.098	0.121	0.221
LASSO	0.704	0.112	0.828	0.055	0.354	0.131	0.831	0.084	0.154	0.217
ENET	0.706	0.112	0.825	0.056	0.355	0.131	0.828	0.085	0.154	0.217
SVRM	0.845	0.080	0.765	0.066	0.396	0.127	0.853	0.075	0.637	0.142
NN	0.844	0.081	0.882	0.043	0.367	0.130	0.795	0.095	0.535	0.161
RF	0.874	0.074	0.909	0.045	0.501	0.115	0.915	0.059	0.972	0.040
RF-RFE	0.876	0.074	0.894	0.046	0.503	0.115	0.915	0.058	0.900	0.084
SVMRFE	0.717	0.120	0.692	0.124	0.378	0.129	0.728	0.151	0.479	0.173

3. Materials and Methods

In order to run RRegrs with full functionality a call to the RRegrs() function is required. All parameters have default values; a detailed list of parameters and functions' descriptions is given in the RRegrs package tutorial available online at https://github.com/enanomapper/RRegrs/blob/master/RRegrs-package-tutorial.pdf. Within the default values a default location for the output files is set, execution of all modelling steps (removal of NA, and near zero variance features, and of correlated features), normalization of the data set, ten splits, ten Y-randomization steps, and running of all ten regression methods. RRegrs function calls can be integrated into complex desktop and web tools for QSAR modelling.

A simple call to the function for a data set file named "MyDataSet.csv" and an output repository "MyResultsFolder" is the following:

```
>library(RRegrs)
>RRegrsResults<-  RRegrs(DataFileName="MyDataSet.csv",
PathDataSet="MyResultsFolder")
```

4. Conclusions

RRegrs integrates results of individual models and decides on the best model given the data set and the user specified parameters. We have demonstrated its performance with five well-established data sets and showed that good performance results are produced in all cases. Its efficiency suggests that RRegrs can be used as a reliable fully-validated and automated predictive modelling framework, and a baseline for comparable results across various studies.

Acknowledgments

This work was supported by the eNanoMapper project, funded by the European Union's Seventh Framework Programme for research, technological development and demonstration (FP7 NMP-2013-SMALL-7) under grant agreement no 604134. The authors acknowledge the support by the Galician Network of Drugs R+D REGID (Xunta de Galicia R2014/025) and by "Collaborative Project on Medical Informatics (CIMED)" P I 13/00280 funded by the Carlos III Health Institute from the

Spanish National plan for Scientific and Technical Research and Innovation 2013–2016 and the European Regional Development Funds (FEDER).

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Tsiliki, G.; Munteanu, C.R.; Seoane, J.A.; Fernandez-Lozano, C.; Sarimveis, H.; Willighagen, E.L. RRegrs: an R package for computer-aided model selection with multiple regression models. *Journal of cheminformatics* **2015**, 7(1): 1-16.
2. Kuhn, M. Building predictive models in R using the caret package. *Journal of Statistical Software* **2008**, 28(5): 1-26.
3. Gramatica, P. Principles of QSAR models validation: internal and external. *QSAR and Combinatorial Science* **2007**, 26(5): 694.
4. UC Irvine Machine Learning Repository. Available online: http://archive.ics.uci.edu/ml/ (accessed on 6th November 2015).
5. Harrison, D.; Rubinfeld, D.L. Hedonic housing prices and the demand for clean air. *Journal of environmental economics and management* **1978,** 5(1): 81-102.
6. Cortez, P.; Cerdeira, A.; Almeida, F.; Matos, T.; Reis, J. Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems* **2009**, 47(4): 547-553.
7. Kibler, D.; Aha, D.W.; Albert, M.K. Instance-based prediction of real-valued attributes. *Computational Intelligence* **1989,** 5(2): 51-57.
8. Tsanas, A.; Little, M.; McSharry, P.E.; Ramig, L.O. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Transactions on Biomedical Engineering* **2010,** 57(4): 884-893.

SciForum
Mol2Net

A Proposal about Normalization of Experimental Designs in Computational Intelligence

Carlos Fernandez-Lozano[1],*, Julián Dorado[1], Marcos Gestal [1]

[1] RNASA-IMEDIR Group, Faculty of Computer Science, University of A Coruna, 15071 A Coruña, Spain, Emails: julian@udc.es, mgestal@udc.es

* Corresponding author: Carlos Fernandez-Lozano, Information and Communication Technologies Department, Faculty of Computer Science, University of A Coruna, 15071 A Coruña, Spain; E-Mail: carlos.fernandez@udc.es; Tel.: +34-881-01-1302; Fax: +34-981-167-160.

Published: 4 December 2015

Abstract: Experimental analysis starts with very similar premises: given a specific problem, we need to either collect or generate a dataset and to choose the best model according to the performance. A set of techniques can be evaluated (i.e. statistical or metaheuristic approaches) as well as results from previous works that should be taken into account. Thus, it is necessary to analyse the behaviour of a method with respect to the others in equality of conditions. Therefore it is necessary to formalize an experimental design to solve as effectively as possible the problem with different approaches and to estimate the error rate; so different results from different methods can be compared. In this work we propose four phases for any experimental design: data extraction, data pre-processing, model learning and the selection of the best model. These generic phases encapsulate the main operations and steps that should be performed during an experimental analysis (some of them mandatory and other optional), independently of the kind of data or method used and are not mandatory and can be adapted to a new specific domain. The proposed experimental design has proven to be a vital contribution to compare different techniques under the same conditions in different scopes.

Keywords: Experimental Design; Statistical Analysis; Computational Intelligence

1. Introduction

Experimental Design in Computational Intelligence is one of the most important aspects on every research so it is crucial to correctly define all the steps that should be address to ensure that we achieve good results. A correct experimental design should also ensure that the results are reproducible for other researchers and that are comparable among different techniques or methods over the same dataset.

This work proposes a generic framework about Normalization of the Experimental Design to address these concerns. Of course, the framework is not a fixed workflow of different phases as it can be adapted to different fields, each of them with its particularities.

Our proposal encapsulates the operations or steps that any researcher should follow to get reproducible and comparable results on their investigations with state-of-the-art approaches or other researcher's results.

2. Materials and methods

This paper normalizes and formalizes experimental design in computational intelligence and proposes and defines four phases: extraction of data, pre-processing of data, learning and selection of the best model, see Figure 1. The following paragraphs describes more in depth each of them:

Data Extraction

Firstly, we should generate the dataset defining its particular characteristics. The definition must contain the variables involved in the study and a brief description of each of them to ensure the reproducibility of the tests by external researchers.

In order to ensure that the data is enough representative of the particular studied problem, the help of experts is needed to define the cases (i.e. regions of interests for medical imaging or case-control patients).

Data Pre-Processing

After the generation of the dataset, data is in a *raw* or *pure* state. Raw data is often difficult to analyze, so it usually requires a preliminary study or pre-processing stage. This study will check that there is no data with incomplete information, outliers or noise. In case that some of the aforementioned appears in the dataset, different approaches should be applied to avoid them. Only once this process finished, it is considered that data is ready to begin the analysis itself. To check the importance of this step, it is often said that 80% of the effort of a data analysis, is spent compiling data correctly for analysis [1].

Furthermore, the variables typically present different scales or sizes, making them difficult to compare in equality of conditions. Thus normalization or standardization techniques are required to made data comparable. Of course, both techniques have their drawbacks and no one in better than the other. Furthermore, it is necessary to study the dataset for each particular problem before applying them. For example, if we try to apply a normalization step and there are outliers in the data (not removed previously), this step will scale useful data to a small interval. This is a non-desirable behavior. After a normalization step, data is scaled in the range [0,1] in case of numeric values. In case we performed a standardization process, data presents an average equal to zero and a standard deviation equal to one so they are independent of the unit of measure. Of course, depending of the kind of data, there are other well-known approaches for minimize the influence of the values.

Model Learning

Maybe the most important step within the process in computational intelligence. First of all a reference model is needed to check the results achieved for a proposed model or technique. This reference model can be extracted from a bibliography study of the field (state-of-the-art model) or constructed from a set of standard data (*gold standards* or *ground truth*) for example. In both cases the results from this reference model will be the ground truth along the following experimental design.

Once the reference model is established, it is time to build and test the model that is intended to develop in order to provide better solutions. The range of techniques available to solve any problem is usually very high. Some of these techniques are dependent on the field of study, so the researcher should review the state-of-the-art in its research field in order to choose the most suitable for his interests.

Some key points arise at this time such as the

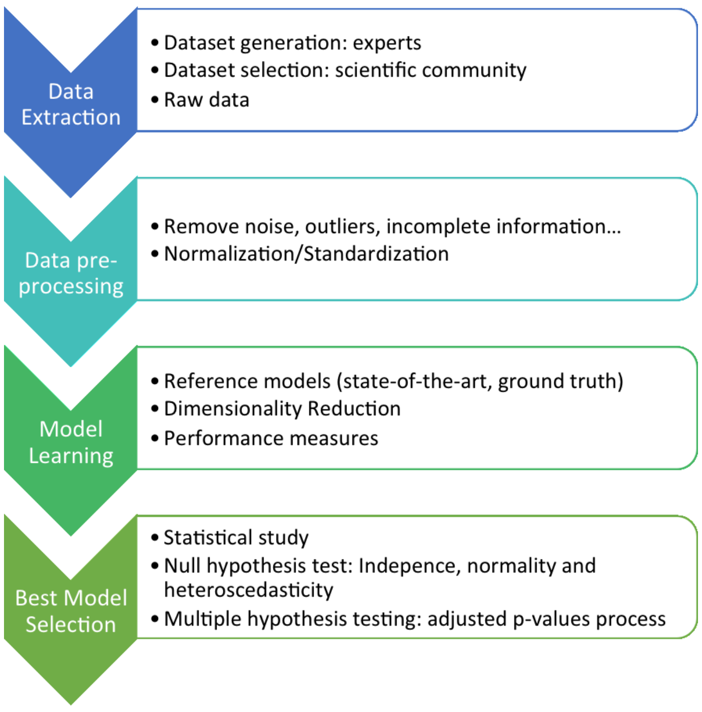

Data Extraction
• Dataset generation: experts
• Dataset selection: scientific community
• Raw data
Data pre-processing
• Remove noise, outliers, incomplete information...
• Normalization/Standardization
Model Learning
• Reference models (state-of-the-art, ground truth)
• Dimensionality Reduction
• Performance measures
Best Model Selection
• Statistical study
• Null hypothesis test: Indepence, normality and heteroscedasticity
• Multiple hypothesis testing: adjusted p-values process

Figure 1. Phases of the proposed Experimental Design

need for some measure of performance that clearly indicates how well the techniques have done this training phase. There are different well-known performance measures such as AUROC, accuracy or F-measure in classification problems or MSE, RMSE o R2 in regression problems. Sometimes it is necessary to evaluate the performance of the model using an ad-hoc measure.

Furthermore, it is desirable to avoid the overtraining of the techniques to the dataset to ensure that the model offers good results with unknown data. Techniques like *cross-validation* [2] can be useful for this point.

Finally the dimensionality of data should be taken into account. The bigger the dimensionality of the input data, the higher the number of examples necessaries for learning. Moreover, techniques for dimensionality reduction are usually interesting [3] for providing the best possible model [4] with the lower dimensionality. Thus, these techniques allow for a complexity reduction of the generated model. Furthermore, it also implies a reduction of the time and improves the overall capacity of the system.

Best Model Selection

In the previous phase, we state that there are several different measures accepted and well known as a good measure of performance for a classifier. This does not means that a researcher is able to compare different classifiers used in the same conditions and with the same dataset with just one run and this measure. At this point, it is needed to run several times each technique in order to ensure that our results are not biased because of the data. With these results per technique and in order to determine whether or not the performance of a particular technique is statistically better than the others, a null hypothesis test is needed. Furthermore, in order to use a parametric or a non-parametric test some required conditions must be checked: independence, normality and heteroscedasticity [5]. Note that these assumptions are not referring to the dataset used as input to the techniques but to the distribution of the performance of the techniques.

As part of a good experimental design for techniques comparisons, it is necessary to apply the proper test, according to the shape of the performance measure distribution. Most of the computational intelligence comparisons in the literature just apply a t-test between the performance measures to check if a technique is significantly better than the others. In some cases, this distribution does not fill the requirements of this parametric test, so a non-parametric test is required. Although the parametric test is perfectly fine to use a non-parametric test when the non-parametric test when the distribution does not fulfil the independency, normality and homoscedasticity assumptions.

Finally, after the null hypothesis test (parametric or non-parametric) is rejected, a post hoc procedure had to be used in order to address the multiple hypothesis testing and to correct the *p-values* with and adjusted *p-values* process (APV).

3. Results and Discussion

Normalization of experimental designs in computational intelligence is demonstrated using x datasets from different scientific fields. We validate this new methodology in cheminformatics and QSAR modeling with three different works. Several different Machine

Learning approaches were tested for finding the first classification model to predict cell death-related proteins [6]. In drug development it is of increased importance to find new molecular targets involved in specific diseases. Therefore, using protein star graphs for the peptide sequence information we find that the final model, reducing from 42 to 11 descriptors the original dataset [7] achieved the better results. Finally, we find for a more accurate ways of predicting residues for complex binding that can be used to model protein structure, dynamics and function [8]. We applied our experimental design as well in other fields such as bioinformatics, for example in image texture analysis problems for classification in a biomedical image texture dataset [9]. Aforementioned work used the four phases of the normalized experimental design, applying different feature selection approaches [3] for dimensionality reduction. Our results show that for all the generated datasets, our methodology reports results that are reproducible, comparable and achieved in equality of conditions. Thus, we are able to state, in each case, that we found the best model for each particular problem.

4. Conclusions

Normalization of experimental design in Computational Intelligence, as well as in other research fields is crucial. In this short communication paper we state that it is crucial to ensure that research is: reproducible, comparable and that our conclusions are based on results achieved in equality of conditions. Furthermore, for the very beginning of a research, authors should be involved in all the process that starts with the generation of the dataset, pre-processing of the data, dimensionality reduction and finally, statistical analysis. We proposed a general framework that could be used and adapted for different scenarios. Four phases could be adapted (crucial phases are mandatory but some steps are optional) for different research fields.

Acknowledgments

The authors acknowledge the support provided by the Galician Network of Drugs R+D REGID (Xunta de Galicia R2014/025) and by the "Collaborative Project on Medical Informatics (CIMED)" PI13/00280 funded by the Carlos III Health Institute from the Spanish National plan for Scientific and Technical Research and Innovation 2013-2016 and the European Regional Development Fund. This work was partially supported by the Galician Network for Colorectal Cancer Research (Red Gallega de Cáncer Colorrectal - REGICC, Ref.: CN 2014/039). This work is supported by the General Directorate of Culture, Education and University Management of Xunta de Galicia (Ref. GRC2014/049) and the European Regional Development Fund.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Tamraparni, D.; Theodore, J. *Exploratory data mining and data cleaning*. John Wiley & Sons, Inc.: 2003; p 203.
2. McLachlan, G.J.; Do, K.-A.; Ambroise, C. *Analyzing microarray gene expression data*. Wiley: 2004.
3. Saeys, Y.; Inza, I.; Larrañaga, P. A review of feature selection techniques in bioinformatics. *Bioinformatics* **2007**, *23*, 2507-2517.
4. Donoho, D.L. In *High-dimensional data analysis: The curses and blessings of dimensionality*, AMS Conference on Math Challenges of the 21st Century, 2000; pp 1-33.
5. García, S.; Fernández, A.; Luengo, J.; Herrera, F. Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Information Sciences* **2010**, *180*, 2044-2064.
6. Fernandez-Lozano, C.; Gestal, M.; González-Díaz, H.; Dorado, J.; Pazos, A.; Munteanu, C.R. Markov mean properties for cell death-related protein classification. *Journal of theoretical biology* **2014**, *349*, 12-21.
7. Fernandez-Lozano, C.; Cuiñas, R.F.; Seoane, J.A.; Fernández-Blanco, E.; Dorado, J.; Munteanu, C.R. Classification of signaling proteins based on molecular star graph descriptors using machine learning models. *Journal of theoretical biology* **2015**, *384*, 50-58.
8. Munteanu, C.R.; Pimenta, A.C.; Fernandez-Lozano, C.; Melo, A.; Cordeiro, M.N.D.S.; Moreira, I.S. Solvent accessible surface area-based hot-spot detection methods for protein–protein and protein–nucleic acid interfaces. *Journal of Chemical Information and Modeling* **2015**, *55*, 1077-1086.
9. Fernandez-Lozano, C.; Seoane, J.; Gestal, M.; Gaunt, T.; Dorado, J.; Campbell, C. Texture classification using feature selection and kernel-based techniques. *Soft Computing* **2015**, 1-12.

SciForum
Mol2Net

Prot-SSP: A Tool for Amino Acid Pairing Pattern Analysis in Secondary Structures

Miguel de Sousa [1,*], Cristian R. Munteanu [2] and Alexandre Magalhães [1]

[1] UCIBIO/REQUIMTE/University of Porto, R. Campo Alegre 687, 4169-007 Porto, Portugal; E-Mail: miguelmsousa@gmail.com (M.S.); almagalh@fc.up.pt (A.M.)
[2] RNASA-IMEDIR group, Computer Science Faculty, University of A Coruna, Campus de Elviña S/N, 15071, A Coruña, Spain (Department of Information and Communication Technologies); E-Mail: crm.publish@gmail.com
* UCIBIO/REQUIMTE/University of Porto, R. Campo Alegre 687, 4169-007 Porto, Portugal; E-Mail: miguelmsousa@gmail.com;
Tel.: +351220402504/+351220402659

Published: 4 December 2015

Abstract: It is known that individual amino acids can have a decisive role in the stabilization of a protein structure. Moreover, it is likely that specific amino acid combinations also fulfil structural and stabilizing roles in protein structure. We present Prot-SSP, an analytical Python tool designed to gather and parse sequence and structural data from sets of PDB files and determine amino acid residue pairing propensities and correlations in alpha helices and beta strands, in various secondary structure contexts.This versatile and user-friendly bioinformatic tool has proven useful for the analysis of a selected set of protein structures as shown in an illustrative example.

Keywords: secondary structure; amino acid pair, alpha helix, beta strand; software

1. Introduction

Understanding the formation, stability and function of protein structures requires the characterization of interaction preferences between amino acid residues in secondary structure motifs. It has been established that specific sequences of amino acids may have important roles in protein folding and stability and, more recently, studies show how amino acid patterns, in particular amino acid pairings patterns, may have a stabilizing or destabilizing influence in beta sheets [1], loop sequences [2] and specific (i, i+4) pairs which stabilize alpha helix structures [3].

Similarly to individual amino acids, whose frequency of occurrence in particular secondary structures varies, it is reasonable to consider that the same principle should also apply to amino acid patterns, with the different propensities of pairs to occur being related to their effect in the formation and stabilization of secondary structures. Likewise, as each residue has an individual part in the stabilization of a secondary structure, the residue distribution is different when considering the different types of positions and contexts (N-terminal, interior and C-terminal) and it is reasonable to consider that this may also apply to amino acid pairing patterns.

To evaluate the preference of a particular amino acid pairing (X_i , Y_{i+n}) to occur at the interior of the secondary structure motifs we used a statistic called global propensity ($P^{SS}{}_{XiYi+n}$), defined as a ratio of the frequency with which that pairing occurs in a given secondary structure and the frequency with which it occurs globally, irrespective of secondary structure:

$$P^{SS}_{X_iY_{i+n}} = \left. \frac{N^{SS}_{X_iY_{i+n}}}{\sum_{A,B} N^{SS}_{A_iB_{i+n}}} \middle/ \frac{N^{all}_{X_iY_{i+n}}}{\sum_{A,B} N^{all}_{A_iB_{i+n}}} \right.$$

To determine and make possible the analysis of this statistic, a novel analytical tool conceived to gather and parse sequence and structural data from user-defined sets of PDB files and determine the pairing propensities of amino acid residue pairings, as well as correlation values, in alpha helices and beta sheets, in various possible contexts of secondary structure motifs. This tool, Prot-SSP, is a GUI Python/wxPython application which, for desktop, can be compiled for the Windows XP/Vista/7 operatic systems. Our working version was compiled for Windows 7.

Prot-SSP uses user-defined inputs to cull protein chain sequence data files from the worldwide Protein Data Bank database [4], extracts protein sequence information into local storage and confirms secondary structure assignment using the DSSP algorithm [5]. The resulting output comes in the form of TXT files in CSV-format which can be opened and edited with notepad, worksheet software or, for ease of viewing, using a simple specialized macro-enabled Excel file (.XLSM) supplied with Prot-SSP.

A wide range of parameters, ranging from sample specifications to motif area under study and residue spacing as well as pattern specificity, are contemplated and can be easily and comprehensively set by the user through the program's GUI.

For alpha helices, the minimum size of protein chains to be analysed is six residues and Prot-SSP can analyse pairings up to five residues apart in three different contexts: N-terminal (considering the first three helical residues as not equivalent), interior and C-terminal (considering the last three helical residues as not equivalent).

In beta sheets, pairings are considered up to two residues apart and the minimum size for a chain to be included in analysis is six residues. Again there is a distinction between N-terminal, interior and C-terminal context and the first and last few residues are considered particular cases of study. Additionally, strands are divided into four categories depending on relative strand orientation: terminal, parallel, anti-parallel and mixed.

2. Results and Discussion

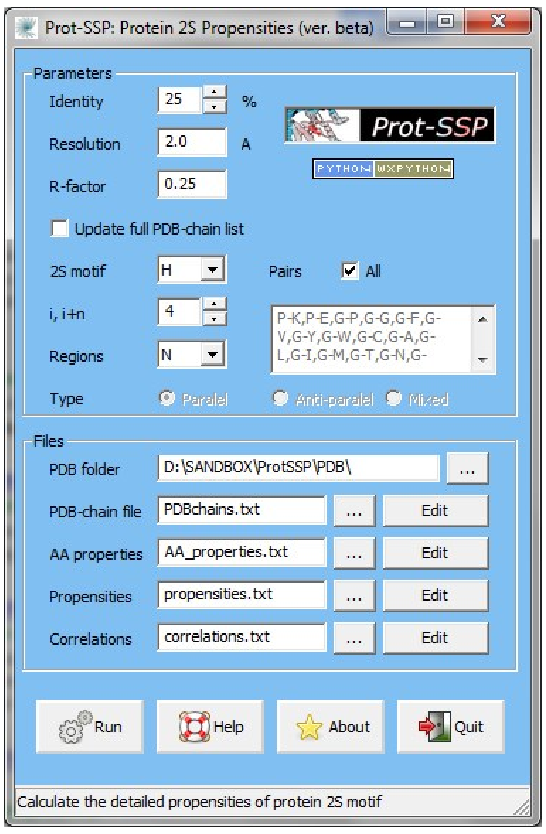
The original form of Prot-SSP, pre-development into an application with GUI, was used in an analysis of pairing propensities of amino acid residues in the interior of alpha helical structures [6] which reproduced and complemented previous research [7].

In its current form, Prot-SSP has been extensively tested for application in calculations of amino acid pairing propensities in alpha helices and in beta sheets, showing reliability as tool with data gathering and parsing capabilities for calculation of statistical analysis parameters.

Table 1 presents a typical example of the processed output of a calculation made using Prot-SSP for the particular case of (i, i+2) amino acid residue pairings in beta sheet secondary structures (results for other cases can be supplied upon request).

Table 1. Global propensities for amino acid residue pairings in beta sheet protein secondary structure. Parameters: 25% identity; 2.0 A resolution; 0.25 R-value; beta 2S motif; i, i+2 residue spacing; Interior region; all pairs considered; no distinction between types of beta strand (All)

	P	G	F	V	Y	W	C	A	L	I	M	T	N	S	H	D	Q	R	K	E	X
P	0.1	0.3	0.3	0.5	0.6	0.3	0.2	0.3	0.3	0.5	0.7	0.5	0.2	0.2	0.3	0.2	0.3	0.4	0.3	0.2	0.0
G	0.2	0.8	1.3	1.7	1.1	0.8	1.9	1.0	1.2	1.4	1.2	0.8	0.5	0.6	0.7	0.4	0.6	0.6	0.5	0.6	0.0
F	0.6	1.3	2.3	2.8	1.7	1.9	2.7	1.2	1.9	2.6	2.3	1.4	1.0	1.1	1.1	0.5	0.9	0.9	0.7	0.7	0.0
V	0.8	1.5	2.9	3.3	2.4	1.9	2.3	1.7	2.1	2.7	2.5	1.9	1.0	1.5	1.5	0.9	1.3	1.3	1.0	1.3	0.0
Y	0.4	1.1	2.3	2.7	2.0	1.4	4.4	1.3	1.9	2.6	2.0	1.6	0.7	1.1	1.1	0.7	1.1	1.2	1.0	1.1	0.0
W	0.2	1.0	2.0	1.8	1.9	1.2	2.2	1.1	1.5	1.7	2.0	1.2	0.9	1.0	1.0	0.6	1.0	1.1	0.4	0.8	0.0
C	0.7	1.3	2.1	4.0	2.2	1.6	1.8	1.5	1.9	2.0	1.7	1.3	0.7	1.2	1.2	0.6	1.0	1.1	1.0	1.2	0.0
A	0.4	0.8	1.4	1.7	1.0	0.9	1.3	0.7	1.0	1.4	1.4	0.8	0.5	0.8	0.8	0.4	0.5	0.5	0.6	0.5	0.0
L	0.4	1.1	2.0	2.6	1.8	1.4	2.9	0.9	1.6	2.2	1.6	1.3	0.6	1.0	0.8	0.4	0.7	0.8	0.7	0.6	0.0
I	0.6	1.3	2.5	2.7	2.2	1.5	2.8	1.5	1.9	2.8	3.2	1.5	0.8	1.4	1.3	0.8	1.1	1.0	0.8	1.1	0.0
M	0.6	1.2	2.4	2.6	2.1	1.8	2.8	1.0	2.0	2.5	1.4	1.3	0.7	1.1	1.1	0.5	0.9	0.7	0.5	0.8	0.0
T	0.3	0.8	1.4	1.8	1.6	1.4	1.4	1.0	1.2	1.5	1.5	2.0	0.8	1.3	1.5	0.7	1.2	1.3	1.1	1.0	0.0
N	0.2	0.5	0.8	0.8	0.9	1.1	1.1	0.6	0.5	0.8	0.5	1.0	0.5	0.8	0.8	0.4	0.7	0.9	0.5	0.5	0.0
S	0.3	0.7	1.3	1.5	1.3	1.0	1.4	0.7	0.8	1.3	1.2	1.4	0.7	1.1	1.0	0.6	0.9	1.0	0.8	0.9	0.0
H	0.4	0.6	1.4	1.6	1.2	1.5	0.5	0.9	1.0	1.4	1.6	1.3	0.7	0.8	0.7	0.6	1.0	0.9	0.9	1.0	0.0
D	0.2	0.5	0.6	0.9	0.9	0.6	0.8	0.4	0.5	0.8	0.5	0.8	0.5	0.5	0.5	0.3	0.7	0.7	0.6	0.4	0.0
Q	0.4	0.4	0.9	1.2	0.9	1.2	0.5	0.6	0.6	1.1	1.1	1.5	0.7	0.8	0.8	0.5	0.6	0.9	0.7	0.7	0.0
R	0.3	0.6	1.2	1.4	1.2	0.8	1.3	0.8	0.8	1.1	1.2	1.3	0.5	1.0	1.2	0.7	1.0	1.2	0.8	1.0	0.0
K	0.2	0.6	0.8	1.0	0.9	1.0	0.7	0.5	0.5	0.9	0.9	1.3	0.5	0.9	1.0	0.6	0.8	1.0	0.7	0.8	0.0
E	0.2	0.5	0.7	1.0	0.7	1.0	0.7	0.5	0.6	0.7	0.8	1.3	0.6	1.0	1.0	0.6	0.7	0.9	0.8	0.9	0.0
X	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0



Prot-SSP: Protein 2S Propensities (ver. beta)

Parameters

Identity 25 %

Resolution 2.0 A

R-factor 0.25

□ Update full PDB-chain list

2S motif H

i, i+n 4

Regions N

Pairs ☑ All

P-K,P-E,G-P,G-G,G-F,G-V,G-Y,G-W,G-C,G-A,G-L,G-I,G-M,G-T,G-N,G-

Type ● Paralel ○ Anti-paralel ○ Mixed

Files

PDB folder D:\SANDBOX\ProtSSP\PDB\

PDB-chain file PDBchains.txt

AA properties AA_properties.txt

Propensities propensities.txt

Correlations correlations.txt

Calculate the detailed propensities of protein 2S motif

Figure 1. Prot-SSP graphic user interface

3. Materials and Methods

Through the GUI main window (**Figure 1**), the user can define and adjust parameters for propensity calculations as well as the input/output files. Details of undergoing operations, progress detail, errors and some calculation details are displayed in the console window.

The thresholds defined when culling PDB using the PISCES [8] server are detailed in "Identity", "Resolution" and "R-factor" and the user has the option of checking PDB for more recent versions of those defined in the PDB-chain file input. The user may also specify the secondary structure motif (alpha helix or beta sheet) under study as well as the amino acid pairing separation and the region under study (N-, C-terminal or interior). Specifically for beta sheet structures, the user can define the relative orientation of the strands to be considered for analysis. Specific pairing possibilities can also be defined by the user.

A TXT file containing the chain listing, which can be created manually or culled from the PISCES server, is used by Prot-SSP to download all the relevant structural files from the PDB into local storage and update any present files if judged necessary. During operation, the program cross-checks the locally stored PDB files with DSSP for the attribution of secondary structure to the structural data and culling of relevant sequences and creates an additional text file with lists and details the sequences culled for analysis.

The names and locations of output TXT files may also be defined through the GUI. The visualization and editing of the resulting files, while possible using text editing software can be easily and comfortably done using a simple macro-enabled Excel XLSM file, supplied with Prot-SSP, to colour-code the entries from the resulting propensity-value tables.

4. Conclusions

Prot-SSP was projected and created to address the need for a tool that could simultaneously address the need for data gathering, parsing, calculation and a degree of analysis of amino acid residue patterns in protein secondary structures. A relatively simple tool to use, its application to carefully selected data sets and the results yielded may prove it to have significant potential – both for immediate analysis and for future applications in the field of protein structure prediction.

Acknowledgments

The authors acknowledge the support provided by the Galician Network of Drugs R+D REGID (Xunta de Galicia R2014/025) and by the "Collaborative Project on Medical Informatics (CIMED)" PI13/00280 funded by the Carlos III Health Institute from the Spanish National plan for Scientific and Technical Research and Innovation 2013-2016 and the European Regional Development Fund / GAIN (FEDER - CONECTAPEME - INTERCONECTA). This work was partially supported by the Galician Network for Colorectal Cancer Research (Red Gallega de Cáncer Colorrectal - REGICC, Ref.: CN2014/039), Institute for Biomedical Informatics of A Coruña (INIBIC), and Center for Research of Information and Communication Technologies (CITIC).

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Fooks, H.M.; Martin, A.C.R.; Woolfson, D.N.; Sessions, R.B.; Hutchinson, E.G. Amino acid pairing preferences in parallel beta-sheets in proteins. *Journal of Molecular Biology* **2006**, *356*, 32-44.

2. Crasto, C.J.; Feng, J.A. Sequence codes for extended conformation: A neighbor-dependent sequence analysis of loops in proteins. *Proteins-Structure Function and Bioinformatics* **2001**, *42*, 399-413.

3. Andrew, C.D.; Penel, S.; Jones, G.R.; Doig, A.J. Stabilizing nonpolar/polar side-chain interactions in the alpha-helix. *Proteins-Structure Function and Genetics* **2001**, *45*, 449-455.

4. Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.B.; Meyer, E.F.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. Protein data bank - computer-based archival file for macromolecular structures. *Journal of Molecular Biology* **1977**, *112*, 535-542.

5. Kabsch, W.; Sander, C. Dictionary of protein secondary structure - pattern-recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577-2637.

6. de Sousa, M.M.; Munteanu, C.R.; Pazos, A.; Fonseca, N.A.; Camacho, R.; Magalhaes, A.L. Amino acid pair- and triplet-wise groupings in the interior of alpha-helical segments in proteins. *J. Theor. Biol.* **2011**, *271*, 136-144.

7. Fonseca, N.A.; Camacho, R.; Magalhaes, A.L. Amino acid pairing at the n- and c-termini of helical segments in proteins. *Proteins-Structure Function and Bioinformatics* **2008**, *70*, 188-196.

8. Wang, G.L.; Dunbrack, R.L. Pisces: Recent improvements to a pdb sequence culling server. *Nucleic Acids Research* **2005**, *33*, W94-W98.

SciForum
Mol2Net

TI2BioP: Topological Indices to BioPolymers

Guillermin Agüero-Chapin[1,2*], Reinaldo Molina-Ruiz[2] and Agostinho Antunes[1,3]

[1] CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Rua dos Bragas, 177, 4050-123 Porto, Portugal; E-Mail: gchapin@ciimar.up.pt; aantunes@ciimar.up.pt

[2] Centro de Bioactivos Químicos, Universidad Central ¨Marta Abreu¨ de Las Villas (UCLV), Santa Clara, 54830, Cuba; E-Mail: reymolina@uclv.edu.cu

[3] Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal; E-Mail: aantunes@ciimar.up.pt

* Author to whom correspondence should be addressed; E-Mail: gchapin@ciimar.up.pt

Published: 4 December 2015

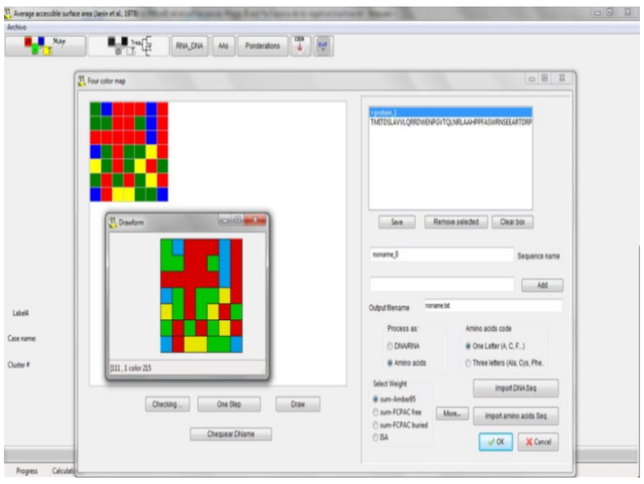
Abstract: TI2BioP (Topological Indices to BioPolymers) is a software to estimate topological indices (TIs) from two-dimensional (2D) graphical approaches for the natural biopolymers DNA, RNA and proteins. The methodology mainly turns long biopolymeric sequences into 2D artificial graphs such as Cartesian and four-color maps but also reads other 2D graphs from the thermodynamic folding of DNA/RNA strings inferred from other programs. The topology of such 2D graphs is either encoded by node or adjacency matrixes for the calculation of the spectral moments as TIs. These numerical indices were used to build up alignment-free models to the functional classification of biosequences and to calculate alignment-free distances for phylogenetic purposes. We released the version 2.0 of the software that can be freely downloaded from http://ti2biop.sourceforge.net/.

Keywords: 2D graphs; Topological indices; Alignment-free models; phylogenetics

1. TI2BioP software

TI2BioP was mainly developed from the **TOPS-MODE** methodology [1] for the estimation of the spectral moments series as TIs, but it takes advantage of the **MARCH-INSIDE** program platform [2]. It was built up on object-oriented Free Pascal IDE Tools (Lazarus) running on either a Windows or Linux operating system. TI2BioP has a friendly interface allowing users to introduce multiple fasta files containing either DNA or protein sequences to select the biopolymer 2D representation type and the calculation of TIs. We released version 2.0 of the software that can be freely downloaded from http://ti2biop.sourceforge.net/. This version contains two main types of 2D artificial representations, one based on Cartesian representation for DNA strings introduced by Nandy [3] and the other inspired by the four-color maps reported by Randic [4] (**Figure 1**).

These two 2D artificial graphs implemented in **TI2BioP** can be applied to nucleotide and amino acid strings as well as to the spectral moments calculations for each type of 2D DNA and protein maps [5]. It is noteworthy that the 2D Cartesian representation was extended to proteins by our group [6] and protein four-color maps were modified according to the amino acid clustering proposed in ref. [6]. Such four-color map modifications allow the speeding up of graph-building and facilitates the calculation of spectral moments as TIs [7].



>protein_1
TMTDSLAVVLQRRDWENPGVTQLNRLAAHPPFASWRNSEEARTDRP

Sequence name: noname_0

Output filename: noname.txt

Process as: ○ DNA/RNA ● Amino acids

Amino acids code: ● One Letter (A, C, F...) ○ Three letters (Ala, Cys, Phe...

Select Weight: ● sum-Amber95 ○ sum-FCPAC free ○ sum-FCPAC buried ○ ISA

Figure 1. TI2BioP window view of the (Topological Indices to BioPolymers) software for the representation of protein four-color maps

TI2BioP can also import files containing 2D structures inferred by other DNA/RNA folding algorithms, e.g. Mfold implemented in the RNA structure software [8], for the calculation of the spectral moments as TIs. **TI2BioP** automatically represents natural biopolymers as 2D graphs and straightforward calculates spectral moments series (TIs) to be used either for statistical classification techniques in building alignment-free models for functional classification or for deriving several alignment-free distance matrices, e.g. Euclidean, Jensen–Shannon, Hamming and Minkowsk for phylogenetic purposes (**Figure 2**)

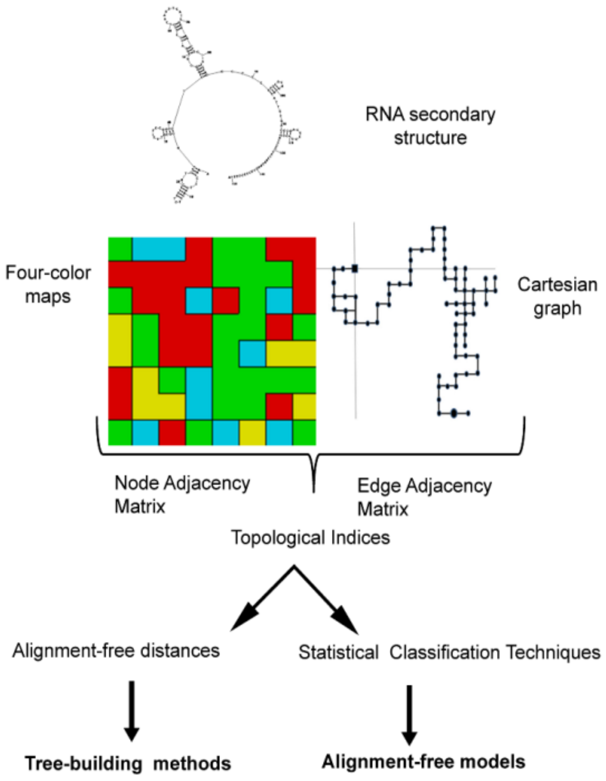


Figure 2. Workflow for the calculation of the topological indices by TI2BioP (Topological Indices to BioPolymers) from several 2D graphs for DNA, RNA and proteins

References

1. Estrada E. On the topological sub-structural molecular design (TOSS-MODE) in QSPR/QSAR and drug design research. SAR QSAR Environ Res. 2000; **11**: 55-73.

2. González-Díaz H, Molina-Ruiz R, Hernandez I. MARCH-INSIDE v3.0 (**MAR**kov **CH**ains **IN**variants for **SI**mulation & **DE**sign). 3.0 ed2007. p. Windows supported version under request to the main author contact email: gonzalezdiazh@yahoo.es.
3. Nandy A. Two-dimensional graphical representation of DNA sequences and intron-exon discrimination in intron-rich sequences. Comput Appl Biosci. 1996; **12**: 55-62.
4. Randic M, Lers N, Plavšić D, Basak S, Balaban A. Four-color map representation of DNA or RNA sequences and their numerical characterization. Chemical Physics Letters 2005; **407**: 205-8.
5. Molina R, Agüero-Chapin G, Pérez-González MP. TI2BioP (Topological Indices to BioPolymers) *version 2.0.*: Molecular Simulation and Drug Design (MSDD), Chemical Bioactives Center, Central University of Las Villas, Cuba; 2011
6. Aguero-Chapin G, Gonzalez-Diaz H, Molina R, Varona-Santos J, Uriarte E, Gonzalez-Diaz Y. Novel 2D maps and coupling numbers for protein sequences. The first QSAR study of polygalacturonases; isolation and prediction of a novel sequence from Psidium guajava L. FEBS Lett. 2006; **580**: 723-30
7. Aguero-Chapin G, Molina-Ruiz R, Maldonado E, de la Riva G, Sanchez-Rodriguez A, Vasconcelos V, et al. Exploring the adenylation domain repertoire of nonribosomal peptide synthetases using an ensemble of sequence-search methods. PLoS One. 2013; **8**: e65926.
8. Mathews DH. RNA secondary structure analysis using RNAstructure. Curr Protoc Bioinformatics. 2006; **Chapter 12**: Unit 12 6.

SciForum
Mol2Net

Conception, Synthesis, Characterization and Antimicrobial Evaluation of New Ferrocene-Based Derivatives Inspired by the Bisacodyl Lead Structure

Meral Görmen[1], Maité Sylla-Iyarreta Veitía[1,*], Fatma Trigui[2], Mehdi El Arbi[2] and Clotilde Ferroud[1]

[1] Equipe de Chimie Moléculaire du Laboratoire CMGPCE, EA 7341, Conservatoire national des arts et métiers, 2 rue Conté,75003, Paris; meralgormen@gmail.com (M.G), clotilde.ferroud@cnam.fr (C.F.)

[2] Centre de Biotechnologie de Sfax, Université de Sfax, Route de Sidi Mansour Km 6, BP 1177, 3018 Sfax, Tunisia; mehdi_arbi@yahoo.fr (M.E.); yangui.trigui.fatma@gmail.com (F.T.)

* Author to whom correspondence should be addressed; E-Mail: maite.sylla@cnam.fr;
Tel.: +33-1-58 80 84 82; Fax: +33-1-40 27 25 84.

Published: 4 December 2015

Abstract: The antibacterial activity of bisacodyl, a drug used in therapeutic as laxative, and its ferrocenyl analogues was investigated against Gram-positive and Gram-negative foodborne pathogens including *Listeria monocytogenes*, *Escherichia coli*, *Enterococcus faecalis*, *Salmonella enterica*, *Micrococcus luteus* and *Staphylococcus aureus*. The results showed that most of these compounds exhibit an excellent antimicrobial activity, and the bisacodyl analogues seemed to be more bactericides than bacteriostatic.

Keywords: bisacodyl; ferrocene; antibacterial activity

1. Introduction

The leaving behind of the antibiotic discovery area by many pharmaceutical companies is one of the major reasons of the discovery decline. Since the year 2000, only eight antibacterial molecules have obtained a marketing authorization. Moreover, despite the discovery over the last twenty years of compounds with an interesting antibiotic activity, few of them belong to new chemical classes or have the required properties to become drugs or to circumvent resistance problems. One of the approaches to overcome drug resistance is the search for new multi-target inhibitors.[1-3]

At present, in order to accelerate the development of drugs with relatively low costs and reduced risks, pharmaceutical companies develop new approaches from existing drugs. This methodology known as drug repurposing allows the development of new indications for existing drugs with well-known pharmacokinetic profiles, known safety profile, already solved manufacturing issues. Concerned by the high interest in infectious disease research and considering the forgoing argues, we decided to evaluate the antimicrobial activity of bisacodyl, drug used in therapeutics as laxative. [4-5]

To our delight, the bisacodyl showed an excellent antimicrobial activity (MIC values of 6.25-12.5 µg/mL; 3.125-12.5 µg/mL and 6.25-12.5 µg/mL against Gram-positive strains *Micrococcus, Staphylococcus* and *Listeria* respectively). These results encouraged us to develop a series of new analogues. We developed the strategy of incorporating an organometallic ferrocenyl moiety. The use of a ferrocene group to enhance the activity of antibiotics was proposed by Edwards et al. in 1976 [6]. The advantages of the introduction of ferrocenyl moiety to increase the antimicrobial activity have been widely described in the literature [7-12]. The use of ferrocene is especially attractive because it is neutral, chemically stable, a nontoxic molecule and can be easily derivatized or functionalized. Several ferrocenyl compounds have been described for their antitumor, antimalarial or antifungal properties [13-15].

All the ferrocenyl compounds that we have synthesized were characterized and evaluated on pathogen bacteria Gram positive and Gram negative. Finally the antimicrobial effect of bisacodyl and one of its analogues was also estimated [16].

2. Results and Discussion

2.1 Synthesis

The ferrocenyl arylethylpyridines and some corresponding *N*-oxide derivatives were prepared via a McMurry coupling reaction. General synthetic methods to obtain the target compounds are outlined in Scheme 1. The detailed synthesis has been described by us [16]. The key step to obtain the desired olefin intermediates involved a McMurry cross-coupling reaction between the ketone **3** and ferrocenecarboxaldehyde to afford the 2-(1-(4-methoxyphenyl)-2-ferrocenylvinyl)pyridine **4**. The olefins are obtained in two separable *E* and *Z* isomers with 30% and 49% yields respectively. These modest yields results of the possible competition between the formation of the desired cross-coupled product and the two homo-coupled compounds [17].

The demethoxylated compounds **5** and **8** were synthesized by reaction with boron tribromide in dichloromethane. The 2-(1-(4-hydroxyphenyl)-2-ferrocenylethyl)pyridine **8** was obtained in 50% yield (non-optimized yield). The 2-(1-(4-hydroxyphenyl)-2-ferrocenylvinyl) pyridine **5**, was obtained with low yields. The isomerization of the double bond can explain this result. It is known that organometallic complexes adjacent to a double bond advantage the stabilization of α carbenium ions by protonation of the double bond in acidic medium. A similar isomerization of analogous organometallic complexes has been described in the literature [18]. Acetates **6** and **9** were obtained with yields of 96% and 76% respectively. *N*-oxide ferrocenyl derivatives **10** and **11** were prepared from the corresponding ferrocenyl pyridines by oxidation with *m*-chloroperbenzoic acid in dichloromethane at room temperature. Compounds **10** and **11** were isolated after purification by flash chromatography on silica gel with non-optimized yields of 7% and 16% respectively. All synthesized compounds were biologically evaluated.

2.2 Biological studies

Bisacodyl and its analogues were screened for antibacterial activity against Gram-positive and Gram-negative pathogens using doxycycline, a broad spectrum antibiotic, as a control. The minimum inhibitory concentration and minimum bactericidal concentration were measured for all compounds (Table 1). The MIC and MBC values for doxycycline were found to be <12.5 µg/mL and 12.5 µg/mL on *Staphylococcus aureus*.

Bisacodyl and its ferrocenyl analogues showed an excellent antimicrobial activity. All tested compounds seemed to be more bactericidal than bacteriostatic, because MBC/MIC ratio is less than or equal to four (≤ 4) [19]. Even if no significant difference in activity against Gram-positive or Gram-negative bacteria was observed, Gram-positive strains, *Micrococcus* and *Staphylococcus aureus* seemed to be more sensitive than Gram-negative strains. Consequently, there is a potential use against Gram-positive bacterial infections for these compounds.

In the ferrocenyl arylvinylpyridine series (compounds **4**, **5**, **6** and **10**) compounds **5** and **6** exhibited the greatest antimicrobial activity with MIC values between [12.5-25] µg/mL against *Micrococcus*, *Staphylococcus* and *Listeria*. No difference for antimicrobial activity was observed between isomers **5**, against both gram-positive and gram-negative bacteria. However for compound **4**, the *Z* isomer showed a higher activity compared with its *E* analogue. Compound **4b** exhibited MIC values between [12.5-25] µg/mL against *Micrococcus* and *Staphylococcus* respectively versus [25-50] µg/mL for compound **4a**. Furthermore, when *N*-oxide moiety was introduced (see compound **10**) the antimicrobial activity was comparable with that of compounds **5** and **6**. These results suggested that the functionalization with a hydroxyl group, an acetoxy group or *N*-oxide moiety can have a positive influence for the antimicrobial activity. It has been described in the literature that the introduction of a pyridine *N*- oxide may play an important role in biological activity because it may increase the metabolic stability and bioavailability. [20-22]

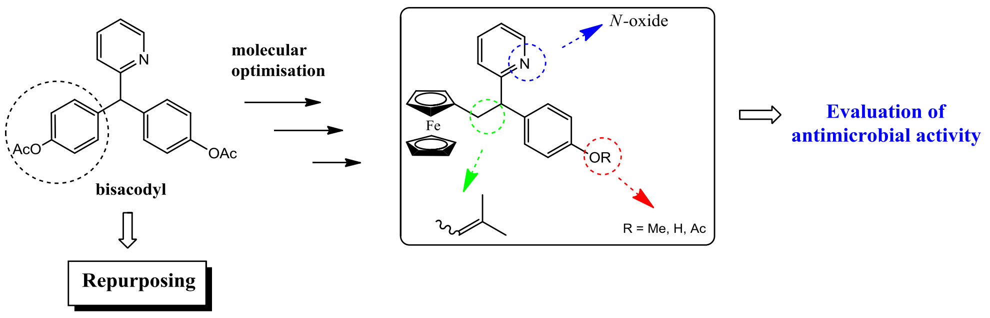
In the ferrocenyl arylethylpyridine series (compounds **7-11**), compound **11** exhibited the greatest antimicrobial activity with MIC values between [12.5-25] µg/mL against Gram-positive strains *Micrococcus*, *Staphylococcus* and *Listeria*. Once again these results suggested that *N*-oxide moiety plays an important role in the antimicrobial activity. Compounds **7**, **8** and **9** were less active against *Listeria*, with MIC values between [25-50] µg/mL. Compound **8** showed less anti-*Micrococcus* activity compared to others ferrocenyl analogues (MIC values between [25-50] µg/mL).

Bisacodyl exhibited the best antimicrobial activity with MIC values between [6.25-12.5] µg/mL; [3.125-12.5] µg/mL, [6.25-12.5] µg/mL for against Gram-positive strains *Micrococcus*, *Staphylococcus* and *Listeria*. Gram-negative strains *Escherichia coli*, *Enterococcus* and *Salmonella* were also sensitive to bisacodyl with MIC values between [3.125-12.5] µg/mL; [3.125-12.5] µg/mL, [6.25-12.5] µg/mL respectively. These results could suggest that an acetoxy group may be necessary to achieve excellent antimicrobial activity.

Finally the antimicrobial effect of bisacodyl and its ferrocenyl acetyl analogue 2-(1-(4-acetoxyphenyl)-2-ferrocenylethyl) pyridine **9** was estimated on *Listeria monocytogenes* and *Salmonella enterica* strains using levofloxacine and fusidic acid as controls. The details of this estimation were described by us [16]. Antimicrobial effect was estimated (Table 2)

It may be noted that the ferrocenyl derivative **9**, is more active against *Listeria monocytogenes* (gram-positive bacteria) with EC_{50} and EC_{90} of 37 and 71 µM respectively than bisacodyl with EC_{50} and EC_{90} of 46 and 89 µM respectively. However, bisacodyl has a better activity against *Salmonella enterica* (gram-negative bacteria) with EC_{50} and EC_{90} of 50 and 69 µM respectively than compound **9** with EC_{50} and EC_{90} values of 51.5 and 85 µM respectively. According to the obtained results bisacodyl and compound **9** are more active against *Listeria monocytogenes* than *Salmonella enterica*. Therefore, these compounds are promising antimicrobials, more effective against gram-positive than against gram-negative bacteria.

Figure 1. Repurposing and pharmacomodulation of bisacodyl



Scheme 1. Synthesis of ferrocenyl derivatives (a): *n*-BuLi, *p*-anisaldehyde, THF, -78°C / r.t., 17 h; (b): NaOH, O_2, toluene, reflux, 18 h; (c): $TiCl_4$, Zn, THF, reflux, 2h then ferrocenecarboxaldehyde, 8 min; (d): H_2, Pd/C, AcOEt, r.t., 36 h; (e): BBr_3, dichloromethane, r.t., 22 h; (f): Ac_2O, NaOH, 20°C, 2 h for compound **6**, 48 h for compound **9** ; (g): *m*-CPBA, dichloromethane, r.t., 5-12 h.

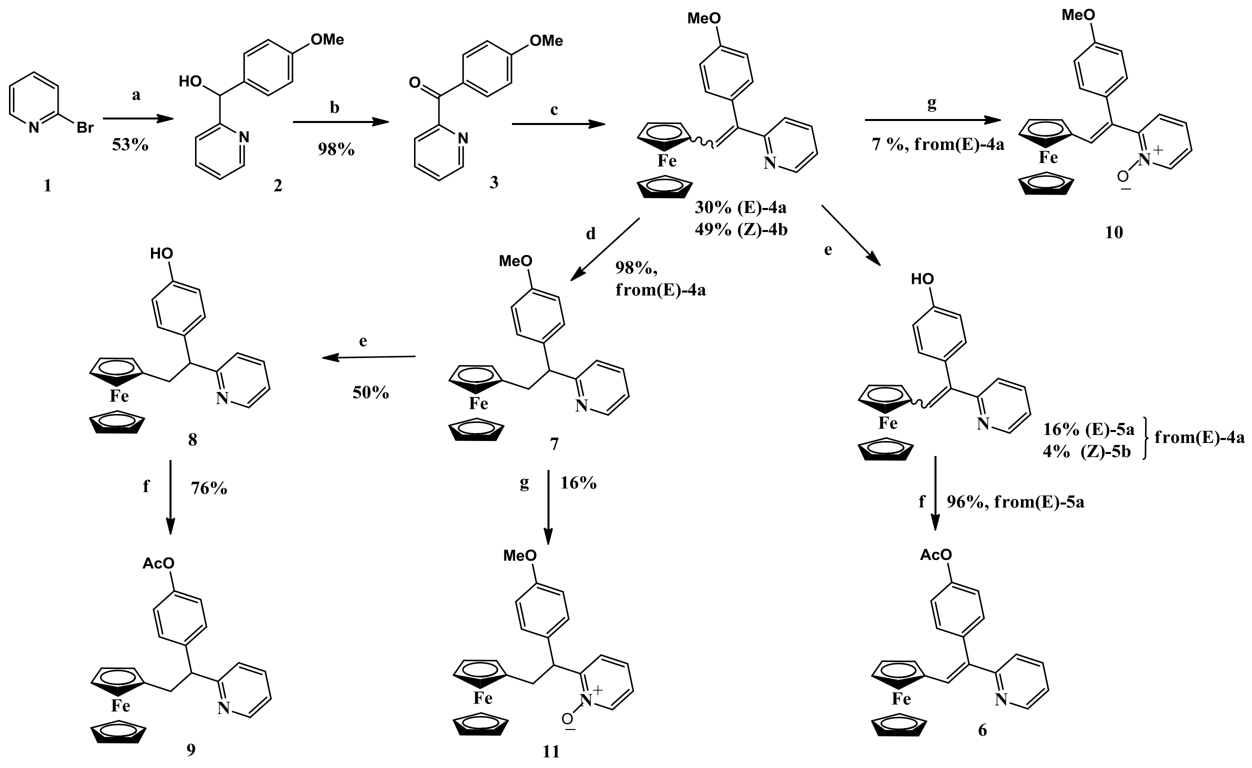


Table 1. Antimicrobial activities of bisacodyl and their ferrocenyl analogues. Minimum inhibitory concentration (MIC) and minimum bactericidal concentration (MBC) in µg/mL.

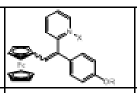
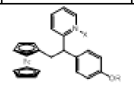
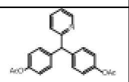
Cpd. Num			**Gram (+)**			**Gram (-)**		
			Micrococcus	*Staphylococcus*	*Listeria*	*E. Coli*	*Entero bacterium*	*Salmonella*
4a	*E isomer* R=Me	MIC	[25-50]	[25-50]	[25-50]	[25-50]	[25-50]	[12.5-25]
		MBC	>100	100	>100	>100	>100	>100
4b	*Z isomer* R=Me	MIC	[12.5-25]	[12.5-25]	[25-50]	[12.5-25]	[25-50]	[25-50]
		MBC	100	>100	50	>100	>100	50
5a	*E isomer* R=H	MIC	[12.5-25]	[12.5-25]	[12.5-25]	[12.5-25]	[12.5-25]	[12.5-25]
		MBC	100	100	100	100	100	100
5b	*Z isomer* R=H	MIC	[12.5-25]	[12.5-25]	[12.5-25]	[12.5-25]	[12.5-25]	[12.5-25]
		MBC	100	100	100	100	100	100
6	*E isomer* R=Ac	MIC	[12.5-25]	[12.5-25]	[12.5-25]	[25-50]	[12.5-25]	[25-50]
		MBC	>100	>100	>100	>100	>100	>100
10	*E isomer* R=Me, X=O⁻	MIC	[12.5-25]	[12.5-25]	[12.5-25]	[25-50]	[12.5-25]	[25-50]
		MBC	>100	>100	50	>100	>100	>100
								
7	R=Me	MIC	[12.5-25]	[12.5-25]	[25-50]	[25-50]	[25-50]	[25-50]
		MBC	100	>100	50	>100	>100	50
8	R=H	MIC	[25-50]	[12.5-25]	[25-50]	[25-50]	[12.5-25]	[25-50]
		MBC	>100	25	>100	>100	>100	100
9	R=Ac	MIC	[12.5-25]	[12.5-25]	[25-50]	[25-50]	[25-50]	[25-50]
		MBC	50	>100	50	>100	>100	50
11	R=Me, X=O⁻	MIC	[12.5-25]	[12.5-25]	[12.5-25]	[12.5-25]	[25-50]	[12.5-25]
		MBC	>100	>100	100	>100	>100	>100
bisacodyl		MIC	[6.25-12.5]	[3.125-6.25]	[6.25-12.5]	[3.125-6.25]	[3.125-6.25]	[6.25-12.5]
		MBC	50	12.5	50	>50	50	25
doxycycline		MIC	-	<12.5	-	-	-	-
		MBC	-	12.5	-	-	-	-

Table 2. EC50 and EC90 values of bisacodyl and its ferrocenyl analogue **9** in µM.

Compounds	**strains**	**EC_{50}**	**EC_{90}**
bisacodyl	*Listeria*	46	89
	Salmonella	50	69
compound **9**	*Listeria*	37	71
	Salmonella	51.5	85
fusidic acid	*Listeria*	38	75
	Salmonella	47	68
levoflaxacine	*Listeria*	3	10
	Salmonella	1	3

3. Materials and Methods

All reagents were obtained from commercial sources unless otherwise noted, and used as received. Heated experiments were conducted using thermostatically controlled oil baths and were performed under an atmosphere oxygen-free in oven-dried glassware. All reactions were monitored by analytical thin layer chromatography (TLC) or by Gas chromatography-Mass spectrometry (GC-MS). TLC was performed on aluminium sheets precoated silica gel plates (60 F_{254}, Merck). TLC plates were visualized using irradiation with light at 254 nm or in an iodine chamber as appropriate. Flash column chromatography was carried out when necessary using silica gel 60 (particle size 0.040-0.063 mm, Merck).

All synthesized compounds were characterized by NMR, IR, MS data and by the TLC behavior.

The experimental procedures and the characterization have been previously described [16].

In vitro antibacterial activity

Microorganism growth inhibition assays were performed using LB (1% Bactotryptone, 0.5% Yeast extract, 0.5% NaCl) cultures *of Listeria monocytogenes* (ATCC 7644), *Escherichia coli* (ATCC 10536), *Enterococcus faecalis* (ATCC 19434), *Salmonella enterica* (ATCC 13314), *Micrococcus luteus* (ATCC 9341) and *Staphylococcus aureus* (ATCC 6538).

All synthesized compounds were tested in triplicate, using microplate dilution method. Minimal inhibitory concentrations (MICs) of compounds were determined according the National Committee for Clinical Laboratory Standard (NCCLS, 2002). The compounds were dissolved in dimethylsulfoxide (DMSO). Serial two fold dilutions of each sample to be evaluated were made to yield volumes of 100 µL per well with final concentrations ranging from 100 to 12.5 µg/mL. 100 µL of bacteria suspension with a concentration of 10^7 CFU/mL were added to each well. Negative control wells contained bacteria only in LB broth medium. After incubation at 30°C for 16 h, the minimal inhibitory concentrations (MICs) were recorded as the lowest concentration of compound in the medium that showed no microbial growth. 3-(4,5-dimethyl thiazol-2-yl)-2,5- diphenyltetrazo lium bromide (MTT) was added to the wells to facilitate reading of the plates. If there is microbial growth, MTT turns to blue if not the medium remains yellow. Solvent medium and positive growth controls were also run simultaneously. Then from each tube, one loopful was cultured on plate count agar and incubated for 24 h at 30°C. The lowest concentration of the compound supporting no colony formation was defined as the MBC.

The estimation of the antimicrobial effect against microbial strains was performed by the method of micro dilution in ELISA plates. A stock solutions of the tested products were prepared in DMSO or water, depending on their solubility. In Elisa plates and for each product a series of nine wells containing 100 µL of culture medium with decreasing concentration of the product was prepared by the successive ½ dilution. A 100 µL of overnight shaking microbial culture, incubated at adequate temperature, depending on bacterial strains, was used to inoculate the plate wells containing different concentrations of compounds. The final concentration of each product, for a series of eight wells was 300 µM, 150 µM, 75 µM, 37.5 µM, 18.75 µM, 9.37 µM, 4.68 µM, 2.34 µM and 1.17 µM

The plates were incubated with shaking overnight at the same temperature and their OD was measured at 600 nm. A negative control (uninoculated wells) and a positive control (seeded and without antimicrobial compound wells) were prepared under the same experimental conditions.

The inhibitory activity of the tested compounds was calculated according to the formula:

IA (%) = 100−100 (OD 600 (x) / OD 600 (i))

where (x) is the microbial culture containing the inhibitor and (i) is the microbial culture without inhibitor.

.

.

4. Conclusions

We described the antimicrobial activity of bisacodyl and its ferrocenyl analogues. The compounds were tested on Gram-positive pathogens: *Micrococcus*, *Staphylococcus*, *Listeria* and Gram-negative pathogens, *Escherichia coli*, *Enterococcus* and *Salmonella*. The results obtained revealed that these compounds are potentially more effective against gram-positive than against gram-negative bacteria. The importance of anti-infective research field motivates us to continue the study of based bisacodyl skeleton compounds in order to find new potential antimicrobial molecules. The influence of ferrocene and other chemical substitutions on the antimicrobial activity will be the subject of further investigations.

Acknowledgments

Authors thank the MESR (Ministère de l'Enseignement et de la Recherche français and the Ministère de l'Enseignement supérieur et de la Recherche Scientifique of Tunisia for financial support.

Author Contributions

The French team, MG, MS-IV and CF, is responsible for the synthesis and characterization of compounds and the Tunisian team, FT and ME is responsible for microbiological testing. All authors contributed to the drafting and revision of the article and approved the final version.

.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. *Nature Reviews Drug Discovery*, **2014**, 13, 165-165.
2. Oldfield, E. and Feng, X. Resistance-resistant antibiotics, *Trends in Pharmacological Sciences*, **2014**, 35, 664-674.
3. Brooks, B. D. and Brooks, A. E. Therapeutic strategies to combat antibiotic resistance. *Advanced Drug Delivery Reviews*, **2014**, 78, 14-27.
4. C. G. Wermuth, Selective optimization of side activities: Another way for drug discovery. *Journal of Medicinal Chemistry*, **2004**, 47, 1303-1314.
5. Rauwerda, H. Roos, M. Hertzberger B. O. and Breit, T. M. The promise of a virtual lab in drug discovery. *Drug Discovery Today*, **2006**, 11, 228-236.
6. Edwards, E. I. Epton R. and Marr, G. A new class of semi-synthetic antibiotics: ferrocenyl-penicillins and cephalosporins. *Journal of Organometallic Chemistry*, **1976**, 107, 351-357.
7. Razafimahefa, D. E. Ralambomanana, D. A. Hammouche, L. Pélinski, L. Lauvagie, S. Bebear C., Brocard J. and Maugein, J. Synthesis and antimycobacterial activity of ferrocenyl ethambutol analogues and ferrocenyl diamines. *Bioorganic & Medicinal Chemistry Letters*, **2005**, 15, 2301-2303.
8. Andrianina Ralambomanana D., Razafimahefa-Ramilison, D. E.. Rakotohova, A. C. M Maugein J. and Pélinski, L. Synthesis and antitubercular activity of ferrocenyl diaminoalcohols and diamines**,** *Bioorganic & Medicinal Chemistry*, **2008**, 16, 9546-9553.

9. Fang, J. Jin, Z. Li Z.and Liu, W. Synthesis, structure and antibacterial activities of novel ferrocenyl-containing 1-phenyl-3-ferrocenyl-4-triazolyl-5-aryl-dihydropyrazole derivatives. *Journal of Organometallic Chemistry*, **2003**, 674, 1-9.
10. Damljanovic, I. Vukicevic, M. Radulovic, N.. Palic, R Ellmerer, E. Ratkovic, Z. Joksovic M. D. and Vukicevic, R. D. Synthesis and antimicrobial activity of some new pyrazole derivatives containing a ferrocene unit. *Bioorganic & Medicinal Chemistry Letters*, **2009**, 19, 1093-1096.
11. Chantson, J. T. Falzacappa, M. V. V. Crovella S. and Metzler-Nolte, N. Antibacterial activities of ferrocenoyl-and cobaltocenium-peptide bioconjugates. *Journal of Organometallic Chemistry,* **2005**, 690, 4564-4572.
12. El Arbi, M. Pigeon, P. Top, S. Rhouma, A. Aifa, S. Rebai, A. VessiÃ¨res, A. Plamont M.-A. and Jaouen, G. R. Evaluation of bactericidal and fungicidal activity of ferrocenyl or phenyl derivatives in the diphenyl butene series. *Journal of Organometallic Chemistry*, **2011**, 696, 1038-1048.
13. Fouda, M. F. R., Abd-Elzaher, M. M. Abdelsamaia, R. A.and Labib, A. A. On the medicinal chemistry of ferrocene. *Applied Organometallic Chemistry*, **2007**, 21, 613-625.
14. Quirante, J. Dubar, F. Gonzalez, A. Lopez, C. Cascante, M. Cortés, R. Forfar, I. Pradines B. and Biot, C. Ferrocene-indole hybrids for cancer and malaria therapy. *Journal of Organometallic Chemistry*, **2011**, 696, 1011-1017.
15. Braga S. S. and Silva, A. M. S.A new age for iron: antitumoral ferrocenes. *Organometallics*, **2013**, 32, 5626-5639.
16. Görmen, M. Sylla-Iyarreta Veitía, M., Trigui, F., El Arbi, M. and Ferroud, C. *Journal of Organometallic Chemistry*, **2015**, 794, 274-281.
17. Görmen, M. Pigeon, P. Top, S. Hillard, E. A. Huché, M. Hartinger, C. G. de Montigny, F. Plamont, M.-A. Vessières A.and Jaouen, G. R. Synthesis, Cytotoxicity, and COMPARE Analysis of Ferrocene and [3]Ferrocenophane Tetrasubstituted Olefin Derivatives against Human Cancer Cells. *ChemMedChem*, **2010**, 5, 2039-2050.
18. S. Top, A. Vessières, G. Leclercq, J. Quivy, J. Tang, J. Vaissermann, M. Huché and G. Jaouen, Synthesis, Biochemical Properties and Molecular Modelling Studies of Organometallic Specific Estrogen Receptor Modulators (SERMs), the Ferrocifens and Hydroxyferrocifens: Evidence for an Antiproliferative Effect of Hydroxyferrocifens on both Hormone-Dependent and Hormone-Independent Breast Cancer Cell Lines**.** *Chemistry – A European Journal*, **2003**, 9, 5223-5236.
19. O'Neill A. J. and Chopra, I. Preclinical evaluation of novel antibacterial agents by microbiological and molecular techniques. *Expert Opinion on Investigational Drugs*, **2004**, 13, 1045-1063.
20. Seto, M. Aramaki, Y. Okawa, T. Miyamoto, N. Aikawa, K. Kanzaki, N. Niwa, S.-i. Iizawa, Y., Baba M. and Shiraishi, M. Orally active CCR5 antagonists as anti-HIV-1 agents: synthesis and biological activity of 1-benzothiepine 1,1-dioxide and 1-benzazepine derivatives containing a tertiary amine moiety. *Chemical and Pharmaceutical Bulletin*, **2004**, 52, 577-590.
21. Guay D., Hamel, P. Blouin, M. Brideau, C. Chan, C. C. Chauret, N. Ducharme, Y. Huang, Z. Girard, M. Jones, T. R. Laliberté, F. Masson, P. McAuliffe, M. Piechuta, H. Silva, J. Young R.

N. and Girard Y. Discovery of L-791,943: A potent, selective, non emetic and orally active phosphodiesterase-4 inhibitor**.** *Bioorganic & Medicinal Chemistry Letters*, 2002, 12, 1457-1461.

22. Haginoya, N. Kobayashi, S. Komoriya, S. Yoshino, T. Nagata, T. Hirokawa Y. and Nagahara, T. Design, synthesis, and biological activity of non-amidine factor Xa inhibitors containing pyridine *N*-oxide and 2-carbamoylthiazole units. *Bioorganic & Medicinal Chemistry*, 2004, 12, 5579-5586.

SciForum
Mol2Net

A Palladium NCP Pincer Complex as an Efficient Catalyst for Intramolecular Direct Arylation

Nerea Conde, Raul SanMartin[2]*, Fátima Churruca,[3] María Teresa Herrero[4] and Esther Domínguez[5]*

[1] Department of Organic Chemistry II, Faculty of Science and Technology, University of the Basque Country (UPV/EHU), Sarriena auzoa, z/g 48940 Leioa, Spain. nerea.conde@ehu.eus
[2] raul.sanmartin@ehu.eus
[3] f.txurruka@googlemail.com
[4] mariateresa.herrero@ehu.eus
[5] esther.dominguez@ehu.eus
*E-Mail: raul.sanmartin@ehu.eus; Tel.: +034-946-014-435; Fax: +034-946-012-748.

Published: 4 December 2015

Abstract: CH functionalization is a convenient methodology with broad application in total synthesis and medicinal chemistry that provides the same products as already well established cross-coupling methodologies, but with the advantage of avoiding the use of metallic species such as Grignard, boron, silicon and tin compounds. Herein, we report an unprecedented palladium-catalyzed intramolecular direct arylation for the general access of phenanthridinones under very low catalyst loadings. With only 0.05 mol%, a palladium NCP pincer complex promotes efficiently the direct functionalization of a series of *N*-arylbenzanilides and *N*-arylsulfonamides which constitutes an effective, versatil and environmentally attractive procedure for the preparation of phenanthridinones, biaryl sultams and related heterocyclic derivatives.

Keywords: biaryl sultams; CH arylation; palladium; phenanthridinones; pincer complexes

1. Introduction

Palladium-catalyzed direct functionalization of C-H bonds has attracted significant attention over the last years, since the development of innovative, environmentally friendly and highly efficient synthetic strategies has became a priority in industry as well as in academia.[1]

In this context, synthetic approaches based on palladium-catalyzed direct functionalization of arenes have been developed to synthesize phenanthridinone derivates and related lactams, scaffolds found in many natural products which exhibit remarkable biological and pharmaceutical properties. Though these methods have proven to be among the most versatile for the synthesis of such structures, they usually require big amounts of catalysts (2-10 mol%),[2] Therefore, the development of a novel palladium-catalyzed approach for the direct functionalization of arenes using very low catalyst loadings would provide a cost-effective and environmentally very attractive procedure. Besides, such method would prevent metal contamination of the product which constitutes an interesting advantage regarding its potential application in medicinal chemistry.

Our group has applied palladium pincer-type complexes as very active catalysts or pre-catalysts (cat. loading ≤ 0.1 mol%) in a variety of chemical transformations.[3] The power of pincer complexes lies in their unique balance of stability vs. reactivity, which confers them extraordinary catalytic performances.[4] Thus, we envisaged that such palladium complexes would be the suitable tool to carry out a more efficient direct arylation of arenes with low catalyst loadings.[5] Herein we report an unprecedented palladium-catalyzed intramolecular direct arylation of *N*-substituted *o*-bromobenzanilides and benzosulfonamides for the general access of phenanthridinones and related biaryl sultams using palladium pincer complex **1** under very low catalyst loadings.

.

2. Results and Discussion

A series of screening experiments were conducted by employing a set of palladium pincer complexes prepared by our group[3] in only 0.1 mol% with *N*-methyl-*N*-phenyl-2-bromobenzamide **2a** as substrate. In contrast with other bases/solvents assayed, the use of PCN catalyst **1** produced the desired phenanthridinone **3a** in a 24% yield using KOAc as base in DMA (entry 1).

The effect of other solvents, the amount of base, solvent concentration, the addition of Brønsted acids, tetrabutylammonium bromide or cationic surfactants and, especially, the catalyst loadings was examined (entries 2-20). After careful experimentation, we succeeded to obtain phenanthridinone **3a** in a very good yield (88%, entry 18) with only 0.05 mol% of palladium pincer complex **1** using 3 equiv of KOAc in a relatively concentrated solution of DMA-H_2O (9:1, 0.3M). Moreover, a further decrease of the catalyst loading down to 0.01 mol% was also possible, affording desired phenanthridinone **3a** although at the cost of lower yields (entries 19-20) even at longer reaction times.

To the best of our knowledge, these values represent the lowest catalyst loadings achieved so far for any biaryl coupling of an aryl halide with a nonfunctionalized arene.[6] It should be also pointed out that the latter reactions were carried out in the air atmosphere with no effect on the reaction yield.

With the establishment of an optimal catalyst loading of 0.05 mol% as the most effective, the generality and scope of the reaction were studied.

As summarized in Table 2, the functional tolerance of this procedure was observed by synthesizing various phenanthridinones and related heterocyclic quinolinones with good to excellent yields. The electronic nature of the substituents seemed to have a little effect on the product yields. Besides, the reaction with different aromatics as naphthalene and heterocycles proceeded selectively in this C(sp^2)-H arylation (78-98%). Even sterically hindered substrates as

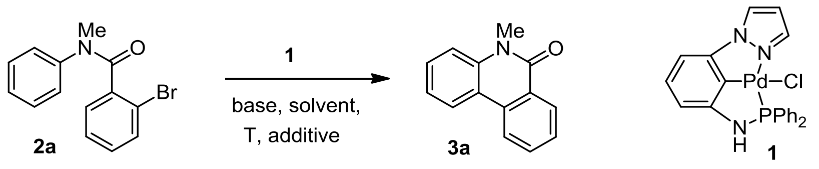
N-cyclohexyl amide **2d** afforded the desired product **3d** in an 86% yield.

We also investigated the applicability of this protocol to structurally related *o*-halo-*N*-arylsulfonamides. Accordingly, a series of *o*-bromo-*N*-(hetero)arylbenzenesulfonamides **2i-s** were readily prepared and reacted with only 0.05 mol% pincer catalyst **1** (Table 2). To our delight, the direct functionalization of *N*-(hetero)arylsulfonamides with such low catalyst loadings afforded regioselectively the corresponding biaryl sultams.

Although the yields obtained in our case are in accordance with literature precedents and, on average, not as high as the ones obtained for the corresponding 6-membered ring sultams **3h-o**, the significantly fewer amounts (0.05-0.09 mol%) of catalyst required turn our protocol into the most effective approach to benzoisothiazoloindoles and related heterocycles.

The measurement of the palladium content in benzothiazinodihydroquinoline **3n** was conducted using IPC-MS and determined as low as 0.29 ppm. Therefore, our method also offers an additional benefit regarding the avoidance of scavenger resins or further purification steps in order to suppress metal contamination in the products.

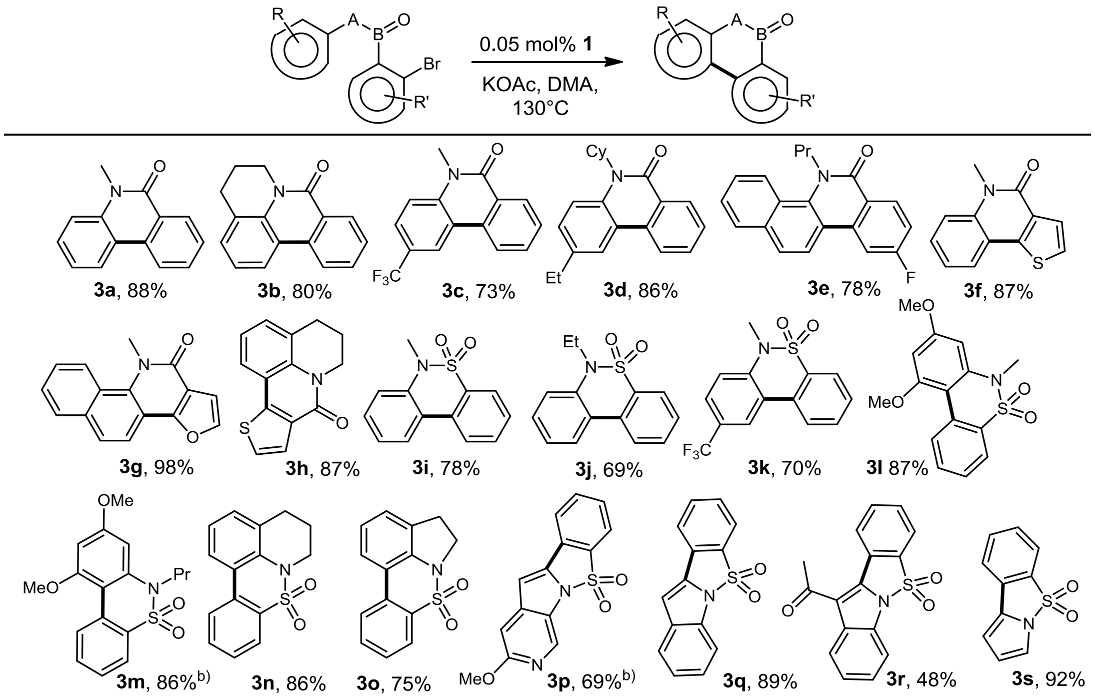
Table 1. Selection of optimization assays.[a]



Entry	Catalyst [Pd]	Base	Additive	Solvent	Yield [%][b]
1	0.1 mol%	KOAc	-	DMA	24
2	0.1 mol%	KOAc	-	DMF	-[c]
3	0.1 mol%	KOAc	-	DMPU	8
4	0.1 mol%	KOAc	-	DMI	-[c]
5	0.1 mol%	K_2CO_3	20 mol% benzoic acid	DMA	17
6	0.1 mol%	KOAc	20 mol% BA[d]	DMA	<5
7	0.1 mol%	KOAc	25 mol% surfactant[e]	DMA	7-18
8	0.1 mol%	KOAc	25 mol% TBAB	DMA	19
9	0.1 mol%	KOAc	-	DMA/*o*-xylene (1:1)	23
10	0.1 mol%	KOAc	-	DMA-THF (1:1)	51
11	0.1 mol%	KOAc	-	DMA-H_2O (9.5:0.5)	57
12	0.1 mol%	KOAc	-	DMA-H_2O (9:1)	68
13	0.1 mol%	KOAc	-	DMA-H_2O (7.5:2.5)	<5
14[f]	0.1 mol%	KOAc	-	DMA-H_2O (9:1)	68
15[g]	0.1 mol%	KOAc	-	DMA-H_2O (9:1)	75
16[f, g]	0.1 mol%	KOAc	-	DMA-H_2O (9:1)	83
17[g]	0.05 mol%	KOAc	-	DMA-H_2O (9:1)	70
18[f, g]	**0.05 mol%**	**KOAc**	**-**	**DMA-H_2O (9:1)**	**89 (88)**
19[f, g, h]	0.03 mol%	KOAc	-	DMA-H_2O (9:1)	78
20[f, g, i]	0.01 mol%	KOAc	-	DMA-H_2O (9:1)	57

[a] Reaction conditions: **2a** (1 equiv.), **1** (0.1 mol%), base (1.5 equiv.), solvent (0.06 M), 130°C, sealed tube, 20h under Ar. DMPU: N,N'-Dimetilpropilenourea; DMI: 1,3-Dimethyl-2-imidazolidinone. [b] Determined by ^{1}H NMR. Diethylene glycol dimethyl ether was used as internal standard. Isolated yield in parentheses. [c] Starting material. [d] BA: Brønsted acids (pivaloic, benzoic and *p*-toluensulfonic acid). [e] CTAB: Hexadecyltrimethylammonium bromide; DDA: Dimethyldioctadecylammonium bromide. [f] 3.0 eq. of KOAc [g] Solvent (0.3 M) [h] 48h. [i] 96h.

Table 2. Intramolecular direct arylation of arenes.[a)]



[a)] Reaction conditions: **2** (1 equiv.), **1** (0.05 mol%), KOAc (3 equiv.), DMA-H_2O (9:1, 0.35M), 130°C, sealed tube, 20h under air. Isolated yields. [b)] 0.09 mol% of **1** was used.

3. Materials and Methods

General procedure for the direct arylation of arenes

DMA (0.8 mL) and water (0.1 mL) were added to a heavy-wall pressure tube charged with substrate **2** (0.35 mmol) and KOAc (1.05 mmol) at room temperature. Then, a solution of pincer complex **1** in DMA (1.75 mM, 0.1 mL, 0.175 μmol of **1**) was added, the tube was closed and it was heated to 130 °C for 20 h. After cooling, the crude was diluted with H_2O (2 mL) and washed with EtOAc (2 x 3 mL). The combined organic phase was dried over anhydrous sodium sulfate and the solvent was evaporated under reduced pressure. The residue was purified by flash column chromatography (EtOAc in hexane) to give the desired product **3**.

.

4. Conclusions

In summary, we have developed a method for the intramolecualr direct arylation of arenes *via* C-H bond functionalization at very low catalyst loadings. With only 0.05 mol%, palladium pincer complex **1** promotes efficiently the direct functionalization of a series of *N*-arylbenzanilides and *N*-arylsulfonamides providing a novel versatile and sustainable access to phenanthridinones, biaryl sultams and related heterocyclic derivatives.

Acknowledgments

This research was supported by the Basque Government (IT-774-13), the Spanish Ministry of Economy and Competitiveness (CTQ2013-46970-P) and the University of the Basque Country (UFI QOSYC 11/12). N.C. thanks the Basque Government for a predoctoral scholarship. The authors also thank Petronor, S.A. for generous donation of hexane. Finally, technical and human support provided by SGIker (UPV/EHU, MICINN, GV/EJ, ESF) is gratefully acknowledged.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. For selected recent reviews see: a) Wencel-Delord, J.¸ Droge, T.; Liu, F.; Glorius, F. *Towards mild metal-catalyzed C–H bond activation Chem. Soc. Rev*. 2011, *40*, 4740-4771; b) Rossi, R. ; Bellina, F.; Lessi, M.; Manzini, C. *Cross-Coupling of Heteroarenes by C—H Functionalization: Recent Progress towards Direct Arylation and Heteroarylation Reactions Involving Heteroarenes Containing One Heteroatom. Adv. Synth. Catal.* 2014, *356*, 17-117.

2. a) Bhakuni, B. S.; Kumar, A.; Balkrishna, S. J.; Sheikh, J. A.; Konar, S.; Kumar, S. *KO^{t}Bu Mediated Synthesis of Phenanthridinones and Dibenzoazepinones. Org. Lett*. 2012, *17*, 2838-2841; b) De, S.; Mishra, S.; Kakde, B. N.; Dey, D.; Bisai, A. *J. Org. Chem*. 2013, *78*, 7823-7844.

3. See for example: a) Churruca, F.; SanMartin, R.; Inés, B.; Tellitu, I.; Domínguez, E. *Expeditious Approach to Pyrrolophenanthridones, Phenanthridines, and Benzo[c]phenanthridines via Organocatalytic Direct Biaryl-Coupling Promoted by Potassium tert-Butoxide. Adv. Synth. Catal.* 2006, *348*, 1836-1840; b) Inés, B.; SanMartin, R.; Churruca, F.; Dominguez, E.; Urtiaga, M. K.; Arriortua, M. I. *A Nonsymmetric Pincer-Type Palladium Catalyst In Suzuki, Sonogashira, and Hiyama Couplings in Neat Water. Organometallics* 2008, *27*, 2833-2839; c) Urgoitia, G.; SanMartin, R.; Herrero, M. T.; Domínguez, E. *Palladium NCN and CNC pincer complexes as exceptionally active catalysts for aerobic oxidation in sustainable media. Green Chem*. 2011, *13*, 2161-2166.

4. *The Chemistry of Pincer Compounds*, 1st ed., Morales-Morales, D.; Jensen, C. M. Eds., Elsevier, Amsterdam, 2007.

5. See for example a) Beccalli, E. M.; Broggini, G.; Martinelli, M.; Paladino, G.; Zoni, C. *Synthesis of Tricyclic Quinolones and Naphthyridones by Intramolecular Heck Cyclization of Functionalized Electron-Rich Heterocycles*. *Eur. J. Org. Chem*. 2005, 2091-2096; b) Bheeter, C. B.; Bera, J. K.; Doucet, H. *Palladium-Catalysed Intramolecular Direct Arylation of 2-Bromobenzenesulfonic Acid Derivatives. Adv. Synth. Catal*. 2012, *354*, 3533-3538.

6. For the only example of biaryl coupling of an aryl halide with a nonfunctionalized arene using < 1 mol% catalyst see: Campeau, L.-C.; Parisien, M.; Leblanc, M.; Fagnou, K. *Biaryl Synthesis via Direct Arylation: Establishment of an Efficient Catalyst for Intramolecular Processes*. *J. Am. Chem. Soc*. 2004, *126*, 9186-9187.

SciForum
Mol2Net

Efficient Aerobic Oxidation of Arylcarbinols and Arylmethylene Compounds Mediated by Nickel (II)/1,2,4-Triazole Ligand Catalyst System

Garazi Urgoitia[1], Raul SanMartin[2]*, María Teresa Herrero[3] and Esther Domínguez[4]*

[1] Department of Organic Chemistry II, Faculty of Science and Technology, University of the Basque Country (UPV/EHU), Sarriena auzoa, z/g 48940 Leioa, Spain. garazi.urgoitia@ehu.eus
[2] raul.sanmartin@ehu.eus
[3] mariateresa.herrero@ehu.eus
[4] esther.dominguez@ehu.eus
*E-Mail: raul.sanmartin@ehu.eus; Tel.: +034-946-014-435; Fax: +034-946-012-748.

Published: 4 December 2015

Abstract: The oxidation of benzylic alcohols is a ubiquitous transformation in organic chemistry due to the relevance of aryl ketones, present in natural products and pharmaceutically active compounds and also intermediates in the synthesis of agrochemicals, medicines and other functional materials. Numerous oxidizing agents are available to effect this key reaction. In most instances, these reagents are required in stoichiometric amounts and are usually toxic, or hazardous, or both. In this context, considering environmental and safety issues, oxygen would be the reagent of choice for such organic transformation. Although much effort has been devoted to aerobic oxidation of benzyl alcohols, the amounts of metal catalyst are still relatively high and, sometimes, oxygen pressures above 5 atms.

In this context, we have discovered that systems based upon 1,2,4-triazole ligands and simple Ni (II) salts show extremely high catalytic activity in the aerobic oxidations of alcohols. In fact, primary and secondary benzylic alcohols have provided the corresponding carbonylic derivatives by using molecular oxygen at atmospheric pressure, in the presence of a combination of nickel(II) bromide and 1,2,4-triazole or, alternatively, a 1,2,4-triazole pincer ligand at such a low catalytic loading as 10^{-5} mol%. This catalytic system has proven to be efficient also for the benzylic C-H oxidation.

Keywords: aerobic oxidation, benzylic alcohols, Nickel, pincer ligand, triazole

Mol2Net YouTube channel*: http://bit.do/mol2net-tube*

1. Introduction

Molecular oxygen is the ideal oxidant for chemical transformations. Therefore, in the last years, a number of metal catalysts for different oxygen-mediated oxidative processes have been developed. Taking alcohol oxidation to carbonyl compounds as an illustrative example, this reaction can be carried out using different transition metals,[1] since the yields and selectivities of chain radical non-catalyzed autoxidation of alcohols and saturated hydrocarbons are often low.[2] Catalyst loadings generally range 0.5-10 mol% of [M], although significantly lower amounts (0.1-0.3 mol%) have been reported in some cases.[3] Regarding reaction media, flammable organic solvents have been mainly employed and just a few examples of oxidations have carried out in more sustainable media.[4]

Removal of metal traces is a serious concern in fine chemicals production. A valuable strategy to achieve this goal is just to decrease the catalyst loading to a point below regulatory requirements, which is particularly difficult for many transition metal catalysts, with very low levels allowed in drug products.[5]

In this paper we wish to present an outstanding catalytic system for the aerobic oxidation of alcohols and arylmethylene compounds based on 1,2,4-triazole type ligands and ($NiBr_2$) that has allowed us to avoid precious metals and fulfill the mentioned regulatory requirements.

2. Results and Discussion

In order to analyze the influence of the coordination degree and chelating effects in the reaction outcome, two triazole ligands, simple 1,2,4-triazole **1** and pincer-type bis-triazolyl ligand **2** were assayed. On the basis of the encouraging results from the initial assays for the nickel-catalyzed oxidation of 1-phenylethanol, we decided to use O_2, $NiBr_2$, **1** or **2**, PEG-400, 120ºC and decrease gradually the catalytic amount down to 10^{-5} mol% of Ni and **1**/**2**. Blank experiments also showed the need of both nickel and triazole ligands, since nickel bromide alone provided low yields (20%) at 0.01 mol%, and no product was detected with lower amounts of the nickel halide.

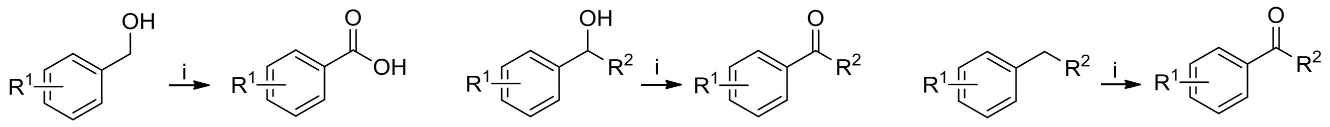
The optimized reaction conditions were accordingly applied to a number of primary and secondary benzyl alcohols, providing good to excellent results regardless the nature of the starting carbinol. As shown in Table 1, oxidation of primary alcohols provided the corresponding carboxylic acids.

Then, we assayed the same conditions in the benzylic C-H bond oxidation of a series of methylene compounds, and to our delight, the carbonyl compounds displayed in Table 1 were obtained in good yields. In both carbinol and methylene oxidation reactions a slightly better catalytic profile was found for the $NiBr_2$/**2** system, as better yields were observed in all cases.

In addition to the fact that oxygen mediates this transformation at atmospheric pressure, it should be also pointed that the reaction is conducted in an environmentally friendly solvent, PEG-400. The role of this solvent in the reported oxidative process might be related to its unusual coordinating properties that remind of crown ethers.[6]

The combination of the above properties and those of 1,2,4-triazoles **1**-**2** enhance nickel catalytic properties to an exceptional catalytic profile that avoid the need of further purification steps in order to remove metal traces.

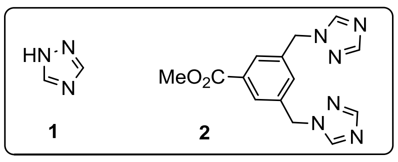
Table 1. Oxidation of alcohols and methylene compounds. General substrate scope.[a



	1 (%)	2 (%)
R^1= H	98	98
R^1= *p*-^{i}Pr	70[b]	94[b]
R^1= *p*-Et	67[b]	94[b]
R^1= *p*-Me	70[b]	90[b]
R^1= *p*-CF_3	90[b]	97[b]
R^1= *m*-OMe	75[c]	90[c]
R^1= *p*-OMe	67[c]	80[c]
R^1= *m*-OPh	60[b]	88[b]

	1 (%)	2 (%)
R^1= H; R^2= Me	80	97
R^1= H; R^2= CN	96	97
R^1= H; R^2= Ph	74	93
R^1= *p*-Me; R^2= Me	90	98
R^1= H; R^2= Et	86	94
R^1= *o*-OMe; R^2= Me	75[b]	98[b]
R^1= H; R^2= ^{t}Bu	70[b]	90[b]
R^1= *p*-Cl; R^2= Me	82	87
R^1= *H*; R^2= *o*-tol	89	88

	1 (%)	2 (%)
R^1= H; R^2= Me	90	97
R^1= H; R^2= Ph	84	97
R^1= H; R^2= 4-Py	50[b]	70[b]
R^1= *p*-Me; R^2= CN	79[b]	82[b]



1 2

[a] Isolated yields. i: O_2 (1 atm), $NiBr_2$ (10^{-5} mol%), **1** or **2** (10^{-5} mol%), NaOAc, PEG-400, 120ºC, 8h. [b] 72h. [c] 96h.

3. Materials and Methods

A round bottom flask equipped with a magnetic stirrer bar was charged with the alcohol (1 mmol), NaOAc (8.0 mg, 0.1 mmol), $NiBr_2$ (2.7 10^{-5} mg, 10^{-7} mmol), **1** (8.6 10^{-6} mg, 10^{-7} mmol,) and PEG 400 (1 mL) at room temperature. The system was purged with molecular oxygen, and an oxygen-filled balloon (1-1.2 atm) was connected. The mixture was heated at 120 ºC under stirring for 24 h. The reaction outcome was monitored by ^{1}H-NMR. Upon completion, the mixture was cooled to room temperature and water was added (50 mL aprox.). The resulting solution was acidified with HCl 1M (pH≈1-2) and extracted with Et_2O (4 x 6 mL) and the combined organic layers were washed with brine, dried over anhydrous Na_2SO_4 and evaporated *in vacuo* to give a residue which was purified by flash column chromatography using hexane:ethyl acetate as eluent.

4. Conclusions

In summary, the combination of nickel(II) bromide and the 1,2,4-triazole pincer ligand **2** constitutes a highly active catalytic system for the aerobic oxidation of arylcarbinols and arylmethylene compounds. In addition to the advantages in terms of safety and sustainability associated to the use of molecular oxygen at atmospheric pressure and PEG-400, the infinitesimal amounts of the metal/ligand system allow isolation of "*metal-free*" compounds for pharmaceutical uses.

Acknowledgments

This research was supported by the Basque Government (IT-774-13), the Spanish Ministry of Economy and Competitiveness (CTQ2013-46970-P) and the University of the Basque Country (UFI QOSYC 11/12). G.U. thanks the University of the Basque Country (UPV/EHU) for a predoctoral scholarship. The authors also thank Petronor, S.A. for generous donation of hexane. Finally, technical and human support provided by SGIker (UPV/EHU, MICINN, GV/EJ, ESF) is gratefully acknowledged.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. See for example: [a] Melero, C.; Shishilov, V; Álvarez, V; Palma, V; Cámpora*,* V. *Well-defined alkylpalladium complexes with pyridine-carboxylate ligands as catalysts for the aerobic oxidation of alcohols. Dalton Trans.* 2012*, 41*, 14087-14100. [b] Verho, O.; Dilenstam, M.D.V.; Kärkäs, M.D.; Johnston, V; Åkermark, T.; Bäckvall, J.E.; Åkermark, B. *Application and Mechanistic Studies of a Water-Oxidation Catalyst in Alcohol Oxidation by Employing Oxygen-Transfer Reagents. Chem. Eur. J.* 2012, *18*, 16947-16954.
2. Shilov, A. E.; Shul'pin, G. B. *Activation and catalytic reactions of saturated hydrocarbons in the presence of metal complexes*, Kluwer Academic, Dordrecht, Holland, 2002**.**
3. See for example: Liu, J.; Ma, S. *Iron-Catalyzed Aerobic Oxidation of Allylic Alcohols: The Issue of C═C Bond Isomerization. Org. Lett.* 2013, *15*, 5150-5153.
4. See for example: Liu, L.; Yu, M.; Wayland, B.B.; Fu, X. *Aerobic oxidation of alcohols catalyzed by rhodium(III) porphyrin complexes in water: reactivity and mechanistic studies*. *Chem. Commun.* 2010, *46*, 6353-6355.
5. See for example: Thayer, A. *Metal scavengers and immobilized catalysts may make for cleaner pharmaceutical products. Chem. Eng. News* 2005, *83*, 55-58.
6. Chen, J.; Spear, V; Huddleston, J. G.; Rogers, R. D. *Polyethylene glycol and solutions of polyethylene glycol as green reaction media*. *Green Chem.* 2005, *7*, 64-82.

SciForum
Mol2Net

Effect of Neuronal Nitric Oxide Synthase Inhibitors and Antioxidants on the Development of Tolerance by Different Opioid Agonists in the Rat Locus Coeruleus

Patricia Pablos [1],*, Aitziber Mendiguren [1] and Joseba Pineda [1]

[1] Department of Pharmacology, Faculty of Medicine and Odontology, University of the Basque Country (UPV/EHU), E-48940 Leioa, Bizkaia, Spain

* Author to whom correspondence should be addressed; e-mail: inespatricia.pablos@ehu.es

Published: 4 December 2015

Abstract: Nitric oxide (NO) is involved in acute μ-opioid receptor (MOR) desensitization in the locus coeruleus (LC) and in the neuroadaptations following chronic morphine administration. However, the role of NO and NO-derived reactive oxygen species (ROS) in the development of cellular tolerance to different opioids remains unclear. Herein, we examined the effect of the selective neuronal nitric oxide synthase (nNOS) inhibitor 7-nitroindazole (7-NI; 30 mg/kg/12 h, i.p.) and the antioxidants Trolox + ascorbic acid (TX+AA; 40 and 100 mg/kg/day, respectively, i.p.) and U-74389G (10 mg/kg/day, i.p.) on the development of cellular tolerance induced by morphine, methadone and fentanyl. For induction of morphine tolerance, rats were treated subcutaneously with a slow release emulsion containing free base morphine (200 mg/kg, 3 days). For methadone (60 mg/kg/day, 6 days) and fentanyl (0.2 mg/kg/day, 7 days), tolerance was induced by s.c. implantation of osmotic pumps. Concentration-effect curves for the inhibitory effect of Met5-enkephalin (ME; 0.05-12.8 μM, 2x, 1 min) on the firing rate were performed by single-unit extracellular recordings of LC neurons from rat brain slices. Morphine, methadone and fentanyl treatments shifted to the right concentration-effect curves for ME and increased the EC_{50} by 2-4 folds. Co-administration of TX+AA or U-74389G in morphine-treated animals prevented the development of tolerance in LC neurons. Conversely, co-treatments with U-74389G or 7-NI failed to affect the induction of cellular tolerance after methadone or fentanyl treatments. Our results suggest that MOR agonists with different intrinsic efficacies cause variable degrees of cellular tolerance in LC cells. Moreover, NO/ROS pathways are differentially involved in opioid tolerance after prolonged treatments with morphine, methadone and fentanyl.

Keywords: slice; μ-opioid receptor; tolerance; reactive oxygen species; locus coeruleus; nitric oxide; morphine; methadone; fentanyl; electrophysiology

1. Introduction

Morphine, methadone or fentanyl are among the most common clinically used opioid agonists. However, their long-term utility is greatly limited due to the development of tolerance and dependence (Inturrisi 2002). Numerous mechanisms have been reported to contribute to opioid tolerance. Thus, tolerance may result from adaptive changes such as enhanced MOR desensitization, receptor internalization or receptor down-regulation, among other mechanisms (Williams et al., 2013). In addition, nitric oxide (NO) has been proposed to be involved in opioid tolerance (Heinzen and Pollack, 2004). In the brain, NO is produced by the neuronal NO synthase (nNOS) and targets the heme group of guanylate cyclase, which elevates 3',5'-cyclic guanosine monophosphate (cGMP) concentrations. Moreover, high, sustained concentrations of NO promote the generation of reactive nitrogen species and reactive oxygen species (ROS), such as the extremely oxidant molecule peroxynitrite (Radi, 2013).

The locus coeruleus (LC), the major noradrenergic nucleus in the brain, has long been used as a model for examining the cellular mechanisms of opioid tolerance and dependence (Nestler et al., 1994). It contains a homogeneous population of neurons that almost exclusively express the MOR (Williams and North, 1984). There is lacking evidence regarding the implication of NO/ROS pathway in the adaptations triggered by chronic treatments with different opioid agonists in the LC. Therefore, the aim of this work was to investigate whether NO and ROS regulate the tolerance induced by opioids with different pharmacologic profiles, such as morphine, methadone and fentanyl.

2. Results

Effect of chronic treatments with morphine, methadone and fentanyl on concentration-effect curves for ME in rat LC neurons

To evaluate the development of tolerance, concentration-effect curves for the inhibitory effect of ME were performed. Subchronic treatment with morphine (200 mg/kg, s.c., 3 days) induced a strong degree of tolerance in the LC, which was revealed by a rightward shift in the concentration-effect curves for ME and an increase in the EC_{50} of about 4 fold, as compared to the corresponding sham group (p < 0.005). Similarly, chronic treatment with methadone (60 mg/kg/day, s.c., 6 days) induced a rightward shift in the concentration-effect curves for ME and increased by about 2 fold the EC_{50} value when compared to the corresponding sham group (p < 0.005). Finally, chronic treatment with fentanyl (0.2 mg/kg/day, s.c., 7 days) also caused a rightward shift in the concentration-effect curves for ME with an increase of the EC_{50} of about 3 fold, as compared to the corresponding sham group (p < 0.005). These results indicate that morphine, methadone and fentanyl induce different degrees of tolerance to the inhibitory effect of ME in LC neurons. In all groups, the maximal effect of ME was 100% of baseline, which corresponded with an absolute inhibition from the basal firing rate.

It can be hypothesized that differences between morphine, methadone and fentanyl in their ability to induce receptor internalization and recycling (Alvarez et al., 2002; Virk and Williams, 2008) may contribute, at least in part, to the different degrees of cellular tolerance observed in this study.

Effect of the nNOS inhibitor 7-NI on opioid tolerance in rat LC neurons

Co-administration of the neuronal NOS inhibitor 7-NI in methadone- and fentanyl-treated rats (methadone/7-NI group and fentanyl/7-NI group) failed to prevent the development of tolerance. 7-NI administration in sham-treated animals (sham-7-NI group) did not modify the concentration-effect curves for ME in the LC as compared to the sham-vehicle group. No differences were found in the basal firing rate among groups.

Effect of antioxidants on opioid tolerance in rat LC neurons

Co-administration of the vitamin E analogue Trolox, together with ascorbic acid (AA), or the structurally unrelated antioxidant U-74389G in morphine-treated rats unclear. Further studies are needed to unmask the underlying mechanisms. (morphine/TX+AA group and morphine/U-74389G group, respectively) attenuated the development of cellular tolerance, which was shown by a blockade of the rightward shift of the concentration-effect curve in this group ($p < 0.05$ and $p < 0.01$, respectively, when compared to the corresponding sham group) (Fig. 1). Administration of TX+AA or U-74389G in sham animals failed to modify the concentration-effect curves for ME when compared to the sham/vehicle group.

On the contrary, co-treatment with U-74389G in rats treated chronically with methadone or fentanyl did not prevent the development of cellular tolerance induced by these opioids (Fig. 2). In all cases, no differences were found in the basal firing rate among groups. The mechanisms by which NO modulates morphine-. but not methadone- or fentanyl-induced tolerance via ROS generation remain

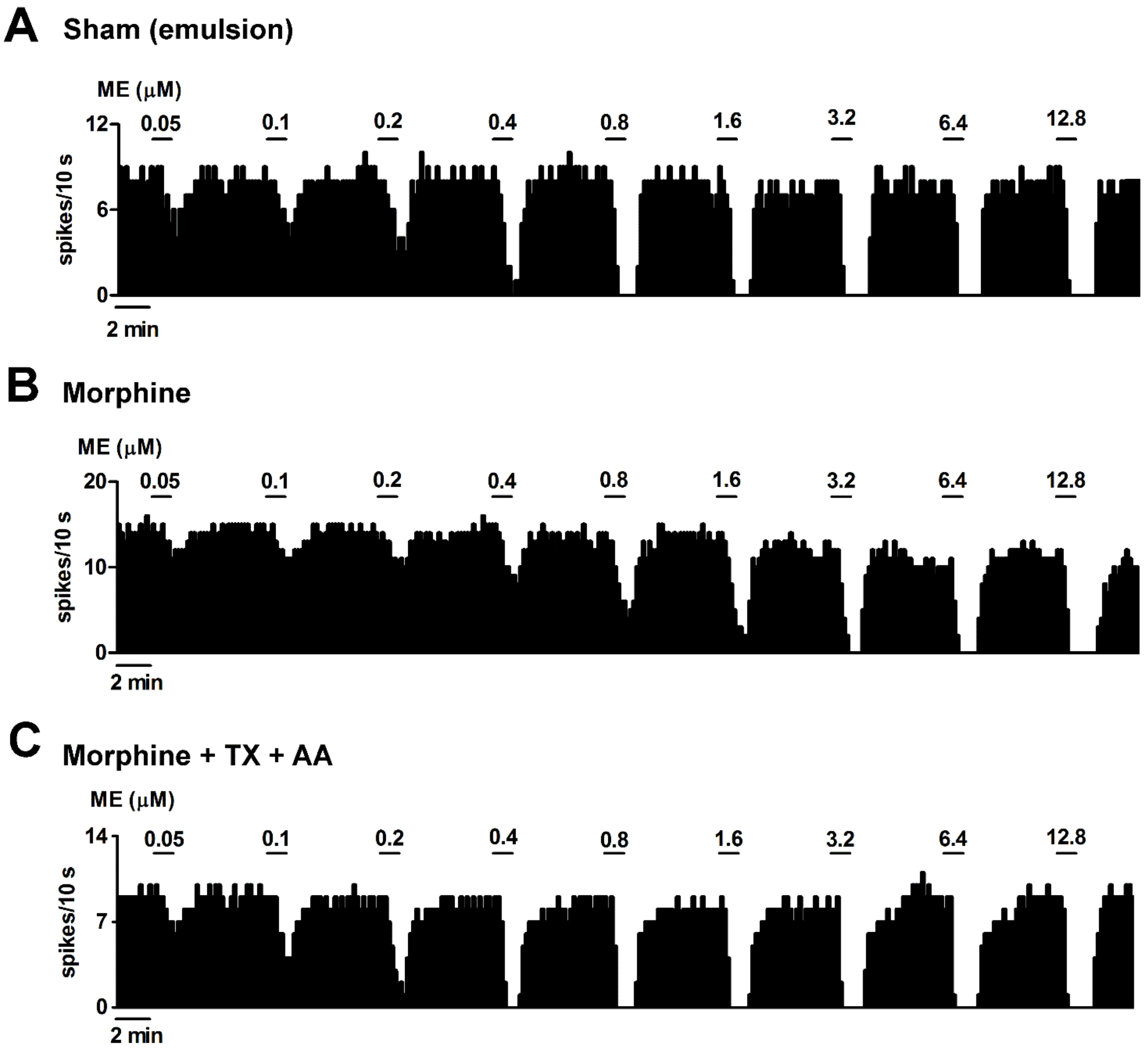


Figure 1. Effect of antioxidants Trolox + ascorbic acid (TX+AA) on the inhibition induced by ME in the LC of morphine-treated rats. **A, B, C**. Representative examples of firing-rate recording of LC cells from rats receiving the following treatments: sham (emulsion) (**A**), morphine (**B**) morphine + TX+AA (**C**). Each horizontal bar represents the period of application of each ME concentration (0.05 - 12.8 μM, 2x) and the vertical lines show the number of spikes recorded every 10 s. The inhibitory effect induced by each application was calculated as a percentage from the basal firing rate. Note that greater concentrations of ME are needed to inhibit the neuron activity in rats treated with morphine, compared with sham animals, which indicates the development of tolerance. Co-administration of TX+AA in morphine-treated rats increases the inhibitory effect of ME when compared to the morphine/vehicle group indicating an attenuation of celullar tolerance.

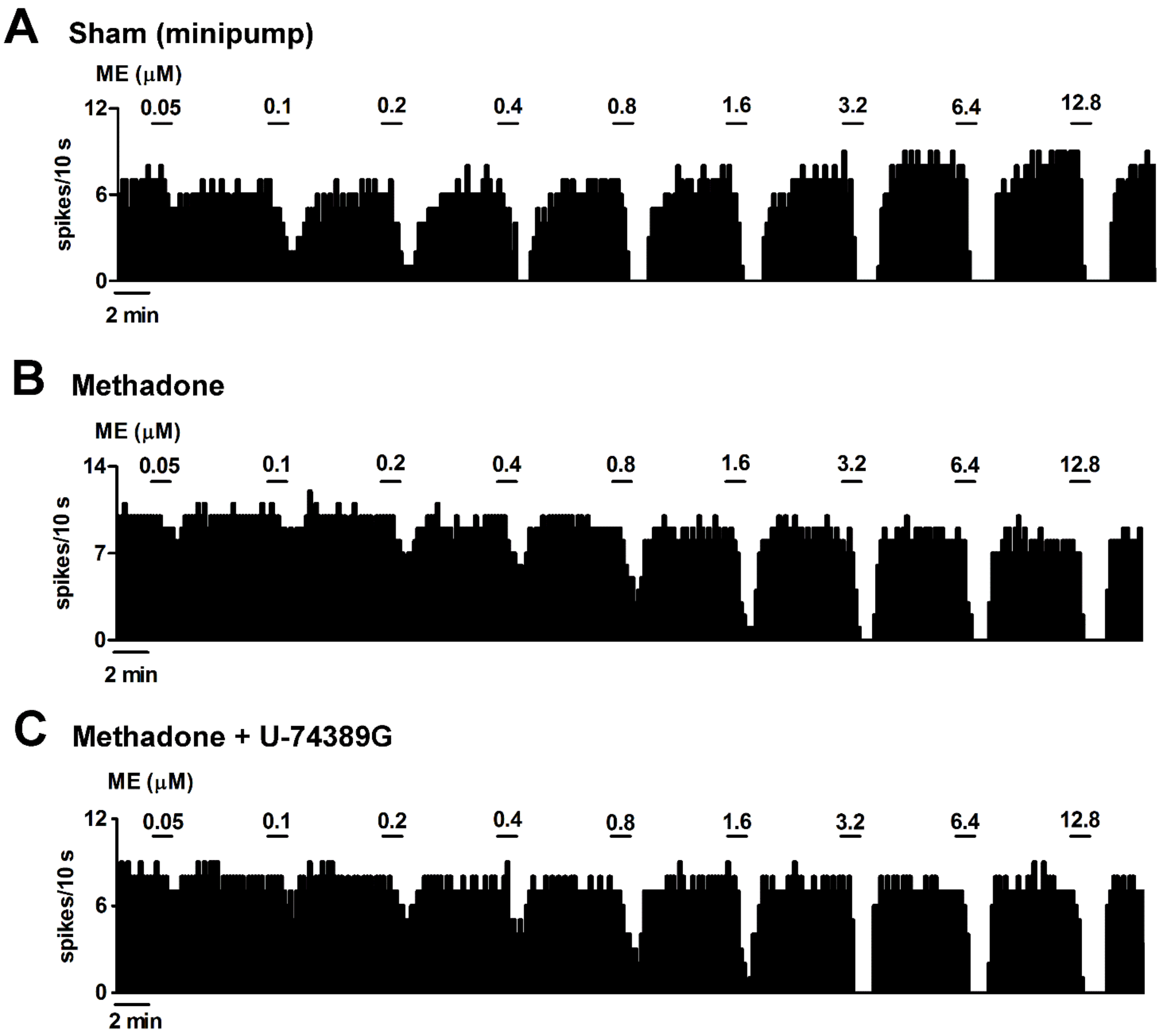


Figure 2. Effect of antioxidant U-74389G on the inhibition induced by ME in the LC of methadone-treated rats. **A, B, C**. Representative examples of firing-rate recording of LC cells from rats receiving the following treatments: sham (minipump) (**A**), methadone (**B**), methadone + U-74389G (**C**). Each horizontal bar represents the period of application of each ME concentration (0.05 - 12.8 μM, 2x) and the vertical lines show the number of spikes recorded every 10 s. The inhibitory effect induced by each application was calculated as a percentage from the basal firing rate. Note that greater concentrations of ME are needed to inhibit the neuron firing in rats treated with methadone, when compared to sham animals, indicative of tolerance However, co-administration of U-74389G in methadone-treated rats failed to prevent the development of cellular tolerance, so that concentration-effect curves for ME were not modified when compared to the methadone/vehicle group.

3. Materials and Methods

Animals and treatments Male adult Sprague-Dawley rats (200–300 g) were housed under standard laboratory conditions (22 ºC and 12-h light/dark cycles) with free access to food and water. Principles of laboratory animal care were followed in all experimental procedures reported in this manuscript. Experimental procedures were carried out in accordance with the European Community Council Directive on "Protection of Animals Used in Experimental and Other Scientific Purposes" (86/609/EEC). The use of animals for this study was also approved by the *Animal Care and Use Committee* of the University of the Basque Country. All the efforts were made to minimize animal suffering and to reduce the number of animals used.

For induction of morphine tolerance, an oily emulsion of morphine base (200 mg/kg) was subcutaneously (s.c.) injected in the back of the rat, under slight anesthesia with chloral hydrate (200 mg/kg, i.p.). Control animals were implanted with a sham emulsion, which contained the same vehicle for morphine (mannide monooleate, liquid paraffin, and NaCl). Then, morphine or sham animals were daily injected with the antioxidants Trolox (40 mg/kg) and ascorbic acid (100 mg/kg) (TX+AA), U-74389G (10 mg/kg) or 0.9% NaCl (saline) intraperitoneally (i.p.). Electrophysiological experiments were performed 72 h after emulsion implantation.

For chronic treatments with methadone or fentanyl, osmotic mini-pumps were implanted subcutaneously in the rat. Animals were treated with methadone (60 mg/kg/day) or fentanyl (0.2 mg/kg/day). Sham, methadone-, or fentanyl-treated animals were treated every 12 h with the nNOS inhibitor 7-NI (30 mg/kg) or its vehicle. In another group of experiments, sham, methadone, or fentanyl-receiving rats were daily injected with the antioxidant U-74389G (10 mg/kg) or its vehicle (saline), i.p.

In vitro electrophysiology

Brain slice preparation

Animals were anaesthetized with chloral hydrate (400 mg/kg, i.p.) and sacrified by decapitation. The brain was rapidly removed and a block of tissue containing the brainstem was placed in ice-cold artificial cerebrospinal fluid (aCSF) containing 130 mM NaCl, 3 mM KCl, 1.25 mM NaH_2PO_4, 10 mM D-glucose, 21 mM $NaHCO_3$, 2 mM $CaCl_2$, and 2 mM $MgSO_4$. Coronal slices of 500–600 µm thickness containing the LC were cut using a vibratome (FHC Inc., Brunswick, USA). The tissue was allowed to recover from the slicing for 90 min in oxygenated aCSF. Next, slices were placed on a nylon mesh in a modified Haas-type interface chamber maintained at 33°C and continuously perfused with oxygenated aCSF saturated with (95% O_2/ 5% CO_2, pH = 7.34–7.38) at a flow rate of 1–1.5 ml/min.

Recording procedures

Single-unit extracellular recordings of LC cells were performed as described previously (Mendiguren and Pineda, 2007). The recording electrode, an Omegadot glass micropipette was pulled and filled with NaCl (50 mM). The tip was broken back to a diameter of 2–5 µm (3–5 MΩ). The electrode was placed in the LC, visually identified as a dark oval area on the lateral borders of the central gray and the fourth ventricle, just anterior to the genu of the facial nerve. The extracellular signal from the electrode was passed through a high-input impedance amplifier and displayed on an oscilloscope and monitored with an audio unit. Individual neuronal spikes were isolated from the background noise with a window discriminator.

The firing rate was analyzed by a PC-based custom-made software, which generated histogram bars representing the cumulative number of spikes in consecutive 10 s bins (HFCP®, Cibertec S.A., Madrid, Spain).

Pharmacological procedures

We performed concentration-effect curves for the inhibitory effect of the MOR agonist ME in sham and opioid-treated animals. Thus, we perfused increasing concentrations of ME (0.05–12.8 μM, each concentration applied for 1 min) at 5-min intervals. ME was chosen because its washout is fast even at high concentrations and its action is mediated almost exclusively by MOR in LC neurons (Williams and North, 1984). The inhibitory effect of each ME concentration was calculated as follows:

$$E(\%) = \frac{FR_{pre} - FR_{post}}{FR_{basal}} \cdot 100$$

where FR_{pre} is the average firing rate for 60 s before each ME application, FR_{post} is the average firing rate for 90 s after each ME perfusion, and FR_{basal} is the firing rate for 60 s of each cell at the beginning of the recording. Curve fitting analysis was performed by the computer program GraphPad Prism (version 5.0 for Windows, San Diego, CA, USA) to obtain the best simple nonlinear fit to the following three-parameter logistic equation:

$$E(\%) = \frac{E_{max}}{1 + \left(\frac{EC_{50}}{A}\right)^{n}} \cdot 100$$

where E is the effect induced by each concentration of ME (A), Emax is the maximal effect, EC_{50} is the concentration of ME needed to elicit a 50% of the maximal effect, and n is the slope factor of the concentration-effect curve. These parameters were determined by the nonlinear analysis. EC_{50} values were finally expressed as negative logarithm (pEC_{50}).

Drugs and reagents

The following drugs were purchased from Sigma-Aldrich Química (Madrid, Spain): Fentanyl, methadone, L-ascorbic acid and 7-nitroindazole (7-NI). Met^5-enkephalin acetate salt (ME) was obtained from Bachem (Weil am Rhein, Germany). Morphine base was purchased from Alcaliber (Madrid, Spain). Trolox and U-74389G were obtained from Enzo Life Sciences (Lausen, Switzerland). For subchronic treatments with morphine, an oily emulsion containing morphine free base (200 mg/ml) in a mixture of mannide monooleate (Sigma-Aldrich Química S.A., Madrid, Spain), liquid paraffin (Sigma-Aldrich Química S.A., Madrid, Spain), and NaCl (0.9 %) (0.08:0.42:0.5, *v/v/v*) was prepared. Methadone, fentanyl and ascorbic acid were dissolved in saline. 7-NI was dissolved in peanut oil. Trolox was dissolved in NaOH (1 M), neutralized with HCl (1 M) and diluted to the last volume with saline. U-74389G was dissolved in 0.05 M HCl. The final pH prior to i.p administration was in all cases 6-7. ME stock solutions were prepared in Milli-Q water, stored at −25 °C and, on the day of the experiment, diluted in aCSF to their final volume.

Data analysis and statistics

Data are expressed as mean ± standard error of the mean (SEM). For statistical analysis, the EC_{50} values were transformed to the corresponding logarithmic data to convert them to a Gaussian distribution. Data among groups were compared by one-way analysis of variance

(ANOVA) followed by Tukey´s *post hoc* test using the computer program GraphPad Prism (version 5.0). A probability level of 0.05 was accepted as statistically significant.

4. Conclusions

Our results suggest that MOR agonists with different intrinsic efficacies cause variable degrees of cellular tolerance in LC cells. Moreover, NO/ROS pathways are differentially involved in opioid tolerance after prolonged treatments with morphine, methadone and fentanyl.

Acknowledgments

This work was supported by the Ministerio de Ciencia e Innovación [Grant SAF2008-03612] and the University of the Basque Country (UPV/EHU) [Grant GIU11/27 and Grant GIU14/29]. Pineda´s research group takes part in a network unit supported by the University of the Basque Country [UFI 11/35]. P. Pablos was supported by a predoctoral fellowship from the Basque Government. The experiments comply with the current laws of Spain.

Author Contributions

- P. Pablos performed the experiments and analyzed the data.
- A. Mendiguren and J. Pineda designed the research study
- P. Pablos, A. Mendiguren and J. Pineda wrote the manuscript

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Alvarez VA, Arttamangkul S, Dang V, Salem A, Whistler JL, Von Zastrow M, Grandy DK, Williams JT (2002) μ-Opioid receptors: Ligand-dependent activation of potassium conductance, desensitization, and internalization. J Neurosci 22:5769–5776.
2. Heinzen EL, Pollack GM (2004) The development of morphine antinociceptive tolerance in nitric oxide synthase-deficient mice. Biochem Pharmacol 67:735–741.
3. Inturrisi CE (2002) Clinical pharmacology of opioids for pain. Clin J Pain 18(Suppl 4), 3S–13S.
4. Mendiguren A, Pineda J (2007). CB(1) cannabinoid receptors inhibit the glutamatergic component of KCl-evoked excitation of locus coeruleus neurons in rat brain slices. Neuropharmacology 52(2):617-25.

5. Nestler EJ, Alreja M, Aghajanian GK (1994) Molecular and cellular mechanisms of opiate action: studies in the rat locus coeruleus. Brain Res Bull 35:521–528.
6. Radi R (2013) Peroxynitrite, a stealthy biological oxidant. J Biol Chem 288:26464–26472.
7. Virk MS, Williams JT (2008) Agonist-specific regulation of mu-opioid receptor desensitization and recovery from desensitization. Mol Pharmacol 73:1301–1308.
8. Williams JT, North RT (1984) Opiate-receptor interactions on single locus coeruleus neurons. Mol Pharmacol 26:489–497.
9. Williams JT, Ingram SL, Henderson G, Chavkin C, von Zastrow M, Schulz S, Koch T, Evans CJ, Christie MJ (2013) Regulation of mu-opioid receptors: desensitization, phosphorylation, internalization, and tolerance. Pharmacol Rev 65:223–254.

SciForum
Mol2Net

Solvent Accessible Surface Area Hot-Spot Detection Method

Cristian R. Munteanu[1,*], António Pimenta[2], Carlos Fernandez-Lozano[1], André Melo[3], Maria Cordeiro[3], Irina S. Moreira[2,*]

[1] Information and Communication Technologies Department, Computer Science Faculty, University of A Coruna, Campus de Elviña s/n, 15071, A Coruña, Spain; E-mail: crm.publish@gmail.com (CR.M.); carlos.fernandez@udc.es (C.FL.)

[2] CNC - Center for Neuroscience and Cell Biology; Rua Larga, FMUC, Polo I, 1ºandar, Universidade de Coimbra, 3004-517; Coimbra, Portugal; E-mail: caesar.m4d@gmail.com (A.P.): irina.moreira@cnc.uc.pt (I.S.M.)

[3] REQUIMTE/Departamento de Química e Bioquímica, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal; E-mail: asmelo@fc.up.pt (A.M.); ncordeir@fc.up.pt (M.C.)

*Author to whom correspondence should be addressed; E-Mail: irina.moreira@cnc.uc.pt or crm.publish@gmail.com.
Tel.: +351-239-820-190 (ext. 123); Fax: +351-239-822-776.

Published: 4 December 2015

Abstract: The natural tendency of proteins to bind to each other, as well as to many different molecules, forming stable and specific complexes is fundamental to all biological processes. The structural and functional description of protein-protein and protein-ligand complexes and their comprehension is a key concept, not only to increase the scientific knowledge in basic terms but also for the application to the biomedical and pharmaceutical industry. In this work we have look for more accurate ways of predicting the crucial residues for complex binding (Hot-spots) that can be used to model protein structure, dynamics and function. We developed an algorithm based in innovative series of descriptors, which have not been used in hot-spot determination and that can be applied to both protein-protein and protein-nucleic acid interfaces HS detection. A web-server for public use of the new methodological approaches was built and can be accessed at http://bio-aims.udc.es/MolStructPred.php

Keywords: Hot-spots; conservation; solvent accessible surface area; machine-learning, protein-protein interfaces; protein-nucleic acid interfaces.

1. Introduction

Protein-protein interactions (PPIs) are fundamental for all life processes and it is vital to understand their dynamics, structural and energetic characteristics in order to find new improved ways to influence these molecular machineries[1]. Traditional mutagenesis approaches, including the use of hybrid receptors and alanine scanning mutagenesis techniques, have led to important insights into the structural basis underlying PPIs. However, experimental mutagenesis scanning of a complete interface is highly costly from a financial and time point of view[1-3]. To overcome this problem it was needed an efficient and fast computational technique that allows the detection of the major binding determinants at a protein-protein interface: the Hot-Spots (HS). HS tend to be conserved residues tightly clustered in the central part of protein-protein interfaces forming a network of specific interactions that are optimized and cooperative[4]. Figure 1 illustrates an example of a protein-protein complex in which HS are highlighted in a vdW red representation and non-HS (called Null-Spots NS) in a yellow one. HS tend to be surrounded by a region of supposedly "less important" residues, largely hydrophobic, that leads to solvent occlusion and results in a lower local dielectric constant environment and enhancement of specific electrostatic and hydrogen bond interactions (Figure 1)[5]. So, according to this theory (O-ring theory), HS regions have a low number of interfacial waters, implying that water entropy effects provide one of the driving forces to complex formation[6] and that occlusion of bulk solvent slows down dissociation. Having these knowledge gathered through the years about HS[1-4,7], we decided to look for a method based on genomic conservation scores and 12 different Solvent Accessible Surface Areas features (described at reference[8]).

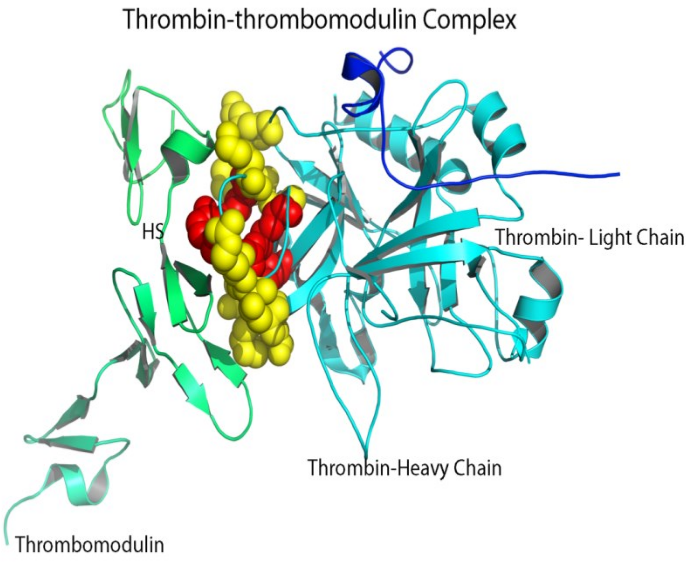


Figure 1. Structural representation of a protein-protein complex (PDBid: 1DX5[9]) in which the HS and NS are highlighted in a red and yellow vdW representation, respectively.

2. Results and Discussion

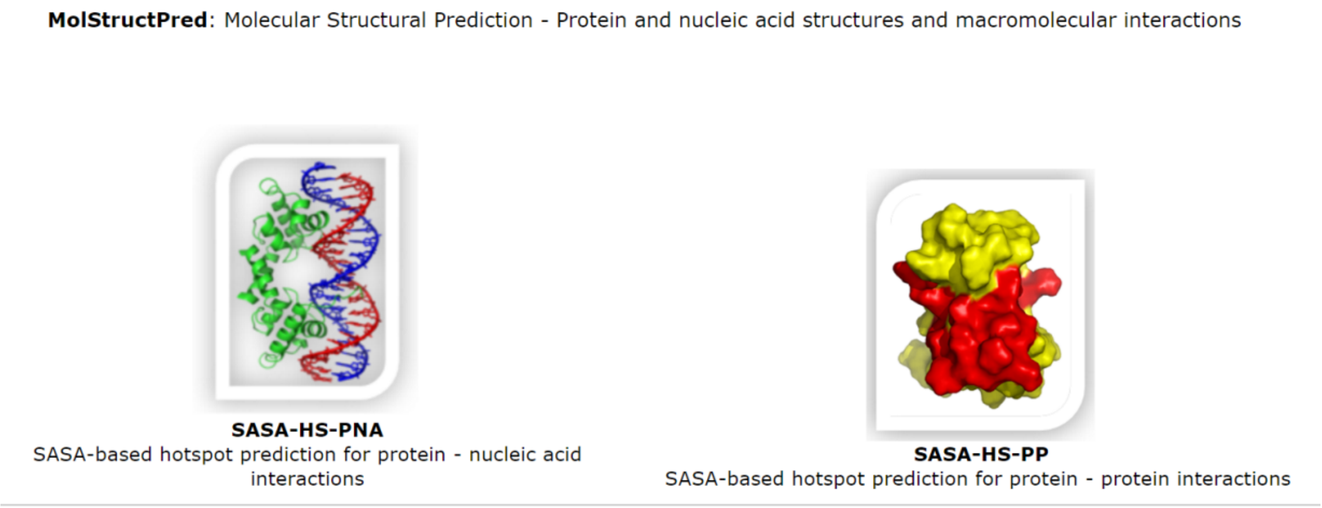
The performance in ML is usually measured using predictive accuracy, which could be problematic if the data is unbalanced[10]. Dataset S1 comprised 71 HS/406 NS, dataset S2 35 HS/56 NS, dataset S3 60 HS/162 NS and dataset S4 20 HS/80 NS, which demonstrates that our datasets (described at reference[8]) are highly unbalanced (classes are not equally represented as HS are less represented in Nature). This way, we evaluated the performance of each model by taking into account Recall (TPR), Precision, Specificity and FPR as well as F1-score and AUROC. We showed that simple Bayes Networks were able to classify HS for protein-protein interactions but only complex methods such as GA-SVM-Full could be used to classify HS for protein-nucleic acid interactions.

The best classifier for protein-protein case uses four features: CONSURF score, $\Delta SASA_i$, ${}_{rel/res}SASA_i$ and ${}_{rel/ave}SASA_{i.}$ (TPR=0.79, FPR=0.21, Precision=0.87, F1-score=0.83 and AUROC=0.85). Our algorithm was assessed against some of the state-of-the-art methods available by web-servers and proven to more accurately predict HS at protein-protein interfaces.

For protein-Nucleic Acid the best classifier uses two features: ConSurf score, ${}_{del/res}SASA_i$ (TPR=0.82, FPR=0.30, Precision=0.82, F1-score=0.85 and AUROC=0.83).

3. Materials and Methods

Three different datasets were used for the protein-protein interfaces: ASEdb,[11] BID[12] and SKEMPI[13] (comprising a total of 790 residues from 58 complexes) and one for protein-nucleic Acid: Pronit[14-16] (a total of 117 residues from 28 complexes). The datasets were constituted by protein complexes for which simultaneously exists experimental alanine scanning mutagenesis data, genetic conservation scores and tridimensional crystallographic structures of the bounded complex. These ones were filtered to ensure that a maximum of 35% sequence identity could be found for at least one protein in each interface[8]. Various machine-learning (ML) techniques were employed for this particular problem and in order to improve the performance and to reduce the number of features in the input space we also performed a Feature Selection (FS) approach as the number and relevance of the input variables can affect the performance of the model. Several statistics analyzes were performed to ensure the achievement of the high accuracy method.



MolStructPred: Molecular Structural Prediction - Protein and nucleic acid structures and macromolecular interactions

SASA-HS-PNA
SASA-based hotspot prediction for protein - nucleic acid interactions

SASA-HS-PP
SASA-based hotspot prediction for protein - protein interactions

Figure 2. Web-server for HS detection.

4. Conclusions

Our methods are accurate and time efficient. Moreover, our method can be applied not only to protein-protein but as well, and for the first time, to protein-nucleic acid complexes[8]. Web-servers were also constructed and made available for the scientific community at BioAIMS portal (http://bio-aims.udc.es/MolStructPred.php). The code of the Web tools is available as pySBHD repository (https://github.com/muntisa/pySBHD).

Acknowledgments

This work was funded by FEDER funds through the Operational Programme for Competitiveness Factors (COMPETE) and by National Funds through FCT (Foundation for Science and Technology) under the projects PEST-C/EQB/LA0006/2013 and COMPETE FCOMP-01-0124-FEDER-37285. This work was further supported by “Collaborative Project on Medical Informatics (CIMED)” PI13/00280 funded by the Carlos III Health Institute from the Spanish National plan for Scientific and Technical Research and Innovation 2013-2016 and the European Regional Development Funds (FEDER). The authors acknowledge the support by the Galician Network of Drugs R+D REGID (Xunta de Galicia R2014/025). ISM was supported by FCT Ciência 2008 and FCT Investigator programmes - IF/00578/2014 (co-financed by European Social Fund and Programa Operacional Potencial Humano).

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Moreira, I.S.; Fernandes, P.A.; Ramos, M.J. Hot spots—a review of the protein–protein interface determinant amino-acid residues. *Proteins: Structure, Function, and Bioinformatics* **2007**, *68*, 803-812.
2. Moreira, I.S.; Martins, J.M.; Ramos, R.M.; Fernandes, P.A.; Ramos, M.J. Understanding the importance of the aromatic amino-acid residues as hot-spots. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2013**, *1834*, 404-414.
3. Moreira, I.S.; Ramos, R.M.; Martins, J.M.; Fernandes, P.A.; Ramos, M.J. Are hot-spots occluded from water? *Journal of Biomolecular Structure and Dynamics* **2013**, *32*, 186-197.
4. Moreira, I.S. The role of water occlusion for the definition of a protein binding hot-spot *Curr Top Med Chem* **2015**, *15*, 2068-2079.
5. Bogan, A.A.; Thorn, K.S. Anatomy of hot spots in protein interfaces. *J Mol Biol* **1998**, *280*, 1-9.
6. Oshima, H.; Yasuda, S.; Yoshidome, T.; Ikeguchi, M.; Kinoshita, M. Crucial importance of the water-entropy effect in predicting hot spots in protein-protein complexes. *Physical Chemistry Chemical Physics* **2011**, *13*, 16236-16246.
7. Martins, J.M.; Ramos, R.M.; Pimenta, A.C.; Moreira, I.S. Solvent-accessible surface area: How well can be applied to hot-spot detection? *Proteins: Structure, Function, and Bioinformatics* **2014**, *82*, 479-490.
8. Munteanu, C.; Pimenta, A.C.; Fernandez-Lozano, C.; Melo, A.; Dias Soeiro Cordeiro, M.N.; Moreira, I.S. Sasa-based hot-spot detection 2 (sbhd2) methods for protein-protein and protein-nucleic acid interfaces. *Journal of Chemical Information and Modeling* **2015**.
9. Fuentes-Prior, P.; Iwanaga, Y.; Huber, R.; Pagila, R.; Rumennik, G.; Seto, M.; Morser, J.; Light, D.R.; Bode, W. Structural basis for the anticoagulant activity of the thrombin-thrombomodulin complex. *Nature* **2000**, *404*, 518-525.
10. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. Smote: Synthetic minority over-sampling technique. *J. Artif. Int. Res.* **2002**, *16*, 321-357.
11. Thorn, K.S.; Bogan, A.A. Asedb: A database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics* **2001**, *17*, 284-285.
12. Fischer, T.B.; Arunachalam, K.V.; Bailey, D.; Mangual, V.; Bakhru, S.; Russo, R.; Huang, D.; Paczkowski, M.; Lalchandani, V.; Ramachandra, C.*, et al.* The binding interface database (bid): A compilation of amino acid hot spots in protein interfaces. *Bioinformatics* **2003**, *19*, 1453-1454.
13. Moal, I.H.; Fernández-Recio, J. Skempi: A structural kinetic and energetic database of mutant protein interactions and its use in empirical models. *Bioinformatics* **2012**, *28*, 2600-2607.
14. Kumar, M.D.S.; Bava, K.A.; Gromiha, M.M.; Prabakaran, P.; Kitajima, K.; Uedaira, H.; Sarai, A. Protherm and pronit: Thermodynamic databases for proteins and protein–nucleic acid interactions. *Nucleic Acids Research* **2006**, *34*, D204-D206.
15. Prabakaran, P.; An, J.; Gromiha, M.M.; Selvaraj, S.; Uedaira, H.; Kono, H.; Sarai, A. Thermodynamic database for protein-nucleic acid interactions (pronit). *Bioinformatics* **2001**, *17*, 1027-1034.

16. Sarai, A.; Gromiha, M.M.; An, J.; Prabakaran, P.; Selvaraj, S.; Kono, H.; Oobatake, M.; Uedaira, H. Thermodynamic databases for proteins and protein-nucleic acid interactions. *Biopolymers* **2001**, *61*, 121-126.

SciForum
Mol2Net

Pharmacological Characterization of the Prostanoid Receptor EP3 in Locus Coeruleus Neurons by Single-Unit Extracellular Recordings in the Rat Brain in Vitro

Amaia Nazabal[1], Aitziber Mendiguren[1] and Joseba Pineda[1]*

[1] Department of Pharmacology, Faculty of Medicine and Odontology, University of the Basque Country (UPV/EHU), E-48940 Leioa, Bizkaia, Spain

* Author to whom correspondence should be addressed; e-mail: joseba.pineda@ehu.es

Published: 4 December 2015

Abstract: Prostanoids are known to regulate several physiological functions and to play an important role in certain pathophysiological situations such as inflammation. Prostaglandin E_2 receptors (EP) are members of the G protein-coupled receptor superfamily. Four subtypes have been described: EP2 and EP4 (coupled to G_s proteins) and EP1 and EP3 (coupled to $G_{i/o}$ proteins). To date, the function of the prostanoid system in the brain has not been well characterized. The locus coeruleus (LC), the main noradrenergic nucleus in the brain, has been described to express the EP3 receptor. The aim of this study was to characterize the functional relevance of EP3 receptors in the LC by single-unit extracellular recordings in rat brain slices. We performed concentration-effect curves for different endogenous derivatives and selective agonists of EP3 receptors. Thus, increasing concentrations of the EP3/EP1 agonist sulprostone (0.3-80 nM) fully inhibited the neuronal activity of LC cells, with an EC_{50} value of 15 nM (n = 9). The EP3 receptor antagonist L-798,106 (10 μM) caused a rightward shift (> 8 fold) in the concentration-effect curve for sulprostone, but the EP2 receptor antagonist PF04418948 (10 μM) or the EP4 receptor antagonist L-161,982 (10 μM) failed to cause any rightward shift of sulprostone effect. On the other hand, perfusion with the endogenous PGE_2 (0.3 nM-1.28 μM) or the PGE_1 analogue misoprostol (0.3-320 nM) induced a concentration-dependent inhibition of the firing rate of LC cells, with EC50 values being 51 nM and 112 nM, respectively. Likewise, only the EP3 antagonist L-798,106 (10 μM) caused a rightward shift (> 8 fold) in the concentration-effect curves for these prostanoid agonists (n = 5). In conclusion, LC neurons are regulated in an inhibitory manner by the prostanoid system likely through the EP3 receptor.

Keywords: in vitro; locus coeruleus; electrophysiology; EP3 receptor; prostanoid; PGE2; sulprostone; slice

1. Introduction

Prostanoids system is involved in the regulation of pain, fever and inflammation processes. During inflammation, membrane phospholipids are transformed into arachidonic acid (AA) by the phospholipase A_2 enzyme. In turn, AA is metabolized by cyclooxygenase (COX) into PGH_2, which is the common prostanoid precursor of other prostaglandins including prostacyclins and thromboxanes (Yagami et al. 2015). COX-blockers or NSAIDs -nonsteroidal anti-inflammatory drug- have been widely used as analgesic, antipyretic and antiinflammatory drugs due to their ability to suppress the production of prostanoids.

Two isoforms of COX (-1 and -2) have been characterized, but the isoform that has been shown to be constitutively expressed in the body is the isoform 1. However, recent studies have suggested that COX-2 could also be constitutively expressed in human and animal brain (Hétu & Riendeau 2005; Martin et al. 2007; Yaksh et al. 2001). This supports the idea of a possible important role of prostaglandins in the central nervous system. Nevertheless, the function of prostaglandins remains unclear since dual effects have been described in the brain. On one hand, PGE_2 seems to promote neuroprotection in injured brain by induction of either BDNF release or reducing the expression of inducible oxide nitric synthase (iNOS) during inflammation (Hutchinson et al. 2009; Levi et al. 1998). On the other hand, EP3 antagonists exert protective effects in the brain after ischemia (Ikeda-Matsuo et al. 2011).

PGE_2 is the most abundant prostaglandin produced in the body and it acts through activation of EP1-4 receptors, which are expressed in neurons and glia (Ito et al. 2001). Classically, EP1 has been described to be coupled to $G_{i/o}$ protein (Ji et al. 2010), although there is recent evidence that supports its coupling to G_q protein (Liu et al. 2010). EP3 is coupled to $G_{i/o}$ protein (Negishi et al. 1995), whereas EP2 and EP4 are coupled to G_s protein (Fujino & Regan 2006; Sugimoto & Narumiya 2007).

The locus coeruleus (LC) is the main source of noradrenergic innervation in the brain. It is involved in the regulation of numerous physiological functions such as sleep-wake cycle, arousal, cognition/memory, pain, cardiovascular control and rewarding behavior. A role of this nucleus in the production of fever has also been proposed (Almeida et al. 2004).

Several findings have suggested an interaction between the noradrenergic/LC system and prostanoid system. First, *in situ* hybridization experiments have demonstrated the presence of mRNA for the EP2 and EP4 receptors in this nucleus (Zhang & Rivest 1999). Furthermore, EP3 expression has also been shown by inmunoreactivity and double hybridization techniques (Ek et al. 2000; Nakagawa et al. 2000; Nakamura et al. 2001). Second, the LC has been identified as a pivotal nucleus in PGE2-induced thermogenesis (Almeida et al. 2004). Third, administration of PGE2 inhibits, via EP3 receptors the release of noradrenaline in the brain (Exner & Schlicker 1995).

Despite several evidences that suggest a possible role of the prostanoid system in the LC, the functional role of EP receptors in this nucleus remained to be studied. Therefore, the aim of our research was to characterize, by single-unit extracellular recordings *in vitro*, the functional role of the prostanoid receptor EP3 in LC neurons from the rat brain. For this purpose, we performed concentration-effect curves for several EP3 receptor agonists in the absence and in the presence of different EP receptor antagonists.

2. Results and Discussion

Effect of the EP3/EP1 receptor agonist sulprostone on the firing rate of LC neurons

To evaluate the effect of prostaglandins on the LC neurons, we applied increasing concentrations of the EP3/EP1 agonist sulprostone (0.3-80 nM) and we found that it fully inhibited the neuronal activity of LC cells, with an EC_{50} value of 15 nM (n = 9). To identify the EP receptor involved in the inhibitory effect induced by sulprostone, we performed the concentration-effect curves for the EP3/EP1 agonist in the presence of specific antagonists of the EP receptors expressed in the LC: EP2 (PF04418948), EP3 (L-798,106) and EP4 (L-161,982) at 3 and 10 µM. The EP3 receptor antagonist L-798,106 (10 µM) caused a rightward shift (> 8 fold) in the concentration effect curve for sulprostone (Fig. 1). Neither the EP2 receptor antagonist PF04418948 (10 µM) nor the EP4 receptor antagonist L-161,982 (10 µM) caused any rightward shift of sulprostone effect.

These results indicate that the inhibitory effect of sulprostone is mediated by the EP3 receptor, which is known to be coupled to $G_{i/o}$ proteins. The inhibitory effect mediated by the EP3 receptor has been shown to occur in other nucleus at the same concentrations used in our study (Ito et al. 2000). Therefore, our results show that EP3 receptor is functionally active in the LC. Future experiments are required to characterize the relevance of other EP receptors to the modulation of the firing activity of LC cells.

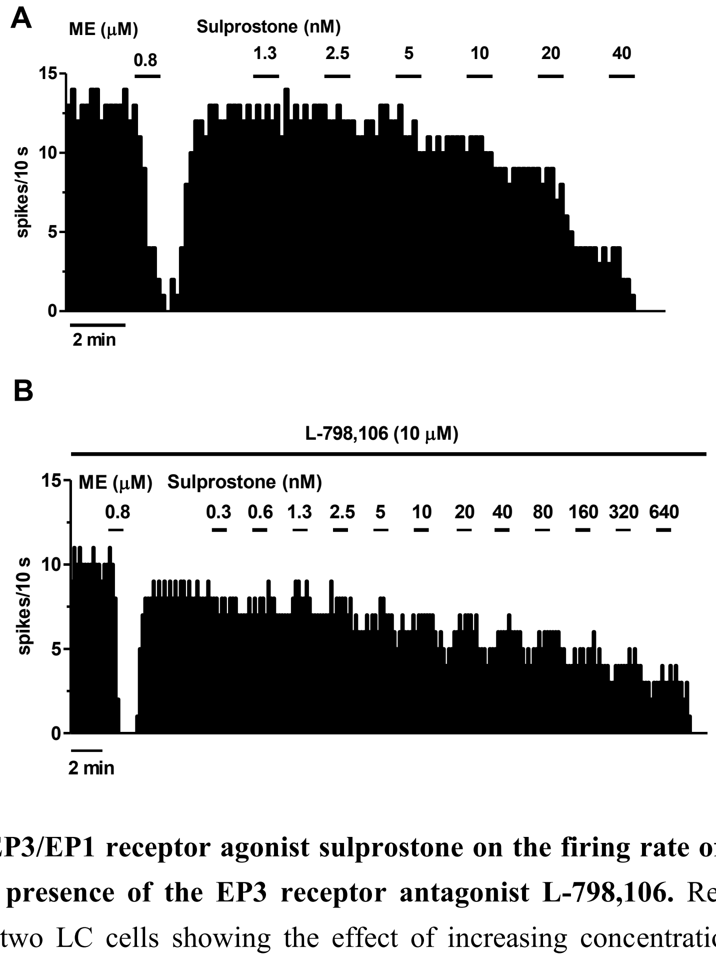


Figure 1. Effect of the EP3/EP1 receptor agonist sulprostone on the firing rate of locus coeruleus cells in the absence and in the presence of the EP3 receptor antagonist L-798,106. Representative examples of firing-rate recordings of two LC cells showing the effect of increasing concentrations of sulprostone in the

absence (**A**) and in the presence (**B**) of the EP3 antagonist L-798,106. Each horizontal bar represents the period of application of each sulprostone concentration and the vertical lines show the number of spikes recorded every 10 s. The inhibitory effect induced by each application was calculated as a percentage from the basal firing rate. Note that sulprostone inhibits the firing rate of LC cells and that its inhibitory effect is diminished in the presence of L-798,106.

Effect of the endogenous PGE_2 and the PGE_1 analogue misoprostol on the firing rate of LC neurons

In order to study whether endogenous prostanoid compounds mimicked the effect observed with sulprostone on the LC, we applied the endogenous PGE_2 and the PGE_1 analogue misoprostol. Perfusion with the endogenous PGE_2 (0.3 nM-1.28 µM) or the PGE_1 analogue misoprostol (0.3-320 nM) induced a concentration-dependent inhibition of the firing rate of LC cells, with EC_{50} values being 51 nM and 112 nM respectively. Likewise, only the EP3 antagonist L-798,106 (10 µM) caused a rightward shift (> 8 fold) in the concentration-effect curves for these prostanoid agonists (Fig. 2; n = 5).

These results indicate that sulprostone and the endogenous derivatives act through activation of the same EP receptor; the EP3 receptor. However, sulprostone shows higher potency than the endogenous prostanoid derivatives to inhibit LC cells.

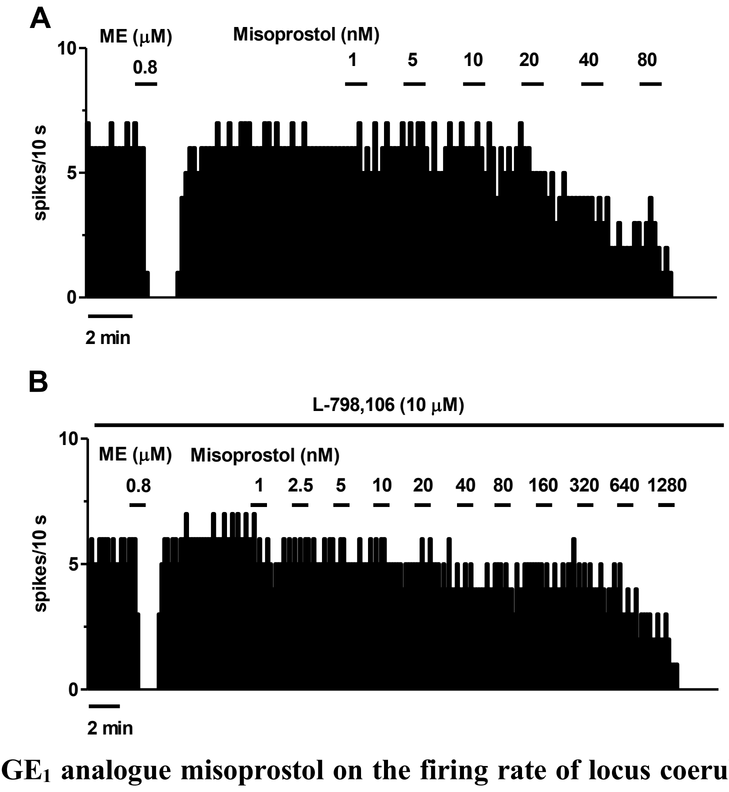


Figure 2. Effect of the PGE_1 analogue misoprostol on the firing rate of locus coeruleus cells in the absence and in the presence the EP3 receptor antagonist L-798,106. Representative examples of firing-rate recordings of two LC cells showing the effect of increasing concentrations of misoprostol in the absence (**A**) and in the presence (**B**) of the EP3 antagonist L-798,106. Each horizontal bar represents the period of application of each misoprostol concentration and the vertical lines show the number of spikes recorded every 10 s. The inhibitory effect induced by each application was calculated as a percentage from the basal firing rate. Note that misoprostol inhibits the firing rate of LC cells and that its inhibitory effect is diminished in the presence of L-798,106.

3. Materials and Methods

3.1. Animals

Adult male Sprague-Dawley rats weighing 200-300 g were housed under controlled laboratory conditions (22 ºC and 12-h light/dark cycles) with free access to food and water. The animals were obtained from the animal house of the University of the Basque Country (Leioa, Spain). All experimental procedures reported in this manuscript were conducted in accordance with U.K. Animals (Scientific Procedures) Act, 1986, and associated guidelines, and with the European Community Council Directive on "Protection of Animals Used in Experimental and Other Scientific Purposes" of 24 November 1986 (86/609/EEC). The procedures were approved by the Animal Care and Use Committee of the University of the Basque Country. All efforts were made to minimize animal suffering and to reduce the number of animals used.

3.2. In vitro electrophysiology

3.2.1. Brain slice preparation

In vitro experiments were performed as previously described (Mendiguren & Pineda 2007). Briefly, animals were first anaesthetized with chloral hydrate (400 mg/kg, i.p.) and sacrificed by decapitation. The brain was removed and a block of tissue containing the brainstem was placed in ice-cold modified artificial cerebrospinal fluid (aCSF) where NaCl was equiosmotically substituted with sucrose to improve neuronal viability. Coronal slices of 600 µm thickness containing the LC were cut by an oscillating vibratome and then allowed to recover from the slicing for 90 min in oxygenated aCSF. Next, slices were placed on a nylon mesh and maintained at 33 ± 0.5 C in a modified Haas-type interface chamber continuously perfused with oxygenated aCSF (95% O_2/5% CO_2, pH=7.38) at a flow rate of 1-1.5 ml/min. The aCSF contained (in mM): NaCl 130, KCl 3, NaH_2PO_4 1.25, D-glucose 10, $NaHCO_3$ 21, $CaCl_2$ 2, and $MgSO_4$ 2.

3.2.2. Recording procedures

Extracellular recordings of single neurons were performed as previously described (Mendiguren & Pineda 2004). The recording electrode consisted of an Omegadot glass micropipette that was pulled and filled with a solution of 50 mM NaCl (tip size of 2-5 µm, 3-5 MΩ). The microelectrode was placed in the LC, which was visually identified in the rostral pons as a dark oval area on the lateral borders of the central gray and the 4th ventricle, just anterior to the genu of the facial nerve. The extracellular signal recorded by the microelectrode was passed through a high-input impedance amplifier system and monitored on an oscilloscope and by an audioanalyzer. Individual (single-unit) neuronal spikes were isolated from the background noise with a window discriminator and counted. The firing rate was represented and analyzed by a PC-based custom-made program, which generated histogram bars representing the cumulative number of spikes in consecutive 10 s bins. Noradrenergic neurons in the LC were identified by the following electrophysiological criteria: a spontaneous and regular discharge, a slow firing rate and a positive-negative biphasic waveform of 3-4 ms duration (Andrade & Aghajanian 1984). We only recorded cells that showed stable firing rates between 0.4 and 1.5 Hz for at least 3-5 min and strong inhibitory effects induced by ME (0.8 mM, 1 min) (higher than 80%). Only one neuron was recorded per slice and only one slice was obtained from each animal.

3.3. Pharmacological procedures

To characterize the effect of prostanoids in the LC neurons, we perfused increasing concentrations of the EP3/EP1 receptor agonist sulprostone (0.3-80 nM) and the endogenous derivatives of prostanoid system (PGE2 and misoprostol, 0.3 nM–1.28 µM). Each concentration of the EP receptor agonists was perfused for at least 1 min. We performed concentration/effect curves for the inhibitory effect of the agonists. In order to study the involvement of EP receptors in the effect observed with the EP receptor agonists, we used specific antagonist for the EP2 receptor (PF04418948), EP3 (L-798,106), and EP4 (L-161,982) at 3 and 10 µM. All the antagonists were perfused for at least 20-30 min before performing the concentration-effect curves for EP receptor agonists.

3.4. Drugs and reagents

Chloral hydrate was obtained from Sigma-Aldrich Química S.A. (Madrid, Spain). Met^5-enkephalin (ME) was obtained from Bachem (Weil am Rhein, Germany). Sulprostone was purchased from Cayman Chemical (Michigan, USA). PF04418948, L-798,106 and L-161,982 were obtained from Tocris Bioscience (Bristol, UK). Drugs were dissolved in the final volume of aCSF just before each assay and applied by turning a threeway valve that switched from aCSF to the test solution. Stock solutions of the EP antagonist were made in DMSO stored at - 25 ºC and, on the day of the experiment, diluted in aCSF to their final volume. The maximal final concentration of DMSO was lower than 0.1%.

3.5. Data analysis and statistics

Values are expressed as the mean ± standard error of the mean (S.E.M) of n experiments. The firing rate of LC cells was recorded before (baseline), during and after drug applications. The inhibitory effect of EP receptor agonists was calculated as follows:

$$E(\%) = \frac{FR_{pre} - FR_{post}}{FR_{basal}} \cdot 100$$

where FR$_{pre}$ is the average firing rate for 60 s before application of each concentration, FR$_{post}$ is the average firing rate after the perfusion of each concentration, and FR$_{basal}$ is the firing rate for 60 s of each cell at the beginning of the recording. Changes induced by drugs in the firing rate were evaluated by a paired Student's t test when compared within the same cell or by a two-sample Student's t test when compared between different cells. Additionally, to evaluate the possible differences between groups the one-way ANOVA test was used. The level of significance was considered as p=0.05. Curve fitting analysis was performed by the computer program GraphPad Prism (version 5.0 for Windows, San Diego, CA, USA) to obtain the best simple nonlinear fit to the following three-parameter logistic equation:

$$E = Emax / [1 + (EC_{50} / A)^{n}]$$

where E is the effect induced by each concentration of the EP3 agonist (A), Emax is the maximal effect, EC$_{50}$ is the concentration of EP3 agonist needed to elicit a 50% of the maximal effect, and *n* is the slope factor of the concentration-effect curve

4. Conclusions

In conclusion, LC neurons are regulated in an inhibitory manner by the prostanoid system likely through the EP3 receptor. Future experiments are required to characterize the relevance of other EP receptors to the modulation of the firing activity of LC cells.

Acknowledgments

This work was supported by the Ministerio de Ciencia e Innovación [Grant SAF2008-03612] and the University of the Basque Country (UPV/EHU) [Grant GIU11/27 and Grant GIU14/29]. Pineda´s research group takes part in a network unit supported by the University of the Basque Country [UFI 11/35]. A. Nazabal was supported by a predoctoral fellowship from the Basque Government. The experiments comply with the current laws of Spain.

Author Contributions

A. Nazabal performed the research and analysed the data. A. Mendiguren and J. Pineda designed the research study. All mentioned authors wrote the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

Almeida, M.C. et al., 2004. Thermoeffector neuronal pathways in fever: a study in rats showing a new role of the locus coeruleus. *The Journal of physiology*, 558(Pt 1), pp.283–94.

Andrade, R. & Aghajanian, G.K., 1984. Locus coeruleus activity in vitro: intrinsic regulation by a calcium-dependent potassium conductance but not alpha 2-adrenoceptors. *The Journal of neuroscience*, 4, pp.161–170.

Ek, M. et al., 2000. Distribution of the EP3 prostaglandin E2 receptor subtype in the rat brain: Relationship to sites of interleukin-1 - Induced cellular responsiveness. *Journal of Comparative Neurology*, 428(August), pp.5–19.

Exner, H.J. & Schlicker, E., 1995. Prostanoid receptors of the EP3 subtype mediate the inhibitory effect of prostaglandin E2 on noradrenaline release in the mouse brain cortex. *Naunyn-Schmiedeberg's archives of pharmacology*, 351(1), pp.46–52.

Fujino, H. & Regan, J.W., 2006. EP 4 Prostanoid Receptor Coupling to a Pertussis Toxin- Sensitive Inhibitory G Protein. *Molecular Pharmacology*, 69(1), pp.5–10.

Hétu, P.-O. & Riendeau, D., 2005. Cyclo-oxygenase-2 contributes to constitutive prostanoid production in rat kidney and brain. *The Biochemical journal*, 391, pp.561–566.

Hutchinson, A.J., Chou, C., Israel, D.D., Xu, W., Regan, J.W., 2009. Activation of EP2 prostanoid receptors in human glial cell lines stimulates the secretion of BDNF. *Neurochem int*, 54(7), pp.439–446.

Ikeda-Matsuo, Y. et al., 2011. Inhibition of prostaglandin E2 EP3 receptors improves stroke injury via anti-inflammatory and anti-apoptotic mechanisms. *Journal of Neuroimmunology*, 238(1-2), pp.34–43.

Ito, S., Okuda-Ashitaka, E. & Minami, T., 2001. Central and peripheral roles of prostaglandins in pain and their interactions with novel neuropeptides nociceptin and nocistatin. *Neuroscience research*, 41, pp.299–332.

Ito, Y. et al., 2000. The prostaglandin E series modulates high-voltage-activated calcium channels probably through the EP3 receptor in rat paratracheal ganglia. *Neuropharmacology*, 39, pp.181–190.

Ji, R. et al., 2010. EP1 Prostanoid Receptor Coupling to G i / o Up-Regulates the Expression of Hypoxia-Inducible Factor-1 ⎵ through Activation of a Phosphoinositide-3 Kinase Signaling Pathway. *Molecular Pharmacology*, 77(6), pp.1025–1036.

Levi, G., Minghetti, L. & Aloisi, F., 1998. Regulation of prostanoid synthesis in microglial cells and effects of prostaglandin E 2 on microglial functions. *Biochimie*, pp.899–904.

Liu, J.-F. et al., 2010. Cyclooxygenase-2 enhances alpha2beta1 integrin expression and cell migration via EP1 dependent signaling pathway in human chondrosarcoma cells. *Molecular cancer*, 9, p.43.

Martin, F. et al., 2007. Constitutive cyclooxygenase-2 is involved in central nociceptive processes in humans. *Anesthesiology*, 106(5), pp.1013–8.

Mendiguren, A. & Pineda, J., 2004. Cannabinoids enhance N-methyl-D-aspartate-induced excitation of locus coeruleus neurons by CB1 receptors in rat brain slices. *Neuroscience letters*, 363, pp.1–5.

Mendiguren, A. & Pineda, J., 2007. CB1 cannabinoid receptors inhibit the glutamatergic component of KCl-evoked excitation of locus coeruleus neurons in rat brain slices. *Neuropharmacology*, 52, pp.617–625.

Nakagawa, T. et al., 2000. Possible involvement of the locus coeruleus in inhibition by prostanoid EP(3) receptor-selective agonists of morphine withdrawal syndrome in rats. *European journal of pharmacology*, 390(3), pp.257–66.

Nakamura, K. et al., 2001. Prostaglandin EP3 receptor protein in serotonin and catecholamine cell groups: a double immunofluorescence study in the rat brain. *Neuroscience*, 103(3), pp.763–75.

Negishi, M. et al., 1995. Selective coupling of prostaglandin E receptor EP3D to Gi and Gs through interaction of α-carboxylic acid of agonist and arginine residue of seventh transmembrane domain. *Journal of Biological Chemistry*, 270, pp.16122–16127.

Sugimoto, Y. & Narumiya, S., 2007. Prostaglandin E receptors. *Journal of Biological Chemistry*, 282(16), pp.11613–11617.

Yagami, T., Koma, H. & Yamamoto, Y., 2015. Pathophysiological Roles of Cyclooxygenases and Prostaglandins in the Central Nervous System. *Mol Neurobiol*, DOI 10.100.

Yaksh, T.L. et al., 2001. The acute antihyperalgesic action of nonsteroidal, anti-inflammatory drugs and release of spinal prostaglandin E2 is mediated by the inhibition of constitutive spinal cyclooxygenase-2 (COX-2) but not COX-1. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 21(16), pp.5847–5853.

Zhang, J. & Rivest, S., 1999. Distribution , regulation and colocalization of the genes encoding the EP 2 - and EP 4 -PGE 2 receptors in the rat brain and neuronal responses to systemic inflammation. *European Journal of Neuroscience*, 11, pp.2651–2668.

SciForum
Mol2Net

An Alternative Approach to Structure Specification Based on Fuzzy Multidimensional Membership Function Using Forward Selection Rule

Sreyasi Ghosh, Assistant Professor, Department of Commerce (Mathematics), The Bhawanipur Education Society College ghosh.sreyasi1@gmail.com,+919433785429

Sarbari Ghosh, Department of Mathematics, Vidyasagar Evening College and Guest Faculty Department of Atmospheric Science, C.U..

Pradip Kumar Sen, Department of Mathematics, J.U.

Subhra Chatterjee, Assistant Professor, Academy of Technology, Kolkata.

Published: 4 December 2015

Abstract: Fuzzy logic first established in July,1964 by Lofti A. Zadeh, is usually used to develop cost-effective approximate solutions to complex real-world problems exploiting the tolerance of imprecision. The present study attempts to develop a general computational technique based on fuzzy multi-dimensional membership function using forward selection rule for discriminating two different situations which is basically non-linear. Incidentally the technique suggested here is validated with atmospheric data. Earlier the fundamental principal component analysis (PCA) technique was applied to identify the significant parameters for the occurrence of pre-monsoon thunderstorms (TS) in Kolkata. They showed how the linear discriminant analysis (LDA) technique alone as well as in conjunction with PCA can be successfully applied for the purpose (Ghosh et al. 1999, 2004; Chatterjee et al., 2009). Also a comparative study was performed between the existing multivariate technique, the linear discriminant analysis and a technique based on fuzzy membership roster method (Chatterjee et al., 2011). Recently a fuzzy –neuro based algorithm for weather prediction has been developed (T. Rahman et al. 2014). The main objective of the study is to address the numerical imprecision of some quantified physical variables. In this rule, a product form is taken to construct the multivariate membership function where the univariate membership function is Gaussian in nature as well as continuously differentiable. Since the parameters may have different units so they are made dimensionless before taking the product. To develop the technique no software package or fuzzy toolbox is used. The program for the study is developed by the authors themselves. This rule is applied to two datasets of different categories consisting of the parameters of the days with convective development and fair weather respectively during pre-monsoon season of Kolkata (22.53ºN, 88.33ºE), India. Basic parameters for discriminating the situation (convective development and fair weather in pre-monsoon season of Kolkata) is constructed from the known data set of 12 years covering the period 1985-1996. The results are validated for the period 1997-1999 using the dataset consisting of variables of unknown nature. The study reveals that the technique can classify the two different situation to give the best possible combination of parameters with atmost 88% success rate. Moreover, the study indicates that the two datasets are structurally different. The technique suggested here is expected to work in any other domain too. It is found that the method works with better accuracy than the existing ones so far the atmospheric parameters are concerned.

INTRODUCTION

Pre monsoon thunder squalls/ Nor'westers are one of the most important events that occupy a major portion in the pre-monsoon weather system over Eastern India. Contribution of different meteorological parameter like temperature, pressure, humidity, etc are the most important variables which play a significant role for the development of the Pre monsoon thunder squalls/ Nor'weasters. Although other factors can also be taken in to consideration for the ideal situation and actual time and place of occurrence of all these thunder squalls during the hot weather period of Summer season(mainly within the optimum period of March to May) , every year, in most cases, over the region of Eastern and North-Eastern states. Normally these Pre monsoon Thunder squalls/ Nor'weasters are very much violent and destructive in nature and appear suddenly in form of dark big clouds with sudden rise in wind speed associated with frequent lightning and thunder. Meteorological Scientist on previous occasions have contributed their valuable ideas and thoughts involving the phenomena by their contributions at different times. However, for forecasting the occurrence of Pre monsoon thunder squalls/ Nor'weaster, some times it is necessary to know the favorable conditions and mechanism of these phenomena.

Expert forecasters use well-developed subjective techniques for weather prediction. They improve their accuracy and skill over time by learning through experiences. Over the years forecasters have a huge collection of dataset and products. So, they can use, intelligent system approaches for data analysis, interpretation, verification etc. Fuzzy logic is one of such intelligent or expert systems, the goal of which is to perform at the level of a human expert by leveraging knowledge and experience gained over time. Ghosh *et al.* have applied Fuzzy multivariate membership function using forward selection rule to the data set of three years to identify significant parameters for the occurrence of pre-monsoon thunderstorms (TS) at Kolkata (22.53ºN, 88.33ºE). The present work aims at the reduction of parameters using some objective method as well as to predict the convective development during pre-monsoon season at Kolkata utilizing radiosonde data of 15 years. In fact, many researchers in different situations have used these multivariate techniques. The principal component analysis (PCA) technique was applied by previous workers to identify the significant parameters for the occurrence of pre-monsoon thunderstorms (TS) in Kolkata. They showed how the linear discriminant analysis

(LDA) technique alone as well as in conjunction with PCA can be successfully applied for the purpose (Ghosh et al. 1999, 2004; Chatterjee et al., 2009)[1] . Also a comparative study was performed between the existing multivariate technique, the linear discriminant analysis and a technique based on fuzzy membership roster method (Chatterjee et al., 2011)[2]. Eigenvector methods have been applied to study the principal anomaly patterns of winter temperature[3]. The principal components derived from a 500 hPa height data set had been linearly transformed to interpret spatial patterns[4-5]. The principal components based on covariance matrix and correlation matrix for a given data set of cyclone frequencies have been compared[6]. Cluster analysis and linear discriminant analysis (LDA) have been utilized to describe a multivariate statistical model for forecasting anomalies of surface pressure present over Europe and North Atlantic[7]. A comparative study of rotated and unrotated PCA has been performed[8]. A composite empirical orthogonal function (EOF) analysis of the monthly sea surface temperature variations and those of precipitation in the tropical Pacific Ocean region was performed[9]. Multiple linear regressions was compared with LDA for making hind casts and real time forecasts of north-east Brazil wet season rainfall using sea surface temperature10. Though a number of attempts were made to establish empirical models for the prediction of atmospheric stability11-12, the work done on Kano is perhaps the first successful attempt for tropical region13. Some statistical forecast models, based on logistic regression, have been reported in the literature. While, in one case, six variables were selected from 27 variables by constructing correlation matrix14, in the other case, variable reduction was done using forward and backward selection procedures15. In India, a number of attempts have been made to describe occurrence of rainfall by two states Markov-chain16-19. Another attempt has been made to predict the occurrence of CD at Dhaka (Bangladesh) in terms of stability indices20. Complex empirical orthogonal function was used to determine vertical wind profiles over the Indian Ocean21. Another study has produced a computational algorithm for one-dimensional cyclostationary empirical orthogonal functions and examined their properties22. A low order barotropic EOF model has also been reformulated23.

Convective developments are strongly favoured by convective instability, abundant moisture at lower levels, strong wind shear, and a dynamical lifting mechanism that can release the instability24. Not only that, the vertical shear of the environmental winds has to match the value of the convective instability for proper development of a large convective cloud25. It has been emphasized that the presence of conditional instability is an essential criterion for supporting electrification and lightning26. Apart from the parameters mentioned, two more parameters, viz. ($^{\theta}$es -$^{\theta}$e) and (P-PLCL) have been included in the present study, where, $^{\theta}$es and $^{\theta}$e, denote the saturated equivalent potential temperature and equivalent potential temperature, respectively; P, is a level pressure; and PLCL, is the pressure at the corresponding lifting condensation level.

The thermodynamic parameter ($^{\theta}$es -$^{\theta}$e) was originally introduced by Betts as a measure of the unsaturation of the atmosphere27. PLCL for the surface parcel was considered as the cloud base and hence, (P-PLCL) has been taken as a forcing factor for the saturation of a parcel28.

2 Data

The number of convective development) and fair weather (FW) days linked with the morning and evening radio sonde / rawin sonde (RS/RW) observations are presented in Tables 1 and 2. These data are used for the construction of discriminant functions. Any convective development occurring within the next 12 hrs of the morning RS/RW observation taken at 0000 hrs GMT (0530 hrs IST) is considered as CD related with morning RS/RW, otherwise it is FW related with the same RS/RW. A similar consideration for evening RS/RW observation taken at 1200 hrs GMT is utilized for the classification of CD or FW linked with evening

RS/RW. On many occasions, the data either at one or more of the significant levels, i.e. 1000, 850, 700, 600 and 500 hPa were not available. Naturally, those occasions could not be taken into consideration.

These limitations have greatly reduced the data size. The linear discriminant functions (morning and evening) for forecasting the convective development at Kolkata have been constructed utilizing all the available radiosonde data of 12 years (1985-1996) and for the validation of these functions, the radiosonde data of 3 years (1997-1999) have been used.

3 METHODOLOGY

The study is performed separately for morning and evening, as two radio soundings are available in a day. The atmospheric layer (1000-500)h Pa is subdivided into the following four layers : - (1000-850)h Pa, (850-700)h Pa, (700-600)h Pa and (600-500)h Pa. A day associated with a sub layer is considered as a vector containing at most five components, among which the first two components represent two thermodynamic parameters and the remaining three are the dynamic parameters. The category or pattern of an unknown day is predicted for next 12 hours from the time of observations depending on its degree of compatibility with the sets of days of two known categories or standard patterns (i.e. convective development and fair-weather). The study is performed separately for morning and evening. The sets of days of known categories are termed as fuzzy sets as the two sets have overlapping area so far the quantified values of the dynamic parameters like conditional instability, convective instability and vertical shear are concerned.

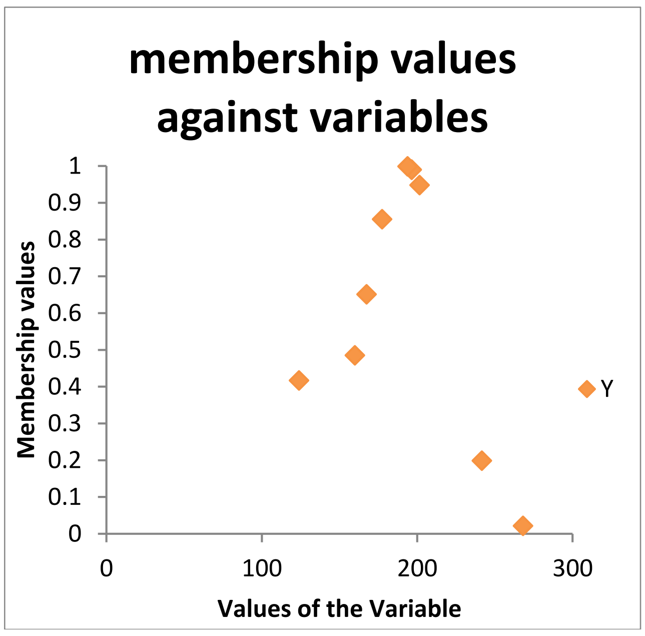
Forward selection

The simplest data-driven model building approach is called forward selection. In this approach, one adds variables to the model one at a time. At each step, each variable that is not already in the model is tested for inclusion in the model. The most significant of these variables is added to the model, so long as it's P-value is below some pre-set level. Thus we begin with a model including the variable that is most significant in the initial analysis, and continue adding variables until none of remaining variables are "significant" when added to the model. We have only verified the result for the layer (1000-850)h Pa. There are five parameters.

A product form is taken to construct the multivariate membership function. Since the parameters have different units so they are made dimensionless before taking the product. To develop the technique no software package or fuzzy toolbox is used. The program for the study is developed by the authors themselves. Since some of the parameters are found to follow Gaussian distribution and usually the physical parameters are assumed to be Gaussian or quasi Gaussian in nature, for each parameter, the

Gaussian membership function has been chosen to construct the one dimensional or univariate degree of compatibility. Gaussian membership functions are continuously differentiable as well as parameterizable. Gaussian membership functions are factorizable. Hence, we may synthesize a multi dimensional or multivariate degree of compatibility as the product of one dimensional or univariate degree of compatibility.

The graph of the membership grade values against the variables also suggests Gaussian nature of the membership function. The graph given below is an example of the relation between the values of the variables and their corresponding membership grade values.



Let us consider two groups X and Y, where X consists of the parameters of FW situations and the elements of Y are the parameters representing the situations of CD.Let us suppose that there are k parameters, Ui (i = 1 to k) on which we have the following two sets of observations:

X = [Xij] (i = 1 to k, j = 1 to m) and

Y = [Yij] (i = 1 to k, j = 1 to n).

In the present study, Ui (i = 1 to 5) denote the above mentioned 5 parameters, Xij denotes the value of the ith parameter on jth FW day and Yij gives the value of the ith parameter on jth CD day.

The work has been performed with k = 5, m = the number of FW days, which is 280 for morning and 201 for evening and n = the number of CD days, which is 123 for morning and 165 for evening. X and Y are termed as fuzzy sets since it is difficult to identify sharp boundaries between these two sets so far the parameters, viz., convective instability, conditional instability and vertical shear are concerned. Then the degrees of compatibility of a parameter, Ui (i=1 to 5) with the standard pattern classes, Y and X are computed by AY(Ui) and AX(Ui), again AY(U) and AX(U) are the product of the uni variates respectively. If, now, an unknown pattern or a day, say U = (U1,U2, …, U5) is given, where Ui is the measurement associated with the ith parameter of the pattern, then the degrees of compatibility of U with the standard patterns, Y and X, denoted by AY(U) and AX(U) respectively, are computed. Next, an unknown pattern or a day, U is classified by the larger value of AY(U) or AX(U), i.e. if AY(U) > AX(U), then there is a possibility for U to be more of the pattern Y than of the pattern X for next 12 hours. Hence it may be predicted that U is possibly a day with convective development for next 12 hours (Klir and Yuan, 2002)[27]. Here, the number of days of unknown category involved in the validation is 44 for MCD, 84 for MFW, 53 for ECD and 65 for EFW. It is worth mentioning in this context that there is no sound principle yet for guiding the choice of membership function or degree of compatibility.

RESULT

No. of Combinations	Nature of Day	No. of Days	Combination	No. of Correct Prediction	% of Correct Result
1	ECD	53	1	29	54.72
	EFW	65	4	40	61.54
	MFW	84	3	45	53.57
	MCD	44	3	24	54.55
2	ECD	53	1,3	45	84.91
	EFW	65	4,3	55	84.62
	MFW	84	3,4	64	76.19
	MCD	44	3,5	34	77.27
3	ECD	53	1,3,2	42	79.25
	EFW	65	4,3,1	56	87.69
	MFW	84	3,4,1	73	86.91
	MCD	44	3,5,4	36	81.82
4	ECD	53	1,3,2,4	47	88.68
	EFW	65	4,3,1,2	57	86.15
	MFW	84	3,4,1,2	70	83.33
	MCD	44	3,5,4,2	32	72.73
5	ECD	53	1,3,2,4,5	44	83.02
	EFW	65	4,3,1,2,5	57	87.69
	MFW	84	3,4,1,2,5	74	86.91
	MCD	44	3,5,4,2,1	39	88.64

Thus we can see from the table, for the layer we have discussed here combination of all the parameters give the best result for Morning convective development, with a whopping 88.64% correct result. Whereas for Evening convective development combination of 4 parameters 1,2,3,4 gives a better result compared to the one with all 5 parameters, with again 88.68% correct result.

CONCLUSION

As any natural phenomenon is inherently complex and multivariate, its exact prediction is difficult. Not only that any natural phenomenon is obviously nonlinear in nature. Many attempts for pre-monsoon weather forecasting of Kolkata(22.53ºN,88.33ºE),India, have been tried by previous workers. But it is worth mentioning that this method works with better accuracy than the existing ones.

We have applied this rule to two datasets (morning and evening) of different situation (convective development and fair weather). The study reveals that the rule can classify the situation with atmost 88% success rate. Moreover, the study indicates that the two datasets are structurally different. The study reveals that the technique used in the present analysis, are efficient for forecasting pre-monsoon weather of Kolkata, India.It might be preferable to the field forecasters as it will reduce the number of parameters.

The technique suggested here is expected to work in any other domain too.

Acknowledgements

The authors wish to express their gratitude to India Meteorological Department, Kolkata for supplying necessary data for the study.

References

1 Ghosh S, et al. Reduction of number of parameters and forecasting convective developments at Kolkata (22.53ºN, 88.33ºE), India during pre-monsoon season: An application of multivariate techniques.(1999) pp 673-681.

2 Ghosh S, et al Comparison between LDA technique and fuzzy membership roster method for pre-monsoon weather forecasting (2011) pp 137-144.

3 Diaz H F & Fullbright D C, Eigenvector analysis of seasonal temperature, precipitation and synoptic-sale system frequency over contiguous United States: Part 1 (Winter), *Mon Weather Rev (USA)*, 109 (1981) pp 1267-1284.

4 Wallace J M & Gutzler D C, Telecommunications in the geopotential height field during the Northern Hemisphere winter, *Mon Weather Rev (USA)*, 109 (1981) pp 784-812.

5 Horel J D, A rotated principal component analysis of the interannual variability of the Northern Hemisphere 500 mb height field, *Mon Weather Rev (USA)*, 109 (1981) pp 2080-2092.

6 Overland J E & Preisendorfer R W, A significance test for principal components applied to a cyclone climatology, *Mon Weather Rev (USA)*, 110 (1982) pp 1- 4.

7 Maryon R H & Storey A H, A multivariate statistical model

for forecasting anomalies of half-monthly mean surface pressure, *Int J Climatol (UK)*, 5 (1985) pp 561-578.

8 Richman M B, Rotation of principal components: Review article, *Int J Climatol (UK)*, 6 (1986) pp 293-335.

9 Weare B C, Relationship between monthly precipitation and SST variation in the tropical Pacific region, *Mon Weather Rev (USA)*, 115 (1987) pp 687-698.

10 Ward M N & Folland C K, Prediction of seasonal rainfall

in the North Nordeste of Brazil using eigenvector of sea- surface temperature, *Int J Climatol (UK)*, 11 (1991) pp 711-743.

11 Showalter A K, A stability index for thunderstorms forecasting, *Bull Am Meteorol Soc (USA)*, 34 (1953) pp 250-252.

12 Darkow G L, The total energy environment of storms, *J Appl Meteorol (USA)*, 7 (1968) pp 199-205

13 Oduro-Afriyie K & Adefolalu D O, Instability indices for severe weather forecasting in West Africa, *Atmos Res (Netherlands)*, 30 (1993) pp 51-68.

14 Reddy P J, Barbarick D E & Osterberg R D, Development of a statistical model for forecasting episodes of visibility degradation in the Denver Metropolitan area, *J Appl Meteorol (USA)*, 34 (1995) pp 616-625.

15 Dasgupta S & De U K, Binary regression models for short term prediction of pre-monsoon Convective developments over

Kolkata (India), *Int J Climatol (UK)*, 27 (2007) pp 831-836.

16 Dasgupta S & De U K, Markov chain models for pre- monsoon thunderstorms in Calcutta, India, *Indian J Radio Space Phys*, 31 (2001) pp 138-142.

17 Kulkarni M K, Kandalgaonkar S S, Tinmake M R & Nath A,Pre-monson season thunderstorms over Pune, *Int J Climatol (UK)*, 22 (2002) pp 1415-1420.

18 Pant B & Shivhare R P, Markov chain model for study of wet / dry spells at AF Stn Sarsawa during SW monsoon season, *Vatavaran (India)*, 22 (1998) pp 37-50.

19 Thiagarajan R, Ramadoss & Ramaraj, Markov chain model for daily rainfall occurrences at east Thanjavur district, *Mausam (India)*, 46 (1995) pp 383-388.

20 Chowdhury A M, Ghosh S & De U K, Analysis of pre- monsoon thunderstorm occurrence at Dhaka from 1983 to 1992 in terms of (es- e) and convergence / divergence at surface, *Indian J Phys*, 70B (1996) pp 357-366.

21 Kishtawal C M, Basu S &d Pandey P C, An algorithm for retrieving vertical wind profiles

.

from satellite-observed winds over the Indian Ocean using complex EOF analysis, *J Appl Meteorol (USA)*, 35 (1996) pp 532-540.

22 Kim K Y, North G R & Huang J, EOFs of one-dimensional cycle stationary time series, computation, examples and stochastic modeling, *J Atmos Sci (USA)*, 53 (1996) pp 1007-1017.

23 Selten F M, A statistical closure of a low-order barotropic model, *J Atmos Sci (USA)*, 54 (1997) pp 1085-1093.

24 Kessler E, *Thunderstorm morphology and dynamics* (US Department of Commerce, USA), 1982, pp 93-95, 146-149.

25 Asnani G C, *Tropical Meteorology Vol 2* (Pune, India), 1992, pp 829-833, 852.

26 Williams E & Renno N, An analysis of the conditional instability of the tropical atmosphere, *Mon Weather Rev (USA)*, 121 (1993) pp 23-26.

27 Klir G. J. and B. O. Yuan, 2002. *Fuzzy sets and fuzzy logic, theory and applications*. Prentice Hall of India Pvt. Ltd.

.

SciForum
Mol2Net

DPPH• Free Radical Scavenging Activity of Coumarin Derivatives. *In silico* and *in vitro* Approach

Elizabeth Goya Jorge[1], Anita Maria Rayar[2], Stephen Jones Barigye[3], María Elisa Jorge Rodríguez[1] and Maite Sylla Iyarreta Veitía [2]*

[1] Pharmacy Department, Faculty of Chemistry and Pharmacy, Central University "Marta Abreu" of Las Villas, C-54830 Santa Clara, Cuba; egoyaj@gmail.com (E.G.J.), elisajorge@yahoo.es (M.E.J.R.)

[2] Equipe de Chimie Moléculaire du Laboratoire CMGPCE, EA 7341, Conservatoire national des arts et métiers, 2 rue Conté, 75003, Paris ; anitarayar@hotmail.fr (A.M.R.), maite.sylla@cnam.fr (M.S.-I.V.)

[3] Department of Chemistry, Federal University of Lavras, P.O. Box 3037, 37200-000 Lavras, MG, Brazil; sjbarigye@gmail.com (S.J.B.)

* Author to whom correspondence should be addressed; E-Mail: maite.sylla@cnam.fr;
Tel.: +33-1-58 80 84 82; Fax: +33-1-40 27 25 84.

Published: 4 December 2015

Abstract: The interest of coumarins as antioxidant agents has attracted much attention in recent years. A quantitative structure-activity relationship (QSAR) study of the DPPH• (*2,2-diphenyl-l-picrylhydrazyl*) radical scavenging ability of chemical compounds, based on the 0-3D DRAGON molecular descriptors and an artificial neural networks (ANN) technique was developed. The built mathematical model showed a correlation coefficient for the training set (R^2) = 0.71, an external correlation coefficient (Q^2_{ext}) = 0.65 and it was used to predict the antioxidant activity of 4-hydroxycoumarin derivatives. Besides, an experimental *in vitro* assay was developed for the reference compound of this group (4-hydroxycoumarin) and the results obtained confirmed the predictions made by the ANN.

Keywords: coumarin; artificial neural networks; antioxidant; QSAR; DPPH; free radical scavengers

1. Introduction

The development of antioxidant agents has attracted much attention in recent years, because oxidative damage is related to many pathological conditions [1]. Several coumarin derivatives have been studied for their biochemical and pharmacological profiles. Some studies suggest that these compounds may significantly affect the function of various mammalian cellular systems. Specifically their antioxidant effect has been explored, because the structural features of this group of compounds suggest that they can probably exhibit this pharmacologic property [2-4]. The antioxidant capacity can be experimentally measured by several *in vitro* assays. One of the best-known method is the one based on the capturing of the DPPH• radical [5,6].

Chemoinformatics tools have been used in the modeling of the antiradical activity, as well as others biological properties, given their advantages in saving time and resources [7,8]. Several statistical and machine learning methods have been widely used in the literature to build models for studying Quantitative Structure Activity Relationships (QSAR). The QSAR studies assess mathematical associations between structural features of the molecules and biological properties. For the last two decades, Artificial Neural Networks (ANN) have increasingly found applicability in QSAR studies, thanks to their ability to map non-linear relations between structural characteristics of chemical compounds and their chemical / biological behavior [9].

The *objective* of this study was to develop an ANN model in order to relate the chemical compounds' scavenging ability of the DPPH• radical with the corresponding structural features, also known as molecular descriptors (MDs). Then, an experiment to predict the antioxidant activity of a group of coumarin derivatives was performed. The coumarin-related compounds used as models in this study; were previously synthesized by the team of Molecular Chemistry at Cnam, Paris.

2. Results and Discussion

2.1. <u>Modeling</u>: The mathematical Multilayer Perceptron (MLP) neural model was constructed using DPPH• scavenging capacity of 1329 compounds reported in the literature. This model showed a correlation coefficient ($\boldsymbol{R^2}$) for the training set of 0.71. The predictive ability of the model was assessed using the external validation procedure, yielding a correlation coefficient ($\boldsymbol{Q^2_{ext}}$) of 0.65. Both values are above the limits established for model acceptance [10], and thus indicate the fitness and predictive power of the obtained ANN model.

2.2. <u>Prediction</u>: Recent advances in drug discovery have enabled a dramatic increase in the number of synthetic and naturally occurring molecules that are available for testing using *in vitro* assays as the scavenging ability of the DPPH• radical. Virtual screening allows for prior assessment of the potential bioactivity of chemical compounds, and thus providing key guidelines in posterior experimental work [11]. In this study, the MLP model obtained was used to predict DPPH• scavenging capacity of coumarin derivatives which were divided into 2 groups, following their structural analogy in function of the posterior analysis of their activity.

-*Group1*: cyclocoumarol analogous (**Cy-analogs**)

-*Group2*: warfarine analogous (**Wf-analogs**).

The results of the predictions for both groups are shown in **Table 1**. *Group1* (compounds **1-7**) and *Group2* (compounds **8-15**) have significantly different values of pIC_{50} as it can be noticed. **Cy-analogs** clearly seems to be less effective in

DPPH• radical capturing, because their values of pIC$_{50}$ are highest (around 4). On the other hand, the pIC$_{50}$ **Wf-analog** values are under 3.3. In the case of 4-hydroxycoumarin (number 15) the pIC$_{50}$ value obtained was of 3.4. These results indicate the superior ability of the compounds *Group2* for scavenging the DPPH• radical. To justify the observed trends, a more detailed analysis of the structural characteristics is required. Firstly, the presence of a free hydroxyl group in **Wf-analogs** may probably favor the higher antioxidant activity of this group. In fact, several research studies have pointed hydroxyl groups as key for antiradical capacity and consequently, antioxidant activity [8, 12-15]. The hydroxyl group is present in the most frequently used reference compounds like: trolox, gallic acid or butylated hydroxytoluene (BHT).

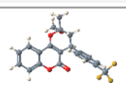
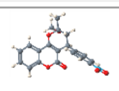
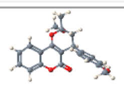
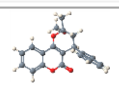
2.3. In vitro Assay: The results obtained with *in silico* modeling were corroborated by an *in vitro* study of DPPH• scavenging capacity for the reference compound, 4-hydroxycoumarin. Selecting this compound was specifically based on the prediction results obtained because, according to the model, this molecule has an intermediate pIC$_{50}$ value. The experimental result of the *in vitro* assay can provide a comparison criterion for evaluating the neural model and its predictions.

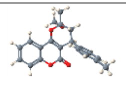
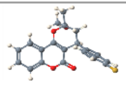
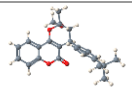
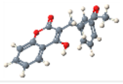
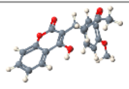
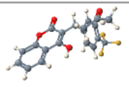
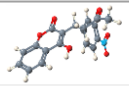
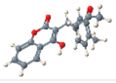
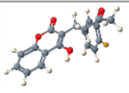
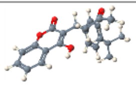
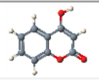
The experimental *in vitro* pIC$_{50}$ value (2.7) obtained according to the method described below for 4-hydroxycoumarin is to be compared with the prediction value 3.4 given by the model. This result is positive, because the neural model as well as the *in vitro* assay are in the scale of high activity according to the statistical range established by the built Data and also compared to the value obtained experimentally for BHT (2.10) use as reference.

It may thus be suggested that the 4-hydroxycoumarin possesses significant radical scavenging ability, and therefore it can be considered as a candidate for antioxidant agent; although more analyses are necessary to ensure that.

The results obtained in the *in vitro* assay confirmed the predicting power of designed ANN model, and consequently its applicability in the search for new antioxidant compounds. An *in vitro* study of antioxidant activity of warfarine analogues is currently underway.

Table 1. Predictions of the pIC$_{50}$ values.

Nº	3D Structures	IUPAC name	Predicted pIC$_{50}$
1		4-(4-(trifluoromethyl)phenyl)-3,4-dihydro-2-methoxy-2-methylpyrano[3,2-c]chromen-5(2H)-one	3,881001
2		3,4-dihydro-2-methoxy-2-methyl-4-(4-nitrophenyl)pyrano[3,2-c]chromen-5(2H)-one	3,951100
3		3,4-dihydro-2-methoxy-4-(4-methoxyphenyl)-2-methylpyrano[3,2-c]chromen-5(2H)-one	3,829745
4		3,4-dihydro-2-methoxy-2-methyl-4-phenylpyrano[3,2-c]chromen-5(2H)-one	3,943571

5		3,4-dihydro-2-methoxy-2-methyl-4-p-tolylpyrano[3,2-c]chromen-5(2H)-one	3,946035
6		4-(4-fluorophenyl)-3,4-dihydro-2-methoxy-2-methylpyrano[3,2-c]chromen-5(2H)-one	3,959642
7		4-(4-tert-butylphenyl)-3,4-dihydro-2-methoxy-2-methylpyrano[3,2-c]chromen-5(2H)-one	3,902871
8		4-hydroxy-3-(3-oxo-1-phenylbutyl)-2H-chromen-2-one	3,282264
9		4-hydroxy-3-(1-(4-methoxyphenyl)-3-oxobutyl)-2H-chromen-2-one	3,253075
10		3-(1-(4-(trifluoromethyl)phenyl)-3-oxobutyl)-4-hydroxy-2H-chromen-2-one	3,253213
11		4-hydroxy-3-(1-(4-nitrophenyl)-3-oxobutyl)-2H-chromen-2-one	3,281052
12		4-hydroxy-3-(3-oxo-1-p-tolylbutyl)-2H-chromen-2-one	3,283764
13		3-(1-(4-fluorophenyl)-3-oxobutyl)-4-hydroxy-2H-chromen-2-one	3,293596
14		3-(1-(4-tert-butylphenyl)-3-oxobutyl)-4-hydroxy-2H-chromen-2-one	3,269560
15		4-hydroxy-2H-chromen-2-one	3,365531

3. Materials and Methods

Data: Experimental results of the scavenging ability of the DPPH• radical (expressed as IC_{50}) for 1329 molecules extracted over 170 scientific reports in the literature; thus yielding a comprehensive and diverse database of compounds for the mathematical analysis. All the structures were optimized using CORINA software. The response variable values (IC_{50}) were transformed to their corresponding pIC_{50} values.

Molecular Descriptors: The parameterization of the structures was performed using 3224 molecular descriptors implemented in the DRAGON software. A wrapper based variable selection procedure was used to obtain a subset of 14 variables for the ANN building: MATS2e, BELe6, HATS3u, H2v, R7v, nN-N, nImidazoles, C-005, C-020, O-057, O-060, GVWAI-50, B02 [O-S] and B07 [O-S]*.

Development of ANN model: The QSAR model was develop using as chemometric tool a Multilayer Perceptron Neural Network implemented in STATISTICA 8.0 software. For the modeling, a Broyden-Fletcher-Goldfarb-Shanno training algorithm was used as the optimization method. The following network architecture was established: fourteen inputs; eight neurons in hidden layer and one output.

*Predictions***:** The coumarin derivatives were optimized following the same configuration previously used and the corresponding MD computed.

In vitro DPPH• assay: The free radical scavenging activity of the 4-hydroxycoumarin was measured using the stable DPPH• radical, according to Blois´s method [16]. Briefly, 0.1 mM solution of DPPH• in methanol was prepared and this solution (1 mL) was added to a sample solution in methanol (3 mL) at different concentrations (150–750 μg/mL). The mixture was shaken vigorously and left to stand for 30 min in the dark, and the absorbance was then measured at 517 nm. BHT was use for comparison. The procedure was triplicate to ensure the results. The capability to scavenge the DPPH• radical was expressed as IC_{50} (concentration of antioxidant that produces 50% of absorbance inhibition).

4. Conclusions

The scavenging capacity of the DPPH• radical is one of the most extended method to evaluate the *in vitro* antiradical activity. An MLP neural network was constructed to relate the structure of 1329 molecules and their antiradical activity. The obtained model showed adequate fitness and a good predictive power and was thus used to predict the antioxidant activity of 15 coumarin derivatives. The *in silico* predictions were further corroborated by an *in vitro* assay for one of the molecules considered as the reference for this set of compounds, and the obtained IC_{50} value was similar to the value predicted by the MLP model.

Acknowledgments

MENRT is gratefully acknowledged for the graduate fellowship awarded to Anita Rayar. The authors gratefully acknowledge support for part of this work through the program PYTHAGORAS II of EPEAEK II (MIS: 97436/073). We would like to thank Dr. C. Hansch and Biobyte Corp. 201 West 4th Street, Suite 204, Claremont, CA 91711, USA for free access to the C-QSAR program. The authors also thank the “Service de Coopération et d'Action Culturelle” of Embassy of France in Cuba for the financial support.

Author Contributions

E.G.J. and S.J.B. are responsible for the model construction and evaluation. M.E.J.R. is responsible for the *in vitro* assay. The French team, A.M.R. and M.S.-I.V., provide the coumarin derivatives structures from their chemical library. All authors contributed to the drafting and revision of the article and approved the final version.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Valko, M.; Leibfritz, D.; Moncol, J.; Cronin, M.T.D.; Mazur, M.; Telser, J. Free radicals and antioxidants in normal physiological functions and human disease. *Int. J. Biochem. Cell Biol.* **2007**, *39*, 44-84.
2. Naceur, H.; Fischmeister, C.; Puerta, M.C.; Valerga, P. A rapid access to new coumarinyl chalcone and substituted chromeno[4,3-c]pyrazol-4(1H)-ones and their antibacterial and DPPH radical scavenging activities. *Med.Chem. Res.* **2011**, *20*, 522-30.
3. Puerta, M.C.; Naceur, H.; Valerga, P. Synthesis, structure, antimicrobial and antioxidant investigations of dicoumarol and related compounds. *Eur. J. Med. Chem.* **2008**, *43*, 2541-8.

*To see the description about each variable, please consult the Handbook of Molecular Descriptors

4. Fylaktakidou, K.; Hadjipavlou-Litina, D.; Litinas, K.; Nicolaides, D. Natural and synthetic coumarin derivatives with antiinflammatory/antioxidant activities. *Curr Pharm* **2004**, 10, 3813-33.
5. Gunars, T.; Grzegorz, B. Determination of antiradical and antioxidant activity: basic principles and new insights. *Acta Biochim.Pol.* **2010**, 57, 139-142.
6. Kedare, S.B.; Singh, R.P. Genesis and development of DPPH method of antioxidant assay. *J. Food Sci. Technol.* **2011**, 48, 412-422.
7. Todeschini, R.; Consonni, V.; Gramatica, P. Chemometrics in QSAR. In: *Comprehensive Chemometrics Chemical and Biochemical Data Analysis*, 1st ed.; Brown S.D, Tauler R., Walczak B.,Eds.; Elsevier: Oxford, UK, 2009; Volume 4, pp. 129-72.
8. Razo-Hernández, R.S.; Pineda-Urbina, K.; Velazco-Medel, M.A.; Villanueva-García, M.; Sumaya-Martínez, T.; Martínez-Martínez, F.J.; et al. QSAR study of the DPPH• radical scavenging activity of coumarin derivatives and xanthine oxidase inhibition by molecular docking. *Cent. Eur. J. Chem.* **2014**, 12, 1067-80.
9. Aoyama, T.; Suzuki, Y.; Ichikawa, H. Neural networks applied to structure-activity relationships. *J Med Chem* **1990**, 33, 905-8.
10. Tropsha, A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Inf.* **2010**, 29, 476−88.
11. McInnes, C. Virtual screening strategies in drug discovery. *Curr. Op. Chem. Biol.* **2007**, 11, 494-502.
12. Mitra, I.; Saha, A.; Roy, K. Chemometric modeling of free radical scavenging activity of flavone derivatives. *Eur. J. Med. Chem.* **2010**, 45, 5071-9.
13. Worachartcheewan, A.; Nantasenamat, C.; Isarankura-Na-Ayudhya, C.; Prachaiasittikul, S.; Prachaiasittikul, V. Predicting the free radical scavenging activity of curcumin derivatives. *Chemometr. Intell. Lab.* **2011**, 109, 207–16.
14. Wright, J.S.; Johnson, E.R.; DiLabio, G.A. Predicting the Activity of Phenolic Antioxidants: Theoretical Method, Analysis of Substituent Effects, and Application to Major Families of Antioxidants. *A. Chem. Soc.* **2001**, 123, 1173-83.
15. Yamagami, C.; Motohashi, N.; Emoto, T.; Hamasaki, A.; Tanahashi, T.; Nagakura, N.; et al. Quantitative structure-activity relationship analyses of antioxidant and free radical scavenging activities for hydroxybenzalacetones. *Bioorg. Med. Chem. Lett.* **2004**, 14, 5629-33.
16. Blois, M.S. Antioxidant determinations by the use of a stable free radical. *Nature* **1958**, 181, 1199–1200.

SciForum
Mol2Net

Regioselective Friedel-Crafts Hydroxyalkylation Using Friendly Conditions: Application to the Synthesis of Unsymmetrical Triarylmethanes

Maité Sylla Iyarreta Veitía *, Céline Rampal, Clotilde Ferroud

Equipe de Chimie Moléculaire du Laboratoire CMGPCE, EA 7341, Conservatoire national des arts et métiers, 2 rue Conté, 75003, Paris; maite.sylla@cnam.fr (M.S.-I.V.), celine.rampal@free.fr (C.R.), clotilde.ferroud@cnam.fr (C.F.)

* Author to whom correspondence should be addressed; E-Mail: maite.sylla@cnam.fr;
Tel.: +33-1-58 80 84 82; Fax: +33-1-40 27 25 84.

Published: 4 December 2015

Abstract: Friedel-Crafts alkylation is one of the most important methods used in organic chemistry to create carbon-carbon bonds. The traditional conditions using alkyl halides activated by a Lewis or Brønsted acid have been widely described in the literature. Nevertheless, the Friedel-Crafts reaction using aldehydes or ketones as substrates, known as hydroxyalkylation, has been poorly described. The development of regioselective Friedel-Crafts conditions is a challenge in organic synthesis. The growing interest in developing “metal and solvent free" reactions justifies the particular attention that the Brønsted acids have received as an alternative to toxic and precious metals. During the last years, triarylmethanes attract considerable attention due to their applications in fine, medicinal and industrial chemistry. They have been used as protecting groups, dyes or photochromic agents. They also exhibit interesting biological properties including anti-tumor and antioxidant activities. Different methods of preparation of symmetrical triarylmethanes have been described in the literature, nonetheless the synthesis of unsymmetrical ones has been very little explored. In this work we describe an efficient regioselective method to prepare unsymmetrical triarylmethanes from enriched aromatic compounds, *via* a Friedel-Crafts hydroxyalkylation catalyzed by Brønsted acids using pyridylarylcarbinols as alkylating agents. The method described here is consistent with the principles of green chemistry and has significant advantages, such as the use of an inexpensive catalyst and mild conditions. This regioselective method offers good yields, shorter reaction times and a possible extension to various substrates.

Keywords: regioselective hydroxyalkylation; triarylmethanes, Friedel-Crafts reaction

1. Introduction

In recent years triarylmethanes (TAMs) have received considerable attention from scientific community due to their numerous applications. In chemical industry they have been used as leuco dyes, photochromic agents, protective groups in organic synthesis or building blocks for generating dendrimers .They have been used in food industry, in textile industry, as well as antifungal agents and parasiticides in the fish farm industry. [1-4] Recent studies have shown a widespread application of TAMs in therapeutics due to their several biological activities such as antiviral, antitumor, antitubercular, antifungal, anticancer and anti-inflammatory agents [5-7].

The synthesis of unsymmetric TAMs is less described in the literature because it generally leads to low yields and to the formation of byproducts principally due to lack of regioselectivity. As a result, the development of new approaches to synthesize unsymmetrical TAMs in simpler, efficient, regioselective and environmentally friendly ways is a real challenge.

The most common method for preparing unsymmetrical TAMs is a Friedel-Crafts hydroxyalkylation from the corresponding carbinols. Unsymmetrical TAMs bearing electrowithdrawing groups have been also prepared *via* the formation of alkylbenzotriazoles. [8-11]. Finally, coupling reactions using metal catalysts (Pd, Cu…), or Friedel-Crafts reactions using sulfone or α-amidosulfones have been also reported [12-14].

In this work we described the regioselective synthesis of unsymmetrical triaylmethanes *via* a Friedel-Crafts hydroxyalkyalation from the corresponding functionalized carbinols and activated aromatic compounds using a Brønsted acid as catalytic system. The scope of this method is also noted herein.

.

2. Results and Discussion

As part of our studies involving the synthesis of bioactive compounds structurally related with a triarylmethane skeleton, we have recently focused our attention on the synthesis of unsymmetrical TAMs with interesting anti-inflammatory properties. We were particularly interested in the synthesis of the *p-p* regioisomers derived from pyridinophenylcarbinols and phenol. In our preliminary studies we noticed that a major regioisomer could be obtained according to the amounts of catalyst used. These results encouraged us to develop an efficient and regioselective method to prepare unsymmetrical TAMs via a Friedel-Crafts hydroxyalkylation. To carry out this study we selected the (4-*tert*-butyl-phenyl) pyridine-2-yl-methanol **1** as a model. Each assay was carried out on a scale of 0.4 mmol with 1.2 equivalents of phenol using sulfuric acid as catalyst. The reactions were monitored by thin layer chromatography (TLC). Finally, the influence of solvent, temperature and the amount of catalyst was studied in details.

2.1 *Screening of solvents*

We decided to use either benzene, toluene, 1,2-dichloroethane or nitrobenzene as solvent. The first experiments were carried out at 80 °C with 4 to 20 equivalents of sulfuric acid. In benzene and toluene with 4 equivalents of sulfuric acid the *p-p* TAM is obtained as a major regioisomer, with a yield not exceeding 45%. In addition, these aromatic solvents may compete with phenol, thereby favoring the formation of byproducts.

In 1,2-dichloroethane, the same trend was observed, i.e. the formation of a major *o-p* regioisomer using 20 equivalents of sulfuric acid and the formation of a major *p-p* regioisomer using 4 equivalents of sulfuric acid. Nonetheless *o-p* and *p-p* regioisomers were obtained with modest yields of 46% and 35% respectively.

The best results were obtained when nitrobenzene was used. It allowed short reaction times, homogeneous reaction medium and it did not react with the corresponding carbinol thus limiting the formation of by-products. Moreover a very good regioselectivity was observed, with total conversion and good yields (98% *o-p* regioisomer with 20 equivalents of sulfuric acid and 72% *p-p* regioisomer with 4 equivalents of sulfuric acid). However it can be noticed that the high boiling point of nitrobenzene (bp = 210 ° C) prevents its removal by distillation and flash chromatography must be used in the purification step. Taking into account these arguments the nitrobenzene has been selected for further studies.

2.2 Screening of temperature

The temperature screening was carried out in nitrobenzene, with 4 or 20 sulfuric acid equivalents. We studied the reaction at 80 ° C, 20 ° C and 0 ° C (Table 1). At 80 °C with 20 equivalents of sulfuric acid only the *o-p* regioisomer is obtained with 98% yield after purification by flash chromatography on silica gel. On the other hand at 80 °C and using 4 equivalents of sulfuric acid the *p-p* regioisomer is principally obtained with 72% yield.

The same conditions were evaluated at 20 °C. In such conditions when 20 equivalents of catalyst were used the *p-p* regioisomer was obtained as the major compound in 5 min with 71% yield. On the other hand the use of 4 equivalents of acid at 20°C reduces the reaction time. After 48 h, *p-p* regioisomer and the *o-p* regioisomer were obtained with yields of 66% and 3% respectively.

At 0 ° C, the reaction was only carried out with 20 equivalents of sulfuric acid. Under these conditions, the *p-p* regioisomer **3** is mainly obtained after 15 min with a good yield of 78%. To summarize, the temperature plays an important role in the regioselectivity of the Friedel and Crafts hydroxyalkylation when 20 equivalents of acid are used. By eating at 80 °C the *o-p* regioisomer **4** is obtained with a yield of 98%. At room temperature or at 0 ° C the *p-p* regioisomer is isolated with respective yields of 72% and 78%.

These results suggest that *p-p* regioisomer would be the kinetic compound since it is obtained at low temperature and *o-p* regioisomer obtained at 80 °C would be the thermodynamic compound. This can be justified by assuming that hydrogen bonds may be formed between the pyridine nitrogen and the hydroxyl function of the phenol, leading to better stability of the molecule (Figure 2).

2.3 Screening of amounts of catalyst

To focus on the influence of the amount of catalyst, we studied this reaction using 0.02; 2; 4; 5 and 20 equivalents of sulfuric acid. In the literature, the synthesis of TAMs is reported using a catalytic amount of acid. [15-18]. In our study when 0.02 or 2 equivalents of acid were used, the reaction did not occur. This could be explained by the protonation of the pyridine carbinol in acidic media. Theoretically almost 2 acid equivalents of acid should be enough. However, 4 equivalents of sulfuric acid are required to completely consume the corresponding carbinol. The *p-p* regioisomer **3** is predominantly isolated using 4 equivalents of sulfuric acid however at 20 ° C the reaction is rather slow in comparison with the reaction carried out at 80 ° C (5 min). The use of 20 equivalents of acid affords exclusively the *o-p* regioisomer with an excellent yield of 98%.

In conclusion, we outline the regioselective conditions for the Friedel-Crafts hydroxyalkylation between phenol and the carbinol **1**. The best conditions use nitrobenzene as solvent and 80 °C. The amount of sulfuric acid influences the formation of the corresponding regioisomers.

The best conditions for synthesizing the *p-p* regioisomer are: 4 equivalents of sulfuric acid in 5 min at 80 ° C. 20 equivalents of sulfuric acid at 80 °C lead to the *o-p* regioisomer in 5 min.

A study on different functionalized carbinols and activated aromatic compounds is currently underway.

Figure 1. Reaction between (4-*tert*-butyl-phenyl) -pyridin-2-yl-methanol and phenol.

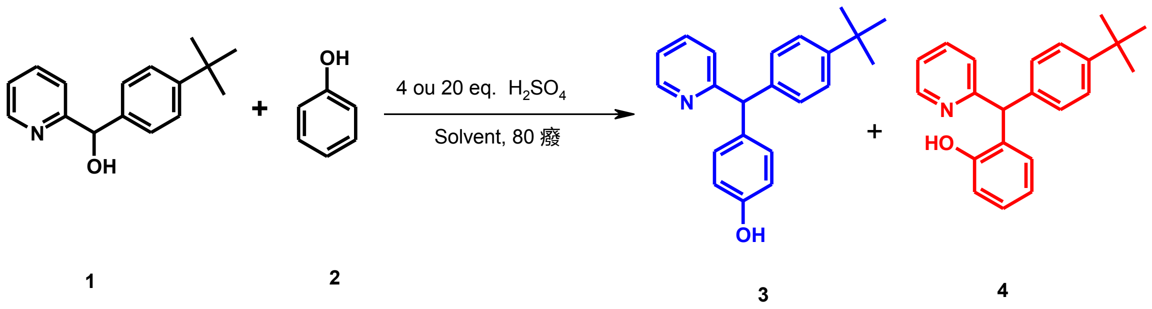
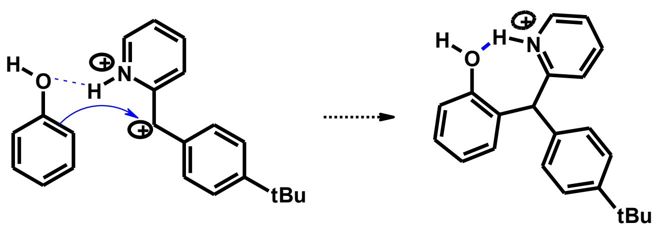


Table 1. Results obtained at different temperatures

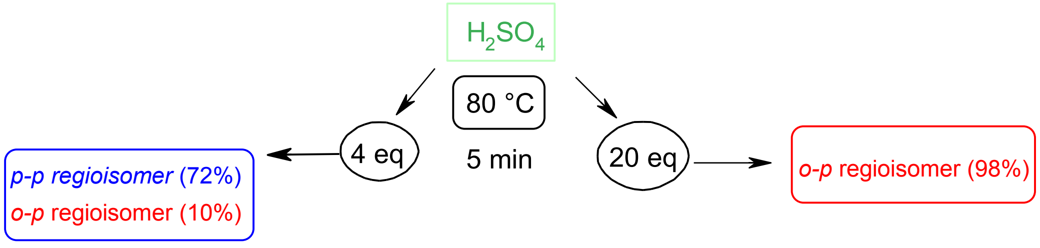
T °C	eq H$_2$SO$_4$	time	yield*% *p-p* **1**	yield* % *o-p* **2**
80	20	5 min	0	98
	4	5 min	72	10
20	20	5 min	72	3
	4	48 h	66	3
		5 h	58	3
0	20	15 min	78	2

* *After purification by silica gel chromatography*

Figure 2. Stabilization by establishing a hydrogen bond between the hydroxyl group and the nitrogen atom of pyridine



Scheme 1 : Selected conditions for the best regioselectivity



3. Materials and Methods

All reagents were obtained from commercial sources unless otherwise noted, and used as received. Heated experiments were conducted using thermostatically controlled oil baths and were performed under an atmosphere oxygen-free in oven-dried glassware. All reactions were monitored by analytical thin layer chromatography (TLC) or by Gas chromatography-Mass spectrometry (GC-MS). TLC was performed on aluminium sheets precoated silica gel plates (60 F_{254}, Merck). TLC plates were visualized using irradiation with light at 254 nm or in an iodine chamber as appropriate. Flash column chromatography was carried out when necessary using silica gel 60 (particle size 0.040-0.063 mm, Merck). All synthesized compounds were characterized by NMR, IR, MS data and by the TLC behavior.

General Procedure for the synthesis of carbinols

To a solution of 2-bromopyridine (1 eq.) in anhydrous THF at -78 ° C, was added a solution of *i*-propylmagnesium chloride 1M. After 2 h at room temperature the corresponding aldehyde (1.03 eq.) was added dropwise. After stirring for 20 hours at room temperature, the mixture was hydrolysed with distilled water and then extracted with dichloromethane. The organic phases were dried over anhydrous Na_2SO_4, filtered and concentrated to give an oil which crystallized when cold. The precipitated product was filtered and rinsed with CyHex/EtOAc mixture and dried. If necessary product was purified by flash chromatography on silica gel.

General procedure for the synthesis of o-p regioisomers

To a solution of the corresponding carbinol (1 eq.) and phenol (1.2 eq.) in nitrobenzene (0.4 M) was added dropwise concentrated sulfuric acid (20 eq.). After 5 min at 80 ° C, the reaction medium was cooled to ambient temperature and then neutralized with a saturated solution of $NaHCO_3$ until pH 7, then extracted with EtOAc. The combined organic phases were dried over anhydrous Na_2SO_4, filtered and concentrated. The crude product was purified by flash chromatography on silica gel.

General procedure for the synthesis of o-p regioisomers

To a solution of the corresponding carbinol (1 eq.) and phenol (1.2 eq.) in nitrobenzene (0.4 M) was poured dropwise concentrated sulfuric acid (4 eq.). After 5 min at 80 ° C the reaction mixture was cooled to room temperature and then neutralized by pouring slowly a saturated solution of $NaHCO_3$ until pH 7-8, then extracted with EtOAc. The combined organic phases were dried over anhydrous Na_2SO_4, filtered and concentrated, filtered and

concentrated. The crude product was purified .
by flash chromatography on silica gel. .

4. Conclusions

The study of the regioselective Friedel-Crafts hydroxyalkylation between a functionalized carbinol and phenol was developed using a Brønsted acid as promoting system. The obtained results highlight an effective method to synthesize unsymmetrical triarylmethanes, regioselectively, under mild conditions with good yields.

Acknowledgments

FONGECIF is gratefully acknowledged for the financial support offered to Céline Rampal.

Author Contributions

C.R. performed experiments and analyzed data; M.S-IV designed experiment, analyzed data and wrote the paper and C. F. wrote the paper and supervised the project. All authors contributed to the drafting and revision of the article and approved the final version.

Conflicts of Interest

The authors declare no conflict of interest

References and

1. Turner, A.R.E.R.F.M., ed. *The condensed chemical dictionary*. sixth ed. **1961**, Reinhold publishing corporation, New York, Chapman and Hall, L. T. D. London. 1026.
2. Kirk-Othmer, ed. *Encyclopedia of chemical technology*. third edition ed. Vol. 8. **1979**, John Wiley and sons, New York.
3. Freund, E., and coll., *Solid-phase synthesis using (allyloxy)carbonyl (alloc) chemistry of a putative heptapeptide intermediate in vancomycin biosynthesis containing m-chloro-3-hydroxytyrosine.* Helv. Chim. Acta, **2000**. 83 (9) : p. 2572-2579.
4. Garcia, M.L., and coll., *Synthesis of new ether glycerophospholipids structurally related to modulator.* Tetrahedron, **1991**. 47 (48) : p. 10023-10034.
5. Kowaltowski, A.J., and coll., *Mitochondrial effects of triarylmethane dyes.* J. Bioenerg. Biomembr., **1999**. 31 (6) : p. 581-590.
6. Panda, G., and coll., *Diaryloxy methano phenanthrenes: a new class of antituberculosis agents.* Bioorg. Med. Chem., **2004**. 12 (20) : p. 5269-5276.
7. Sumoto, K., et coll., *Synthesis of 2,2'-dihydroxybisphenols and antiviral activity of some bisphenol derivatives.* Chem. Pharm. Bull., **2002**. 50 (2) : p. 298-300.
8. Katritzky, A.R., X. Lan, and J.N. Lam, *Benzotriazole mediated synthesis of methylenebisanilines.* Synthesis, **1990**, (4) : p. 341-346.

9. Katritzky, A.R., X. Lan, and J.N. Lam, *Benzotriazole as a synthetic auxiliary: advantageous syntheses of substituted diarylmethanes and heterocyclic analogs.* J. Org. Chem., **1991**. 56 (14) : p. 4397-4403.
10. Katritzky, A.R. and X. Lan, *Benzotriazole-mediated arylalkylation and heteroarylalkylation.* Chem. Soc. Rev., **1994**. 23 (6) : p. 363-373.
11. Katritzky, A.R., S.A. Henderson, and B. Yang, *Applications of benzotriazole methodology in heterocycle ring synthesis and substituent introduction and modification.* J. Heterocycl. Chem., **1998**. 35 (5) : p. 1123-1159.
12. Zhang, J., and coll., *Palladium-Catalyzed C(sp3)-H Arylation of Diarylmethanes at Room Temperature: Synthesis of Triarylmethanes via Deprotonative-Cross-Coupling Processes.* J. Am. Chem. Soc., **2012**. 134 (33) : p. 13765-13772.
13. Thirupathi, P., L.N. Neupane, and K.-H. Lee, *Tris(pentafluorophenyl)borane [B(C_6F_5)$_3$]-catalyzed Friedel-Crafts reactions of activated arenes and heteroarenes with a-amido-sulfones: the synthesis of unsymmetrical triarylmethanes.* Tetrahedron, **2011**. 67 (38) : p. 7301-7310.
14. Thirupathi, P. and S.S. Kim, *Regioselective Arylations of a-Amido Sulfones with Electron-Rich Arenes through Friedel-Crafts Alkylations Catalyzed by Ferric Chloride Hexahydrate: Synthesis of Unsymmetrical and Bis-Symmetrical Triarylmethanes.* Eur. J. Org. Chem., **2010**. (9) : p. 1798-1808.
15. Parai, M.K., and coll., *Thiophene containing triarylmethanes as antitubercular agents.* Bioorg. Med. Chem. Lett., **2008**. 18 (1) : p. 289-292.
16. Das, S.K., S. Panda, and G. Panda, *An easy access to unsymmetric trisubstituted methane derivatives (TRSMs).* Tetrahedron Lett., **2005**. 46 (17) : p. 3097-3102.
17. Shagufta, and coll., *Substituted phenanthrenes with basic amino side chains: A new series of anti-breast cancer agents.* Bioorg. Med. Chem., **2006**. 14 (5) : p. 1497-1505.
18. Panda, G., and coll., *Design, synthesis and antitubercular activity of compounds containing aryl and heteroaryl groups with alkylaminoethyl chains.* Indian J. Chem., Sect. B: Org. Chem. Incl. Med. Chem., **2009**. 48B (8) : p. 1121-1127.

SciForum
Mol2Net

QSAR for the Characterization of Drug Resistance: Differential QSAR (DiffQSAR) Using Mathematical Chemodescriptors

Subhash C. Basak

[1]University of Minnesota Duluth-Natural Resources Research Institute (UMD-NRRI) and Department of Chemistry and Biochemistry, University of Minnesota Duluth, 5013 Miller Trunk Highway, Duluth, MN 55811, USA; sbasak@nrri.umn.edu; Tel.: +1-218-727-1335

Published: 4 December 2015

Abstract: Drug resistance is a serious issue that compromises the efficacy of many drugs and antibiotics. One mechanism underlying the development of resistance is the alternation in the target enzyme or receptor resulting in gradual silencing of the target to the effects of the ligands. Basak et al developed a method called differential QSAR (DiffQSAR) whereby test data of drugs on their effects on the sensitive and resistant targets are used to characterize the phenomenon of drug resistance using mathematical molecular descriptors. This paper will summarize our research in this area.

Keywords: Quantitative structure-activity relationship (QSAR); Differential QSAR (DiffQSAR); Drug resistance; Silencing of target; Mathematical descriptors; Ridge regression (RR); Dihydrofolate reductase (DHFR); Plasmodium falciparum (Pf); Topological Indices; Atom pairs

1. Introduction

Drug resistance is a phenomenon that is creating problems in the continuing clinical efficacy of drugs through the development of gradually declining potency of drugs. To give just a couple of examples, resistance to drugs have developed for diseases like cancer [1], H_5N_1 pandemic Bird Flu infection [2], and malaria [3]. In the case of malaria, dihydrofolate reductase (DHFR) of *Plasmodium falciparum (Pf)* is an important target for antimalarial drug discovery because it catalyzes a critical step in the biochemical pathway of the parasite, viz., the reduction of dihydrofolate to tetrahydrofolate, which has a critical role in the DNA synthesis of the parasite [3-5]. As a result, various modelling methods have been used in understanding the structural basis of the antimalarial activity of DHFR inhibitors [6].

Recently, Sivaprakasam et al. [7] carried out quantitative structure-activity relationship (QSAR) and docking studies of cycloguanil PfDHFR-TS inhibitors.

The above indicates that fast screening of chemical databases for their activity against target macromolecules in *Pf* is essential for effective antimalarial drug discovery. This can be accomplished if the screening models are based on molecular descriptors which can be calculated fast and directly from the molecular structure without the input of any other experimental data. Therefore, we carried out a QSAR analysis of 58 cycloguanil *Pf*DHFR inhibitors using computed topological descriptors [8-16].

2 Results and Discussion

Data on cycloguanil derivatives and mathematical descriptors were used to develop QSARs that can be used for the screening of chemicals. . The results of statistical analysis showed that one compound, compound #22, in ref [7] was an influential outlier. The same conclusion was drawn from the QSAR studies of Sivaprakasam et al. [7]. Both topological indices (TIs) and TI plus atom pair (AP) combination give good QSARs for the binding affinity (K_i) of the cycloguanil derivatives. Some improvement in model quality was observed after the addition of APs to the set of topological indices. Of the three statistical methods used for modelling, viz., ridge regression (RR), partial least square (PLS) , and principal components regression (PCR), RR outperformed the other two This is in line with our numerous previous observations with QSAR modelling of various property/bioactivity data.
A total of 369 TIs were calculated using programs including POLLY, Triplet, and Molconn-Z. Atom pairs were calculated by APProbe [14]. A look at the top 20 descriptors extracted from the QSARs of the sensitive versus the resistant strains sorted by their t values show that only two descriptors, viz., AZV4 and ANV3, are common between the two models. For details see [16].

Thus it can be said that the descriptor space created by the set of calculated mathematical descriptors can provide a subsets of descriptors which can differentiate the chemical-biological interactions between the sensitive versus the resistant forms of the drug target, PfDHFR.

3. Materials and Methods

For materials and methods of data collection and statistical analyses, see [16].

.

.

4. Conclusions

Using high dimensional structure space consisting of calculated topological indices and atom pairs, ridge regression was applied to develop QSARs for the sensitive and resistant forms of dihydrofolate reductase (DHFR) inhibitors of *Plasmodium falciparum* (Pf). The top 20 descriptors extracted from the QSAR models developed by ridge regression showed that the subsets of non-overlapping descriptors are capable of characterizing the silencing of the target arising out of mutation in the genetic apparatus of the organism, *Plasmodium falciparum* (Pf). It is expected that such research can be carried out with other drug targets to characterize the molecular basis of resistance based on computed properties of the ligands involved in the interactions with biochemical targets.

Acknowledgments

The author is grateful to Douglas Hawkins (School of Statistics, University of Minnesota, TC campus), Denise Mills (former collaborator of Basak at NRRI), and Greg Grunwald (NRRI) for sustained collaboration in his QSAR research program at UMD-NRRI.

Author Contributions

Since the early 1970sn Subhash C. Basak has been involved in the development of novel topological indices and their applications in QSARs pertaining to the estimation of property/ bioactivity/ toxicity of chemicals.

Conflicts of Interest

This author declares no conflict of interest.

References and Notes

[1] Choi, Y.L. et al. EML4-ALK mutations in lung cancer that confer resistance to ALK inhibitors. N. Engl. J. Med., 2010, 363, 1734–1739.

[2] deJong, M. D. et al. Oseltamivir resistance during treatment of influenza A (h5N1) infection. N. Engl. J. Med., 2005, 353, 2667–2672.

[3] Hyde, J. E. Drug-resistant malaria, Trends Parasitol., 2005, 21, 494–498.

[4] Gregson, A.; Plowe, C. V. Mechanisms of resistance of malaria parasites to antifolates, Pharmacol. Rev., 2005, 57, 117–145.

[5] Yuthavong, Y., et al Malarial (Plasmodium falciparum) dihydrofolate reductase thymidylate synthase: structural basis for antifolate resistance and development of effective inhibitors, Parasitology, 2005, 130, 249–259.

[6] Adane, L.; Bharatam, P. V. Modelling and informatics in the analysis of P-falciparum DHFR Enzyme inhibitors, Curr. Med. Chem. 15 (2008), pp. 1552–1569.

[7] Sivaprakasam, P.; Tosso, P. N.; Doerksen, R. J. Structure-activity relationship and comparative docking studies for cycloguanil analogs as PfDHFR-TS inhibitors, J. Chem. Inf. Model., 2009, 49, 1787–1796.

[8] Basak, S. C., Restrepo, G. and Villaveces, J. L., Eds, Advances in Mathematical Chemistry and Applications, volume 1 & 2 , Bentham eBooks, Bentham Science Publishers and Elsevier, , 2015.

[9] Devillers, J.; Balaban, A.T., Eds. Topological Indices and Related Descriptors in QSAR and QSPR; Gordon and Breach: Amsterdam, 1999, pp. 811.

[10] Gonzalez-Diaz, H.; Munteanu, C. R. (Editors), Topological Indices for Medicinal Chemistry, Biology, Parasitology, Neurological and Social Newworks, Transworld Research Neywork, 2011.

[11] Basak, S. C.; Harriss, D. K.; Magnuson, V. R. 1988. POLLY v. 2.3: 1988; Copyright of the University of Minnesota.

[12] MolConnZ, Version 4.05, 2003; Hall Ass. Consult. ; Quincy, MA.

[13] Basak, S.C.; Grunwald, G.D.; Balaban, A.T.TRIPLET: Copyright of the Regents of the University of Minnesota, 1993

[14] Basak, S. C.; Grunwald, G. D., APProbe. 1993; Copyright of the University of Minnesota

[15] Basak, S. C.; Grunwald, G. D., A COMPARATIVE STUDY OF GRAPH INVARIANTS, TOTAL SURFACE AREA AND VOLUME IN PREDICTING BOILING POINTS OF ALKANES, Math Modelling & Sci. Computing, 1993, 2, 735-740.

[16] Basak, S. C; Mills, D. Quantitative structure-activity relationships for cycloguanil analogs as PfDHFR inhibitors using mathematical molecular descriptors. SAR and QSAR in Environmental Research, 2010, 21, 215–229.

SciForum
Mol2Net

Pharmacokinetics and Toxicological Profiling of Surfactin A: An In silico Approach

Rajeev K Singla, Ashok K Dubey *

Division of Biotechnology, Netaji Subhas Institute of Technology, Sec-3, Dwarka, New Delhi-110078, India. E-Mail: rajeevsingla26@gmail.com

* Author to whom correspondence should be addressed; Ashok K Dubey, Division of Biotechnology, Netaji Subhas Institute of Technology, Sec-3, Dwarka, New Delhi-110078, India. E-Mail: adubey.nsit@gmail.com

Published: 4 December 2015

Abstract: Surfactin A, a cyclic lipopeptide from *Bacillus subtilis,* exhibited a wide spectrum therapeutic profile. But it's drug likeness has not been thoroughly assessed yet. Thus, the objective of the present work was to simulate it's drug likeness by predicting the pharmacokinetic and toxicological profiling parameters. ADME profiling was carried out by using StarDrop. Integrated Derek Nexus with StarDrop was used for the toxicological prediction against 40 toxicological end points. Metabolism of Surfactin A was modelled with three isoforms of cytochrome P450: 3A4, 2D6 and 2C9; and composite site lability (CSL) was analyzed. Only the alkyl regions in Surfactin A were found to be moderately labile to metabolism, indicating their tendency to get oxidized and form dealkylated Surfactin A. Toxicological prediction suggested that Surfactin A is not carcinogenic, mutagenic, teratogenic, hepatotoxic, neurotoxic, nephrotoxic and even clean for rest of the toxicological end points. Good human intestinal absorption (HIA) and poor blood brain barrier (BBB) crossing ability were also predicted for Surfactin A.

Keywords: Surfactin A; Cyclic Lipopeptide; ADMET; Drug Likeness; *Bacillus subtilis*

1. Introduction

Biosurfactants have potential industrial applications, for example, in digestion of persistent organic pollutants (POPs), as stabilizing agents in food processing and foam formation to name a few [1]. *Bacillus subtilis* yields a well known biosurfactant as secondary metabolite, Surfactin which exists in four isomeric forms viz. Surfactin A - D. Physiologically, it acts as anti-fibrin clotting agent, as an agent for cell lysis, anti-diabetic adjuvant, anti-inflammatory agent etc [2-4]. But so far, its drug metabolism, pharmacokinetics and toxicological parameters have not be reported. In this study, we predict these features with the aid of computational tools.

2. Results and Discussion

Composite site lability of 0.9219 predicted the high rate of metabolism with 3A4 isoform of cytochrome P450 (**Table 1**). The CYP3A4 is the main isoform of P450 superfamily which resides majorly in the liver and is responsible for the metabolism of many drugs. Hence, the lability of surfactin with this isoform predicts the significant affinity. Interaction of surfactin A with CYP3A4 may increase/decrease its physiological effects which need to be assessed further. Amphiphillic nature of the surfactin A can be computed from its LogS and LogP values. Aliphatic elements like hydrocarbon side chain influence the hydrophobicity positively while the peptidal bonds have some negative influence on the hydrophobicity (**Figure 1**). Carboxyl group of aspartic acid and glutamic acid improved the solubility of surfactin A. Further, the tendency to get absorbed through human intestinal barrier and the inability to cross blood brain barrier makes it a potential candidate for oral and non-CNS drug.

Derek Nexus software predicted that the surfactin A doesn’t have any potential site which can cause toxicity against any of the 40 tested toxicological end points.

Table 1. Drug Metabolism, Pharmacokinetics and Toxicological Profiling of Surfactin A, a cyclic lipopeptide.

Parameters	Surfactin A	Parameters	Surfactin A	Parameters	Surfactin A
Composite Site Lability against CYP3A4	0.9219	Mutagenicity in vitro	No report	Irritation (of the gastrointestinal tract)	No report
logS	2.459	Mutagenicity in vivo	No report	Irritation (of the respiratory tract)	No report
logS @ pH7.4	2.903	Photomutagenicity in vitro	No report	Irritation (of the skin)	No report
logP	3.11	alpha-2-mu-Globulin nephropathy	No report	Lachrymation	No report
logD	2.88	Anaphylaxis	No report	HERG channel inhibition in vitro	No report
2C9 pKi	4.858	Bladder urothelial hyperplasia	No report	Hepatotoxicity	No report
hERG pIC50	1.794	Cardiotoxicity	No report	Genotoxicity in vitro	No report
BBB log([brain]:[blood])	-0.6902	Cerebral oedema	No report	Genotoxicity in vivo	No report
BBB category	-	Chloracne	No report	Photogenotoxicit	No

				y in vitro	report
HIA category	+	Cholinesterase inhibition	No report	Photogenotoxicity in vivo	No report
P-gp category	yes	Cumulative effect on white cell count and immunology	No report	Chromosome damage in vitro	No report
2D6 affinity category	high	Cyanide-type effects	No report	Chromosome damage in vivo	No report
PPB90 category	high	High acute toxicity	No report	Photo-induced chromosome damage in vitro	No report
Developmental tox. category	Non-toxic	Methaemoglobinaemia	No report	Carcinogenicity	No report
Thyroid toxicity	No report	Nephrotoxicity	No report	Photocarcinogenicity	No report
Photoallergenicity	No report	Neurotoxicity	No report	Pulmonary toxicity	No report
Skin sensitisation	No report	Oestrogenicity	No report	Uncoupler of oxidative phosphorylation	No report
Occupational asthma	No report	Peroxisome proliferation	No report	Irritation (of the eye)	No report
Respiratory sensitisation	No report	Phospholipidosis	No report	Testicular toxicity	No report
Developmental toxicity	No report	Phototoxicity	No report	Ocular toxicity	No report

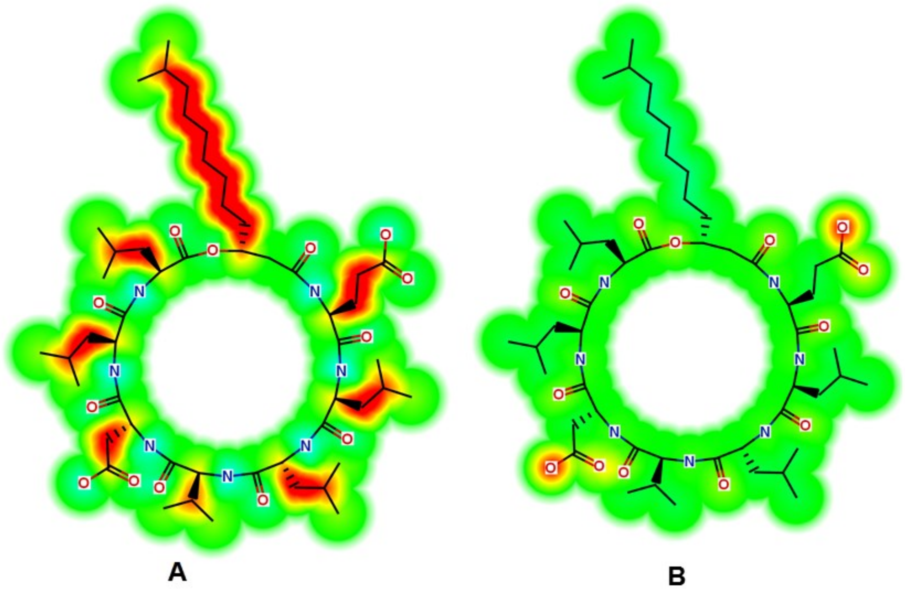


Figure 1. Responsible Structural Elements of Surfactin A. A: LogP and B: LogS @ pH 7.4. Green section: Neutral Region; Red section: Positively Influencing Elements; Yellow Section: Intermediary Infleuncing Elements; Blue Section: Negatively Influencing Elements.

3. Materials and Methods

Drug Metabolism & Pharmacokinetics

Drug metabolism and pharmacokinetics parameters were predicted using StarDrop (Optibrium Ltd., United Kingdom). Parameters studied were LogS, LogS@pH7.4, LogP, LogD, 2C9 pKi, hERG pIC50, BBB Log ([brain]:[blood]), BBB category, HIA category, P-gp category, 2D6 affinity category, PPB90 category, developmental toxicological category and composite site lability of these molecules on 3A4 isoform of cytochrome P450 [5-6].

Toxicological Studies

Derek Nexus module of LHASA Ltd. was used to calculate toxicological endpoints like carcinogenicity, photo-carcinogenicity, chromosome damage in vitro, chromosome damage in vivo, photo-induced chromosome damage in vitro, genotoxicity in vitro, genotoxicity in vivo, photogenotoxicity in vitro, photogenotoxicity in vivo, hepatotoxicity, irritation (of the eye), irritation (of the gastrointestinal tract), irritation (of the respiratory tract), irritation (of the skin), lachrymation, HERG channel inhibition in vitro, alpha-2-mu-globulin nephropathy, anaphylaxis, bladder urothelial hyperplasia, cardiotoxicity, cerebral oedema, chloracne, cholinesterase inhibition, cumulative effect on white cell count and immunology, cyanide-type effects, high acute toxicity, methaemoglobinaemia, nephrotoxicity, neurotoxicity, oestrogenicity, peroxisome proliferation, phospholipidosis, phototoxicity, pulmonary toxicity, uncoupler of oxidative phosphorylation, developmental toxicity, teratogenicity, testicular toxicity, ocular toxicity, mutagenicity in vitro, mutagenicity in vivo, photomutagenicity in vitro, thyroid toxicity, photoallergenicity, skin sensitization, occupational asthma and respiratory sensitization [5-6].

4. Conclusions

In this study, prediction of the drug metabolism, pharmacokinetic, and toxicological parameters were investigated for the well known biosurfactant, Surfactin A produced by *Bacillus subtilis*. Computataional tools predicted Surfactin A as a molecule, which can be studied and explored further for lead optimization.

Acknowledgments

Authors would like to express their thanks to Optibrium Ltd (U. K.) for providing trial version of StarDrop and Derek Nexus for the current study.

Author Contributions

RKS collected the data and prepared initial draft. AKD initialize the data, analyzed, and finalize the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Chen, W-C.; Juang, R-S.; Wei, Y-H. Applications of a lipopeptide biosurfactant, surfactin, produced by microorganisms. *Biochemical Engineering Journal* **2015**, *103*, 158-169.
2. Singla, R.K.; Dubey, H.D.; Dubey, A.K. Therapeutic spectrum of bacterial metabolites. *Indo Global Journal of Pharmaceutical Sciences* **2014**, *4*, 52-64.
3. Okumura, K.; Iwakawa, S.; Yoshida, T.; Seki, T.; Komada, F. Intratracheal delivery of insulin absorption from solution and aerosol by rat lung. *International Journal of Pharmceutics* **1992**, *88*, 63-73.
4. Tang, J-S.; Zhao, F.; Gao, H.; Dai, Y.; Yao, Z-H.; Hong, K.; Li, J.; Ye, W-C.; Yao, X-S. Characterization and online detection of surfactin isomers based on HPLC-MS^n analyses and their inhibitory effects on the overproduction of nitric oxide and the release of TNF-α and IL-6 in LPS induced macrophages. *Marine Drugs* **2010**, *8*, 2605-2618.
5. Aggarwal, B.; Singla, R.K.; Ali, M.; Singh, V.; Igoli, J.O.; Gundamaraju, R.; Kim, K.H. Triterpenic and monoterpenic esters from stems of *Ichnocarpus frutescens* and their drug likeness potential. *Medicinal Chemistry Research* **2015**, *24*, 1427-1437.
6. Singla, R.K. In silico Drug Design & Medicinal Chemistry. *Current Topics in Medicinal Chemistry*, **2015**, *15*, 971-972.

SciForum
Mol2Net

Do We Use Well Benzodiazepines in Elderly? a Case Report

Maria José Díaz Gutiérrez
Community pharmacist in a the Pharmacy office of Ines Barrenetxea, Alango street 7.in Getxo 48992, Spain; E-Mail: marijo72@euskaltel.ent

Published: 4 December 2015

Abstract:

A 75 year old man comes to the pharmacy to pick up the medications n prescribed after a 2 week hospitalization period due to a fall at his habitual residence, with the result of a broken femur. The patient lives in a nursing home where staff prepare his medication in customized dispensing systems. The unique new prescribed drug is paracetamol (1g) only if pain appears and with maximum dose of 3g per day. Tha patient daily consumes:

Mirtazapine 30 mg
Escitalopram 15 mg
Ketazolam 30 mg
Lorazepam 5 mg
Dutasteride 0.5 mg
Omeprazol 20 mg

We note that for their anxious-depressive symptoms he consumes two benzodiazepines at higher than recommended dose for his age along with two antidepressants, one of which has high doses sedative effect (mirtazapine). We do not know how long he has been taken with all these drugs but it refers than more than four months.

Because it is an retrospective evaluation we cannot establish a causal relationship of treatment with the fall, but we may suspect that the fall was triggered by an overdose of benzodiazepines. We get in touch with the doctor of the nursing home e to discuss the case who decides to withdraw ketazolam treatment and subsequently valued reduction in the dose of mirtazapine according to the patient's response.

The elderly population is a special risk group for drug adverse events, due to factors such as changes in pharmacokinetic and pharmacodynamic processes, with frequent presence of multiple pathologies and polypharmacy.

We must to remember the importance of the review of the dose and duration of treatment with benzodiazepines in the elderly and follow the recommendations of clinical guidelines for selecting those with short or ultra-short BZD of life, at the lowest possible dose for the shortest time.

Keywords: benzodizepines, elderly, overdose, accidental falls

1. Introduction

The beneficial effects of BZDs are often disputed and concerns expressed about their adverse events and high rates of prescription in older adults (1,2). Certainly, prescription decisions have to be made on a case-by-case basis and patients should be informed of both the risks and benefits of any prescribed medication (3).

There is an age-related increase in the rate and severity of adverse effects of drugs that act on the central nervous system, which often results from a decrease in the number of neurons and synapses and greater permeability of the blood-brain barrier (4). BZDs are one of the most commonly prescribed drugs in older adults because of their proven efficacy, but care must be taken as their use and abuse may lead to unwanted effects, including cognitive deterioration (5), motor incoordination, ataxia, falls (6-8) and respiratory failure (9-11).

2. Results and Discussion

A 75 year old man comes to the pharmacy to pick up the medications n prescribed after a 2 week hospitalization period due to a fall at his habitual residence, with the result of a broken femur. The patient lives in a nursing home where staff prepare his medication in customized dispensing systems. The unique new prescribed drug is paracetamol (1g) only if pain appears and with maximum dose of 3g per day. Table 1 shows the drugs taken by patients daily.

We note that for their anxious-depressive symptoms he consumes two benzodiazepines at higher than recommended dose for his age along with two antidepressants, one of which has high doses sedative effect (mirtazapine). We do not know how long he has been taken with all these drugs but it refers than more than four months..

Because it is an retrospective evaluation we can not establish a causal relationship of treatment with the fall, but we may suspect that the fall was triggered by an overdose of benzodiazepines. We get in touch with the doctor of the nursing home e to discuss the case who decides to withdraw ketazolam treatment and subsequently valued reduction in the dose of mirtazapine according to the patient's response.

Table 1. Situation before pharmaceutical intervention

Health problem	**Drug**	**Unit dose**	**Prescription**	**Total Daily dose**
Anxiety-depressive syndrome	Mirtazapine	30 mg	0-0-1	30 mg
Anxiety-depressive syndrome	Escitalopram	15mg	1-0-0	15 mg
Anxiety-depressive syndrome	Ketazolam	30 mg	0-0-1	30 mg
Insomnia	Lorazepam	1 mg	2-1-2	5 mg
Prostatic syndrome	Dutasteride	0.5 mg	1-0-0	0.5 mg
Gastric dyspepsia	Omeprazole	20 mg	1-0-0	20 mg

3. Materials and Methods

.

4. Conclusions

The elderly population is a special risk group for drug adverse events, due to factors such as changes in pharmacokinetic and pharmacodynamic processes, with frequent presence of multiple pathologies and polypharmacy.

We mus to remember the importance of the review of the dose and duration of treatment with benzodiazepines in the elderly and follow the recommendations of clinical guidelines for selecting those with short or ultrashort BZD of life, at the lowest possible dose for the shortest time..

Author Contributions

MDG reported the case and wrote this report

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Touitou Y. Sleep disorders and hypnotic agents: medical, social and economical impact. Ann Pharm Fr. 2007 ;65:230–8.
2. Bourin M. The problems with the use of benzodiazepines in elderly patients. L'Encéphale. 2010;36:340–7.

3. Balon R, Fava GA, Rickels K. Need for a realistic appraisal of benzodiazepines. World Psychiatry 2015 ;14:243–4.
4. Oakley R, Tharakan B. Vascular hyperpermeability and aging. Aging Dis 2014;5:114–25.
5. Stewart SA. The effects of benzodiazepines on cognition. J Clin Psychiatry 2005;66 (Suppl 2):9–13.
6. Woolcott JC, Richardson KJ, Wiens MO, Patel B, Marin J, Khan KM, et al. Meta-analysis of the impact of 9 medication classes on falls in elderly persons. Arch Intern Med 2009;169:1952–60.
7. Ungar A, Rafanelli M, Iacomelli I, Brunetti MA, Ceccofiglio A, Tesi F, et al. Fall prevention in the elderly. Clin Cases Miner Bone Metab 2013;10:91–5.
8. Huang AR, Mallet L, Rochefort CM, Eguale T, Buckeridge DL, Tamblyn R. Medication-related falls in the elderly: causative factors and preventive strategies. Drugs Aging 2012;29:359–76.
9. Gueye PN, Lofaso F, Borron SW, Mellerio F, Vicaut E, Harf A, et al. Mechanism of respiratory insufficiency in pure or mixed drug-induced coma involving benzodiazepines. J Toxicol Clin Toxicol 2002;40:35–47.
10. Kamijo Y, Hayashi I, Nishikawa T, Yoshimura K, Soma K. Pharmacokinetics of the active metabolites of ethyl loflazepate in elderly patients who died of asphyxia associated with benzodiazepine-related toxicity. J Anal Toxicol 2005;29:140–4.
11. Guilleminault C. Benzodiazepines, breathing, and sleep. Am J Med 1990;88:25S – 28S.

SciForum
Mol2Net

Machine Learning and Atom-Based Quadratic Indices for Proteasome Inhibition Prediction

Gerardo M. Casañola Martin,[1,2,3]* Huong Le-Thi-Thu,[4] Facundo Perez-Gimenez,[2] and Concepción Abad[1]

[1] Departament de Bioquímica i Biologia Molecular, Universitat de València, E-46100 Burjassot, Spain; emails: gerardo.casanola@uv.es (G.M.C.M) ; concepción.abad@uv.es (C.A)

[2] Unidad de Investigación de Diseño de Fármacos y Conectividad Molecular, Departamento de Química Física, Facultad de Farmacia, Universitat de València, Spain. emails: gerardo.casanola@uv.es (G.M.C.M); facundo.perez@uv.es (F.P.G)

[3] Universidad Estatal Amazónica, Facultad de Ingeniería Ambiental, Paso lateral km 2 1/2 via Napo, Puyo, Ecuador gcasanola@uea.edu.ec (G.M.C.M)

[4] School of Medicine and Pharmacy, Vietnam National University, Hanoi (VNU) 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam ltthuong1017@gmail.com (H.L.T.T)

* Author to whom correspondence should be addressed; E-Mail: gerardo.casanola@uv.es ; Tel.: +34-963543156

Published: 4 December 2015

Abstract: The atom-based quadratic indices are used in this work together with some machine learning techniques that includes: support vector machine, artificial neural network, random forest and k-nearest neighbor. This methodology is used for the development of two quantitative structure-activity relationship (QSAR) studies for the prediction of proteasome inhibition. A first set consisting of active and non-active classes was predicted with model performances above 85% and 80% in training and validation series, respectively. These results provided new approaches on proteasome inhibitor identification encouraged by virtual screenings procedures.
.

Keywords: Atom-based quadratic index, classification and regression model, machine learning, proteasome inhibition, QSAR, TOMOCOMD-CARDD software

Mol2Net YouTube channel*: http://bit.do/mol2net-tube*

1. Introduction

The ubiquitin-proteasome pathway (UPP) is responsible for the selective degradation of the majority of the intracellular proteins in eukaryotic cells and regulates nearly all cellular processes [1]. Disfunction of the ubiquitination machinery or the proteolytic activity of the proteasome is associated with many human diseases [2]. Proteasome inhibitors have been developed being effective for some disorders but sometimes show detrimental effects and resistance. Therefore, efforts are currently directed to the development of new therapeutics with adequated potency and safety properties that target enzyme components of the UPP [3,4].

Ligand-based molecular design and QSAR approaches are promising fields with several applications in drug development, which use a battery of novel molecular descriptors and different classification algorithms for in silico virtual drug screening studies [5,6]. In the present research, we use and compare a set of different machine learning (ML) techniques using the 2D atom-based quadratic indices as attributes with the objective to perform the QSAR modeling of two datasets. The first dataset allows to separate molecules with proteasome inhibitory activity from inactive ones, and the second provides the numerical prediction of the EC_{50}.

2. Results and Discussion

In the case of our classification study, we reduced the inactive subset removing all the cases that fall outside of the applicability domain of our model. Therefore, the dataset remains with 705 chemicals, being 258 active and the rest 447 inactive ones. The first 705 dataset used for classification studies generates 529 in the training set (TS) and 176 compounds in the prediction set (PS). Based on the aspects mentioned above for our case a first step with non-supervised feature reduction filtering was done, by using the Shannon´s entropy as a measure keeping c.a. the 30% of the features (4 143). In a second step a supervised feature reduction filtering was done. In this stage, the process was carried out for the class problem. In this case the features were reduced a 70%, keeping a total of 1248 for the class data. These feature selection processes were carried out with the IMMAN software an “in house” program. Later, in the two-class data the best subset search was done resulting in 43 selected variables. Then wrapper methods associated with the ML techniques were applied to reduce data sets giving different data subsets combinations. Finally, all these subsets were used to generate diverse ML-QSAR models keeping those with the best results for each algorithm. The results for each ML technique used to develop classification QSAR models to predict proteasome inhibitors are shown in Fig. 1.

As it can be observed in Fig. 1 for the TS the fitted models using RF and MLP techniques showed the best accuracies (Ac = 90.17% and Ac = 89.22%) with Mathew´s correlation coefficient (MCC) values of 0.79 and 0.77, respectively. In the case of the PS, the performance of these two QSAR models was of 86.36% (MCC=0.70) and 83.52% (MCC=0.64), respectively. Moreover, can be observed low values of false positive rates, which ensures a good performance at time to perform virtual high-throughput screenings, disminissing the wrong evaluation of predicted positive cases. In the same Fig. 1 can also be noted that RF outperforms other models in most of the quality parameters. Besides, the rest of the models also depicted adequate performances with accuracies values above 85% in the case of the TS and 80 % for the PS.

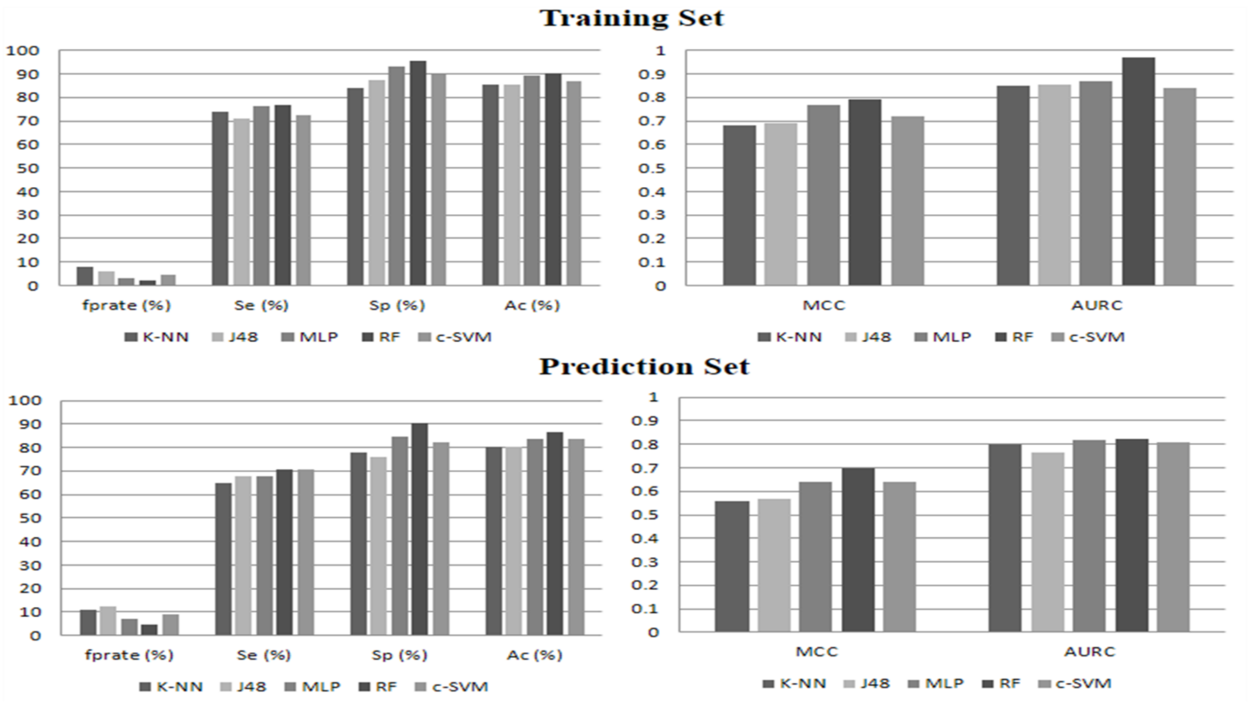


Figure 1. Performance of the ML-based QSAR classifiers

3. Materials and Methods

In this study the molecular descriptors atom-based quadratic indices were calculated using the TOMOCOMD software version 1.0 [7]. We also attempt the different feature selection methods implemented in the IMMAN software [8]. Moreover, the attribute selection method based on BestSubset Search (BSS) of LDA discriminant analysis was used [9]. Later, the wrapper and ranker methods of Waikato environment for knowledge analysis (WEKA) [10] were considered. As a final stage, the parameter tuning optimization for each ML technique was performed to find the best ML-QSAR models.

A dataset derived from a luminescent cell-based dose titration retest counterscreen assay to identify inhibitors of the proteasome pathway was selected from PubChem BioAssay (AID 2486) where the name, structures, compound identifier (CID), and activities can be found. First, a curation process on the database was assessed removing salts, and inorganic compounds. The main difficulty of the ML approaches is to select attributes from a large list of candidates to describe the data. This is because the complete set of molecular descriptors is not needed for the description of the proteasome inhibition. In this sense, the addition of non-relevant attributes can cause noise to the ML systems [10]. Therefore, the feature selection approaches are very suitable to deal with this kind of problem. In this work, different schemes of attribute selection including filter and wrapper approaches implemented in WEKA [10] are examined to select the best attribute subset for each ML technique. Some details, advantages and drawbacks of the two approaches can be reviewed in many works dealing with this subject [11-13].

The machine learning methods shows impressive performances a wide diversity of studies involving automated, text classification and drug design [14-16]. Based on this the machine learning approaches selected were: support vector machine, artificial neural network and k-nearest neighbor also included in the list of the top ten algorithms used in data mining [17]. Besides the random forest technique was included because is fast and robust approach with recent succesfull application into many problems [18-20]. For each ML method applied in this study, various schemes of selecting attributes were examined and for each selected subset, various models were developed and checked out.

4. Conclusions

In this work, a QSAR study on a diverse and enlarged proteasome inhibitor database collected from the PubChem Bioassay is shown for the first time. The random forest algorithm demonstrates to be the best technique for the modeling of the proteasome inhibitory activity with high accuracies values in the training and test set. The low false positive rates observed validates the presented workflow based on ML-QSAR for the prediction of active proteasome inhibitors compounds from inactive ones.

Acknowledgments

Casañola-Martin. G.M. and Castillo-Garit. J.A thank the program 'Estades Temporals per a Investigadors Convidats' for a fellowship to research University (2013-2014) at Valencia. Marrero-Ponce, Y. thanks to the program 'International Professor' for a fellowship to work at Cartagena University in 2013-2014. Le-Thi-Thu, H. gratefully acknowledge support from the National Vietnam National University, Hanoi

Conflicts of Interest

"The authors declare no conflict of interest".

References and Notes

1. Varshavsky, A. The ubiquitin system, an immense realm. *Annu. Rev. Biochem* **2012**, *81*, 167-176.
2. Rastogi, N.; Mishra, D.P. Therapeutic targeting of cancer cell cycle using proteasome inhibitors. *Cell Division* **2012**, *7*, 26.
3. de Bettignies, G.; Coux, O. Proteasome inhibitors: Dozens of molecules and still counting. *Biochimie* **2010**, *92*, 1530-1545.
4. Pevzner, Y.; Metcalf, R.; Kantor, M.; Sagaro, D.; Daniel, K. Recent advances in proteasome inhibitor discovery. *Expert Opinion on Drug Discovery* **2013**, *8*, 537-568.
5. Rescigno, A.; Casañola-Martin, G.M.; Sanjust, E.; Zucca, P.; Marrero-Ponce, Y. Vanilloid derivatives as tyrosinase inhibitors driven by virtual screening-based qsar models. *Drug Test Anal* **2011**, *3*, 176-181.
6. Kumar, D.; Kapoor, A.; Thangadurai, A.; Kumar, P.; Narasimhan, B. Synthesis, antimicrobial evaluation and qsar studies of 3-ethoxy-4-hydroxybenzylidene/4-nitrobenzylidene hydrazides. *Chin. Chem. Lett* **2011**, *22*, 1293-1296.
7. Marrero-Ponce, Y.; Valdés-Martini, J.R.; García Jacas, C.R. *Tomocomd-cardd qubils software qubils-mas. Version 1.0*, CAMD-BIR Unit, Universidad Central "Marta Abreu" de Las Villas, 2012.
8. Barigye, S.J.; Pino Urias, R.W.; Marrero-Ponce, Y. *Imman (information theory based chemometric analysis) version 1.0.*, 2011.
9. *Statistica (data analysis software system) vs 6.0*, StatSoft Inc: Tulsa,OK:, 2001.

10. Witten, I.H.; Frank, E. *Data mining: Practical machine learning tools and techniques*. 2nd ed. ed.; Morgan Kaufmann: Burlington, MA, 2005.
11. Ben Meskina, S. In *On the effect of data reduction on classification accuracy*, 2013.
12. Shahlaei, M. Descriptor selection methods in quantitative structure-activity relationship studies: A review study. *Chemical Reviews* **2013**, *113*, 8093-8103.
13. Inza, I.; Larrañaga, P.; Blanco, R.; Cerrolaza, A.J. Filter versus wrapper gene selection approaches in DNA microarray domains. *Artificial Intelligence in Medicine* **2004**, *31*, 91-103.
14. Baumes, L.A.; Ranilla, J. A study on factors affecting the reproducibility of a chemical tongue analysis responding to amino acids. *Combinatorial Chemistry and High Throughput Screening* **2013**, *16*, 572-583.
15. Gertrudes, J.C.; Maltarollo, V.G.; Silva, R.A.; Oliveira, P.R.; Honório, K.M.; Da Silva, A.B.F. Machine learning techniques and drug design. *Current Medicinal Chemistry* **2012**, *19*, 4289-4297.
16. Le-Thi-Thu, H.; Marrero-Ponce, Y.; Casañola-Martin, G.M.; Cardoso, G.C.; Chávez, M.D.C.; Garcia, M.M.; Morell, C.; Torrens, F.; Abad, C. A comparative study of nonlinear machine learning for the "in silico" depiction of tyrosinase inhibitory activity from molecular structure. *Molecular Informatics* **2011**, *30*, 527-537.
17. Wu, X.; Kumar, V.; Ross, Q.J.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.J.; Ng, A.; Liu, B.; Yu, P.S.*, et al.* Top 10 algorithms in data mining. *Knowledge and Information Systems* **2008**, *14*, 1-37.
18. Ziegler, A.; König, I.R. Mining data with random forests: Current options for real-world applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2014**, *4*, 55-63.
19. Verikas, A.; Gelzinis, A.; Bacauskiene, M. Mining data with random forests: A survey and results of new tests. *Pattern Recognition* **2011**, *44*, 330-349.
20. Chen, X.; Ishwaran, H. Random forests for genomic data analysis. *Genomics* **2012**, *99*, 323-329.

SciForum
Mol2Net

New Insights from the CoMSIA Analysis within the Framework of Density Functional Theory

Alejandro Morales-Bayuelo * and Julio Caballero

Centro de Bioinformática y Simulación Molecular (CBSM), Universidad de Talca, 2 Norte 685, Casilla 721, Talca, Chile

* Author to whom correspondence should be addressed; E-Mail: alejandr.morales@uandresbello.edu.

Published: 4 December 2015

Abstract: Today, one of the main aims in the pharmaceutical companies is seek new methodologies to understand the biological activity in molecules from the computational point of view. In this sense, understand the traditional tools (3D QSAR) such as the Comparative Molecular Similarity Analysis (CoMSIA) within the quantum chemistry framework, can be relevant. In this context, the quantification of steric and electrostatic effects on a serie of antimalarials chalcones was performed on the basis of the descriptors from the molecular quantum similarity field and chemical reactivity supported in DFT. The steric and electrostatic effects were studied using scales of convergence quantitative alpha (α) and beta (β), respectively. To deal the problem of relative molecular orientation in the quantum similarity field the Topo-Geometrical Superposition Algorithms (TGSA) was used as molecular alignment method. Finally, a chemical reactivity analysis using global and local descriptors such as chemical hardness, softness, electrophilicity, and Fukui Functions was developed.

.

.

Keywords: Antimalarials Chalcones, Comparative Molecular Similarity Indexes Analysis (CoMSIA), 3D-QSAR, Molecular Quantum Similarity (MQS), Chemical Reactivity Descriptors, Density Functional Theory (DFT)

1. Introduction

In a recent publication our researcher group shown as the Comparative Molecular Field Analysis (CoMFA) can be understood in terms of Molecular Quantum Similarity (MQS) and Density Functional Theory (DFT)-based reactivity descriptors **[1]**. The CoMFA analysis have many applications in the three-Dimensional Quantitative Structure-Activity Relationships (3D QSAR) studies, yet this method is commonly associated with the Comparative Molecular Similarity Indexes Analysis (CoMSIA) by this reason in this work is studied the CoMSIA analysis in terms of MQS and chemical reactivity descriptors to search new insights within the DFT framework.

The MQS field was introduced by Carbó and co-workers approximately 35 years ago **[2-5]**, this is a topic which has been widely considered and applied on chemical phenomenon study such as electron delocalization and aromaticity **[6]**, modeling 3D QSAR **[7]**, topological studies **[8]**, among others. The MQS field the main variable is the density function **[9-11]**; of this form can be related with the chemical reactivity descriptors such as chemical hardness (η), softness (S), electrophilicity (ω) and Fukui Functions. Therefore, using this hybrid methodology (joining the MQS and chemical reactivity) we hope show how the CoMSIA results can be related with the DFT context.

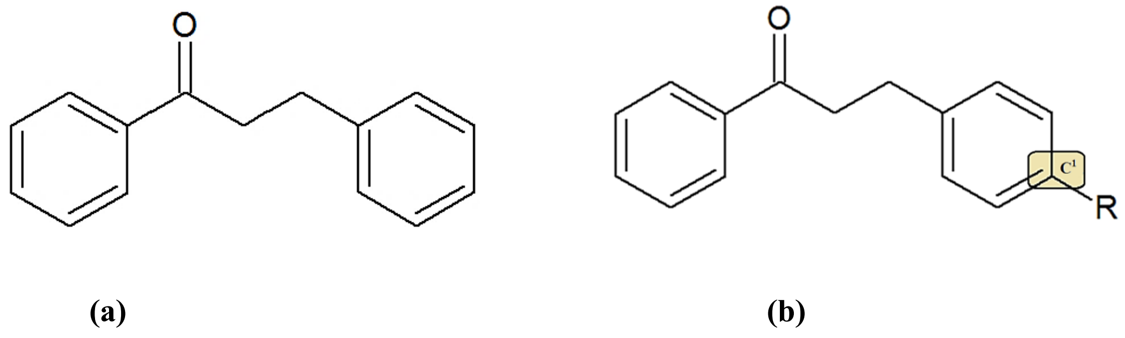
To the carry out these goals, we used the CoMSIA results reported by Xue and co-workers **[12]**. They development a 3D QSAR studies on antimalarial alkoxylated and hydroxylated chalcones by CoMFA and CoMSIA to determine the factors required for the activity of these compounds, this study shown that the CoMSIA analysis presents better physical-chemistry parameters to understand the antimalarial activity using five physical-chemistry properties (steric, electrostatic, hydrophobic, and hydrogen-bond donor or acceptor properties). In line, with this reported we will use the hybrid methodology proposed to modeling and study these CoMSIA outcomes using DFT.

Furthermore, the entropic contributions to the binding affinity are more difficult to describe using the CoMSIA methodology, because a major factor arises from the solvent-to-protein transfer. This portion approximately correlates with the size of the hydrophobic surface area of the drug molecule **[13, 14]**. For these reasons, show new methodologies are relevant in the QSAR field.

2. Molecular Set

A series of chalcones studied by Xue and co-workers **[12]** were used in this study. The biological activity IC_{50} values μM (for inhibition of [3H] hypoxanthine uptake into *P. falciparum* (K1) in the presence of drug) were expressed as pIC_{50} (the –log IC_{50}), these biological values were reported by Liu and co-workers **[15]**, the theoretical values from the CoMSIA method are shown in **Table 1 [12]**.

Table 1. Compounds, biological activities and theoretical predictions from the CoMSIA method in the molecular set **[12]**.



(a) **(b)**

Figure 1. (a) Molecular recognition skeleton (compound 1) used for the molecular alignment and **(b)** Local structural differences (to the substituent effect analysis).

Compound	R	pIC$_{50}$ [b]	CoMSIA [c]	
			Pred.	Δ (error)
1	H [a]	4,80	4,88	0,08
2	chloro	4,84	4,78	-0,06
3	nitro	4,65	4,35	-0,30
4	phenyl	4,58	4,64	0,06
5	fluoro	5,02	4,78	-0,24
6	methoxy	4,60	4,88	0,28
7	quinolinyl	5,70	5,70	0,00
8	ethyl	4,78	4,86	0,08
9	methyl	4,59	4,90	0,31
10	trifluoromethyl	5,52	5,29	-0,23
11	dimethylamino	4,74	4,70	-0,04

[a] Reference compound

[b] Experimental values reported by Liu and co-workers **[15]**.

[c] Theoretical predictions from the CoMSIA method **[12]**.

The CoMSIA method on the molecular set studied is calculated at the intersections of a regularly spaced lattice (1.1 and 2 Å spacing), the similarity indices $A_{F,k}$ between the compounds of interest and a probe atom have been calculated according to:

$$A^{q}_{F,k}(j)=\sum\nolimits_{i} w_{probe,k}\, w_{ik}\, e^{-\alpha r^{2}_{kq}} \qquad \textbf{(1)}$$

Where A is the similarity index at grid point q, summed over all atoms i of the molecule j under investigation; $w_{probe,k}$ probe atom with charge +1, radius 1 Å, hydrophobicity +1, H-bond donor and acceptor property +1; α: attenuation factor; r_{iq}: mutual distance between probe atom at grid point q and atom i of the test molecule **[13]**.

Analysing the equation 1 is possible see the Gaussian function behavior, large values of α will result in a strong attenuation of the distance-dependent consideration of molecular similarity (low global similarity in its neighborhood). The opposite effects (reducing α) means that also the remote parts of each molecule will be experienced by the probe atom (high global similarity in its neighborhood).

3. Theory and Computational Details.

3.1 Molecular Quantum Similarity Indexes.

With the main aim fixed in study the CoMSIA results from the DFT framework, we used the Molecular Quantum Similarity Measures (MQSM). A general definition of MQSM has been made in various papers **[16-19]** the quantum similarity measure Z_{AB} between compounds A and B, with electron density $\rho_A(r_1)$ and $\rho_B(r_2)$ respectively, can be studied on the idea of the minimizing of the expression for the Euclidean distance as:

$$D_{AB}=\left(\int\left|\rho_A(r)-\rho_B(r)\right|^2 dr\right)^{1/2}=\left(\int(\rho_A(r_1))^2dr_1+\int(\rho_B(r_2))^2dr_2-2\int\rho_A(r_1)\rho_B(r_2)dr_1dr_2\right)^{1/2}$$
$$=\sqrt{Z_{AA}+Z_{BB}-2Z_{AB}} \quad \textbf{(2)}$$

Where Z_{AB} is the overlap integral between the electron density of the compound A and B, Z_{AA} and Z_{BB} are the self-similarity of compounds A and B **[20]**.

In this researcher we have used the Carbó index due to that is very used in the quantum similarity context **[16-20]**:

$$I_{AB}=\frac{\int\int\rho_A(r_1)\rho_B(r_2)dr_1dr_2}{\sqrt{\left(\int\rho_A(r_1)dr_1\right)^2\left(\int(\rho_B(r_2))^2dr_2\right)^2}} \quad \textbf{(3)}$$

The main structural difference on the compounds studied (see **Figure 1**) is the carbon atom (C^1), therefore the similarity features can be associated from the local point of view, in this order of ideas is used the Hirshfeld approach to study the local quantum similarity

One of the more useful methods to partitioning of electron density in DFT is the Hirshfeld approach **[21]**. This approach is based on partitioning of electron density $\rho(r)$ in contributions $\rho_{C^1}(r)$. These contributions allow define a concept of atom in a reference system and study its (dis)similarity on a molecular set (i.e.; substituent effect analysis). On the other hand, these contributions are proportional to the weight $w_C(r)$ of the electron density of the isolated compound in the so-called *promolecular density* **[22,23]**. The promolecular density is defined as:

$$\rho_{C^1}^{\mathrm{Pr}om}(r)=\sum_x\rho_x^0(r) \quad \textbf{(4)}$$

To calculate the contribution of carbon atom (C) in the electron density in a molecule A $\rho_A(r)$ is according to:

$$\rho_{C^1}(r)=w_{C^1}(r)\rho_A(r) \quad \textbf{(5)}$$

In this form, the weight ($w_C(r)$) is obtained as:

$$w_{C^1}(r)=\frac{\rho_{C^1}^0}{\sum_x\rho_x^0(r)} \quad \textbf{(6)}$$

Here $\rho_{C^1}^0(r)$ is the electron density of the isolated carbon atom C^1, (i.e.; the reference electron density) **[24]**. In this sense, the contribution atomic of other carbon atom (C^2) in a molecule B is obtained as:

$$\rho_{C^2,B}(r)=w_{C^2}(r)\rho_B(r) \quad \textbf{(7)}$$

with

$$w_{C^2,B}=\frac{\rho_{C^2,B}^0(r)}{\sum_x\rho_x^0(r)} \quad \textbf{(8)}$$

So we can write the contribution of the asymmetric carbon atom products $\rho_A(r)\rho_B(r)$ as:

$$\rho_{C,AB}(r)=w_{C,AB}(r)\rho_A(r)\rho_B(r) \quad \textbf{(9)}$$

Using the equations (4-9) we can express the numerator Z_{AB} in the Carbó index (equation 3) as:

$$Z_{A,B}^{Local,C} = \frac{Z_{AB}}{\sqrt{Z_{AA}Z_{BB}}} = \frac{\iint w_{C,AB}\rho_A(r)\rho_B(r)dr_A dr_B}{\sqrt{\left(\int w_{C,A}(r)\rho_A(r)dr_A\right)^2 \left(\int w_{C,B}(r)\rho_B(r)dr_B\right)^2}} \qquad \textbf{(10)}$$

where we can write the global index (equation 3) as local contributions. In this context, using these equations we hope study the local similarity and the substituent effects on the reference carbon atom C^1 (see **Figure 1**).

3.2 Reactivity descriptors in the DFT framework.

Due to the fundamental variable in the MQS field is the electron density naturally there is a relationship between MQS and chemical reactivity, moreover the key feature of quantum similarity lies in the use of the electron density of a molecule. From the DFT point of view physical-chemistry properties such as electrostatic, hydrophobic and hydrogen-bond donor or acceptor properties can be related with global chemical descriptors as chemical potential, hardness, electrophilicity index and local reactivity descriptors as the Fukui Functions.

The chemical potential (μ) can be understood as the tendency that have the electrons to exit of the electron cloud and is calculate according to the equation:

$$\mu \approx \frac{\varepsilon_H + \varepsilon_L}{2} \qquad \textbf{(11)}$$

Where (εH) is the energy of the (HOMO) and (εL) is the energy of the (LUMO) **[25, 26]**. The chemical hardness is defined using the equation (11) according to Pearson et. al. **[27]** and is understood as the opposition to distort the electron cloud of the system according to the equation:

$$\eta \approx \varepsilon_L - \varepsilon_H \qquad \textbf{(12)}$$

Using the equation (12), we obtain the softness **[28]** as:

$$S = \frac{1}{\eta} \qquad \textbf{(13)}$$

Finally, using the equations 11 and 12 is obtaining the electrophilicity index (ω) **[29, 30]**. This index is understood as the measure of the stabilization energy of the system when it is saturated by electrons from the external environment and is calculated as follows:

$$\omega = \frac{\mu^2}{2\eta} \qquad \textbf{(14)}$$

The quantities defined in equations (11-14) are called global reactivity indexes and provide information about the reactivity or stability of a chemical system front to external perturbations. To study the chemical reactivity from the local point of view are used the Fukui Functions. The Fukui Functions (equation 15 and 16, $f(r)$) are defined as the derivative of the electronic density with respect to the number of electrons at constant external potential:

$$f_k^+ \approx \int_k \left[\rho_{N+1}(\vec{r}) - \rho_N(\vec{r})\right] = \left[q_k(N+1) - q_k(N)\right] \qquad \textbf{(15)}$$

$$f_k^- \approx \int_k \left[\rho_N(\vec{r}) - \rho_{N-1}(\vec{r})\right] = \left[q_k(N) - q_k(N-1)\right] \quad \textbf{(16)}$$

Where q_k refers to the electron population at k^{th} atomic site in a molecule. Here, we adopted natural population analysis (NPA) scheme to evaluate atomic charge. (f_k^+) governing the susceptibility for nucleophilic attack and (f_k^-) governing the susceptibility for electrophilic attack **[31-33]**.

In this sense, using these global and local reactivity schemes is possible study the selectivity and substituent effect on the molecular set from DFT framework.

3.3 Alignment Method and Computational details.

Similar to the CoMSIA method the MQS also need an optimal alignment methodology, to deal with the problem of the relative molecular alignment is used the Topo-Geometrical Superposition Algorithm (TGSA) **[34]**. This alignment method tries to overlap as many structural elements as possible. These structural elements correspond to chemical bonds and sequences of two chemical bonds, always involving the same type of atoms in both molecules compared **[35-37]**. All the compounds were optimized using B3LYP exchange-correlation functional **[38(a,b)]** at 6-31G(d,p) level of theory. All the optimizations were carried out using Gaussian 09 **[39]**.

Using the Dirac delta distribution $\Omega(r_1,r_2)=\delta(r_1,r_2)$ **[40]** is possible define the so called overlap molecular quantum similarity measure and relates the volume associated with the overlap of the two densities $\rho_A(r)$ and $\rho_B(r)$ according to the equation:

$$Z_{AB}(\Omega)=\iint \rho_A(r_1)\delta(r_1-r_2)\rho_B(r_2)dr_1dr_2=\int \rho_A(r)\rho_B(r)dr \quad \textbf{(17)}$$

Equation 17 provides the information about the electron concentration in the molecule and indicates the degree of overlap between the compared compounds.

When the $\Omega(r_1,r_2)$ operator is the coulomb operator $\Omega(r_1,r_2)=|r_1-r_2|^{-1}$ it represents the electronic coulomb repulsion energy between molecular densities $\rho_A(r)$ and $\rho_B(r)$ as:

$$Z_{AB}(\Omega)=\iint \rho_A(r_1)\frac{1}{|r_1-r_2|}\rho_B(r_2)dr_1dr_2 \quad \textbf{(18)}$$

Using these operators (equations 17 and 18) we calculate the local quantum similarity through the equation 10. In this sense, the Carbó index is restricted to the range (0,1) where C_{AB}=0 means dis(similarity) and C_{AB}=1 self-similarity, according to the Schwartz integral.

$$\left[\int \rho_A(r)\rho_B(r)dr\right]^2 \le \int \rho_A^2(r)dr\int \rho_B^2(r)dr \quad \textbf{(19)}$$

Using these methodologies we hope study the substituent effects in the molecular set and shows news insight on the selectivity and chemical reactivity of these antimalarial chalcones.

4. Results and Discussion

The CoMSIA method is a very reliable method for study the structure-activity trend within biological sets. It is a statistic approach that seeks to correlate relative differences in molecular descriptor values to a dependent property (e.g.; the binding affinity). However, the complexity and complications to understand this 3D QSAR results are increasing. One of the forms to deal these problems can be using the Quantum Similarity field and reactivity descriptors supported on DFT.

In this context, in **Table 2** are shows the local molecular quantum similarity indexes using the operator of overlap (17) and the equation 10. These measures can be related with the steric effects along the molecular set.

The highest values in the local similarity of overlap is between compounds 2 and 9 (0,991) with an euclidean distance of (0,450, see **Table 3**) while the lowest value is between the compounds **7** and **10** (0,682) with an euclidean distance of 3,523. The diagonal corresponds to the self-similarity according to the range of the Carbó index, the main difference between the Carbó indexes and the euclidean distances is that these last can take values from zero to infinity (0,∞). To understand these trends in the molecular set with respect to the reference compound 1 are used the scales of convergence quantitative alpha (α) to steric effects using the **Tables 2** and **3**, respectively (see **Figure 2**).

Despite the steric effect by the chloro atom (compound **2**) with respect to the hydrogen atom (compound **1**), in this **Figure 2** the highest similarity is between these compounds (0,976) with an euclidean distance (3,108), the substituent with most steric effect is trifluoromethyl (compound **10**), and this substituent decreases the quantum similarity in 0,735. Finally, in both trends we can see the same behavior. To analyses the electrostatic effects along the molecular set is shows the **Tables 4** and **5** using the equations 10 and 18.

As **Table 2**, in **Table 4** the highest values in the coulomb similarity is between the compounds **2** and **9** (0,999) with an euclidean distance (0,791, see **Table 5)**, the lowest value is between the compounds **5** and **7** (0,913) with an euclidean distance (20,221), these values shows as the resonance effects cause (dis)similarity along the molecular set. In general, comparing the overlap and coulomb indexes we can see highest values in these last. Therefore, the electrostatic effects can be more relevant than the steric to explain the antimalarial activity.

To study the trends on the molecular set using the coulomb operator with respect to the reference compound **1** is shows in **Figure 3** the scales of convergence quantitative (β) to study the electrostatic effects. The most active compound **7** (see **Table 1**) has the highest values of euclidean distance (22,474) with the compound **1**, this values is agrees with the size of the quinolinyl group and it resonance effect. These good similarity values (**Tables 2-5**) can be related with the cross-validated correlation coefficient (q^2 = 0.704) of the CoMSIA results reported by Xue **[12]** in this context the hybrid methodology (MQS and Chemical reactivity) reported can be independent of the number of molecules used.

In **Figure 3** the highest value is between the compounds **1** and **2** (0,994) with an euclidean distance (22,474) this result is agree with **Figure 2** while the lowest value is between the compounds **1** and **7** (0,893) and an euclidean distance of (4,191). To understand as the MQSM can be considered as QSAR descriptors we used the equation reported by Carbó and co-workers **[41]**. In this equation any physical-chemical property (e.g.; entropy) or biological activity of a molecule (π_I) can be considered to be the expectation value of an unknown quantum-mechanical observable

$$\pi_I = \left\langle \Omega(x) \right\rangle_I = \int \Omega(x)\rho_I(x)dx = \left\langle \Omega \middle| \rho_I \right\rangle \quad \textbf{(20)}$$

Being (ρ_I) the density function of molecules I, ($\mathbf{\Omega}$) represent some quantum-mechanical operator. Using the mean of MQSM is possible obtain the molecular density function projected into a n-dimensional point-molecule vector $\mathbf{Z_I}$, in this context we can approximate the operator ($\mathbf{\Omega}$) through a vector **w**.

$$\pi_I = \left\langle \Omega \right\rangle_I \approx \mathbf{w}^{\mathbf{T}}\mathbf{z}_{\mathbf{I}} \quad \textbf{(21)}$$

In this equation the point operator **w** is unknown a priori; yet its elements can be evaluate using the least-squares fitting for a molecular training set. This equation (21) shows a possible relationship between MQSM and the QSAR field **[42]**.

Due to that the coulomb operator has more incidence in the molecular set (the highest values in the Carbó index see **Tables 2** and **4**) we used the chemical reactivity descriptors. In **Table 6** are shows the global reactivity descriptors such as chemical potential (μ), hardness (η), softness (S) and electrophilicity (ω).

In **Table 5** the reference compound **1** has a chemical potential (μ=-3,9550 eV), hardness (η=4,894 eV), softness (S=0,204 eV^{-1}) and electrophilicity (ω=1,598 eV). However, the compound **7** (quinolinyl as substituent) has the highest chemical potential (μ=-3,622 eV), while that the compound **10** (trifluoromethyl as substituent) has the highest hardness (η=5,133 eV) with softness (S=0,195 eV^{-1}), finally the compound 3 (nitro as substituent) has the highest electrophilicity with (ω=2,353 eV).

Although the compound **3** has the lowest biological activity (pIC$_{50}$=4,65), this compound has the highest electrophilicity value (ω=2,353 eV). On the other hand, the most active compound **7** (pIC$_{50}$=5,70) has the highest chemical potential (μ=-3,622 eV) and lowest electrophilicity (ω=1,5351 eV) these results can be related with the no-covalent interactions associated to these antimalarial compounds **[12, 43]**. With these descriptors we can see as the acceptor and donor groups can have influence on the reactivity parameters along the molecular set. To analyses the local reactivity, in **Figure 4** is shows the Fukui Functions on the carbon atom C^1 (see **Figure 1**).

In **Figure 4** are highlight the Fukui Functions $f^{+/-}(r)$ regions, these regions shows the type of stabilization of these compounds on the active site. In this sense, the substituents analyzed increase the chemical activity and the retrodonor process. Additionally, this retrodonor process can determine the stabilization in the active site and the antimalarial activity presented. On the other hand, these Fukui regions are agrees with the docking studies reported by Oliveira and co-workers **[44]** and other works about structure-activity relationship **[45-46]**.

One of the important goals into the QSAR studies is the quantitative correlation of molecular structure with the binding constant and subsequently the prediction of this property for novel compounds. In this sense, this methodology can help to characterize those spatial features that are responsible for activity changes in a series of drug molecules when the receptor is known or not.

Additionally, the entopic changes associated to the molecular set can be understood in term of quantum similarity. Furthermore, the target property to be correlated and predicted in a comparative analysis is a free energy value. It can be imagined that enthalpic contributions to the binding constant are covered by molecular descriptors that explore the capabilities of molecules to perform intermolecular interactions with a putative receptor, these insights also can be understand in terms of chemical reactivity.

Table 2. Local molecular quantum similarity matrix using the overlap operator (equation 18).

C[a],Ove.[b]	1	2	3	4	5	6	7	8	9	10	11
1	1,000										
2	0,976	1,000									
3	0,849	0,903	1,000								
4	0,866	0,899	0,840	1,000							
5	0,919	0,935	0,867	0,847	1,000						
6	0,902	0,923	0,902	0,886	0,873	1,000					
7	0,798	0,823	0,763	0,791	0,774	0,769	1,000				
8	0,940	0,964	0,881	0,880	0,911	0,901	0,811	1,000			
9	0,964	0,991	0,909	0,897	0,939	0,947	0,819	0,958	1,000		
10	0,735	0,784	0,820	0,753	0,711	0,881	0,682	0,756	0,830	1,000	
11	0,884	0,917	0,893	0,841	0,858	0,867	0,824	0,892	0,921	0,810	1,000

[a] C: compound.

[b] Ove: Overlap Index.

Table 3. Local molecular quantum similarity matrix (MQSM) using the euclidean distance of overlap.

C[a], DO[b]	1	2	3	4	5	6	7	8	9	10	11
1	0,000										
2	0,757	0,000									
3	2,115	1,731	0,000								
4	1,926	1,685	2,229	0,000							
5	1,497	1,337	2,019	2,113	0,000						
6	1,604	1,429	1,727	1,808	1,886	0,000					
7	2,531	2,382	2,830	2,621	2,708	2,728	0,000				
8	1,219	0,943	1,894	1,831	1,565	1,625	2,458	0,000			
9	0,916	0,450	1,672	1,698	1,299	1,201	2,408	1,021	0,000		
10	3,108	2,850	2,639	3,041	3,269	2,190	3,523	3,004	2,586	0,000	
11	1,765	1,509	1,814	2,151	2,014	1,937	2,398	1,719	1,473	2,688	0,000

[a] C: compound.

[b] DO: Euclidean Distance of Overlap.

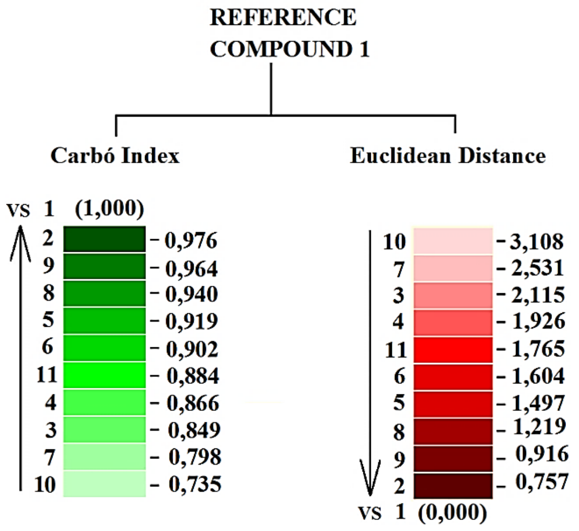


Figure 2. Scales of convergence quantitative α to steric effects proposed to the reference compound 1.

Table 4. Local molecular quantum similarity matrix using the coulomb operator (equation 19).

C[a],Cou.[b]	1	2	3	4	5	6	7	8	9	10	11
1	1,000										
2	0,994	1,000									
3	0,964	0,986	1,000								
4	0,939	0,962	0,975	1,000							
5	0,990	0,997	0,983	0,955	1,000						
6	0,978	0,991	0,991	0,973	0,988	1,000					
7	0,893	0,920	0,939	0,961	0,913	0,932	1,000				
8	0,982	0,993	0,987	0,971	0,989	0,991	0,932	1,000			
9	0,993	0,999	0,986	0,962	0,997	0,991	0,920	0,993	1,000		
10	0,936	0,965	0,986	0,971	0,959	0,981	0,944	0,973	0,967	1,000	
11	0,965	0,984	0,992	0,975	0,979	0,984	0,952	0,984	0,984	0,984	1,000

[a] C: compound.

[b] Cou: Coulomb Index.

Table 5. Local molecular quantum similarity matrix (MQSM) using the Euclidean Distance of coulomb.

C[a], DC[b]	1	2	3	4	5	6	7	8	9	10	11
1	0,000										
2	4,191	0,000									
3	10,896	7,135	0,000								
4	15,129	12,11	9,325	0,000							
5	5,331	2,920	7,472	12,741	0,000						
6	8,178	5,096	5,228	9,946	5,973	0,000					
7	22,474	19,799	16,786	13,299	20,221	17,989	0,000				
8	7,431	4,433	6,315	10,422	5,361	4,821	18,046	0,000			
9	4,318	0,791	7,120	12,172	2,857	4,939	19,831	4,443	0,000		
10	15,507	11,891	7,284	9,841	12,422	8,779	15,519	10,271	11,671	0,000	
11	10,851	7,474	4,863	9,190	8,171	6,902	15,268	6,817	7,435	7,472	0,000

[a] C: compound.

[b] DC: Euclidean Distance of Coulomb.

5. Conclusions

This work shows how the CoMSIA results can be understood in terms of (MQS) and Chemical reactivity descriptors. To carry out this aim, were used the CoMSIA results reported by Xue and coworkers **[12]**, this CoMSIA study is carried out on a serie of antimalarials chalcones.

The hybrid methodology reported, shows the steric and electrostatic effects in form of the scales of convergence quantitative convergence alpha (α) to steric effects and beta (β) to electrostatic effects, these scales allow study the substituent effects and were constructed using the reference compound 1 (hydrogen as substituent on the reference carbon C^1). These results were completed with a reactivity study using global and local descriptors such as chemical hardness, softness, electrophilicity, and Fukui Functions.

In this sense, the CoMSIA results reported by Xue **[12]** were modeled joining MQS and chemical reactivity; in this context these outcomes can be applied in QSAR correlations and docking studies to understand the antimalarial activity of these compounds. Taking into account that this methodologies can be used when the receptor is known or even when it is not known.

Acknowledgments

A.M.B thanks to the Universidad de Talca (CBSM)) for the continuous support to this investigation and to the postdoctoral project N^0 3150035 (FONDECYT, CHILE).

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Morales-Bayuelo A, Matute R A, Caballero J 2015 J. Mol. Model. 21, 156.
2. Carbó-Dorca R, Arnau M, Leyda L 1980 Int. J. Quant. Chem. 17, 1185.
3. Amat L, Carbó-Dorca R 2002 Int. J. Quant. Chem. 87, 59.
4. Gironés X, Carbó-Dorca R 2006 QSAR Comb. Sci. 25, 579.
5. Carbó-Dorca R, Gironés X 2005 Int. J. Quat. Chem. 101,8.
6. Bultinck P, Rafat M, Ponec R, Gheluwe B V, Carbó-Dorca R, Popelier P 2006 J. Phys. Chem. A. 110, 7642.
7. Robert D, Amat L, Carbó-Dorca R 1999 J. Chem. Inf. Comp. Sci. 39, 333.
8. (a) Morales-Bayuelo A, Vivas-Reyes R. 2013 J. Math. Chem. 51, 125. (b) Morales-Bayuelo A, Vivas-Reyes R. 2013 J. Math. Chem. 51, 1835. (c) Morales-Bayuelo A, Torres J, Vivas-Reyes R 2012 J. Theor. Comput. Chem. 11, 1. (d) Morales-Bayuelo A, Vivas-Reyes R. 2014 J. Quant. Chem. Article ID 239845, 19 pages. (e) Morales-Bayuelo A, Vivas-Reyes R. J. Quant. Chem. 2014, Article ID 850163, 12 pages.
9. Parr R G, Yang W 1989 Density Functional Theory of Atoms and Compounds; Oxford University Press: New York.
10. Geerlings P, De Proft F, Langenaeker W, 2003 Chem. Rev. 103, 1793.
11. Parr RG, Chattaraj PK. 1991 J. Am. Chem. Soc. 113, 1854.
12. Xue CX, Cui SY, Liu MC, Hu ZD, Fan BT 2004 Eur. J. Med. Chem. 39, 745.
13. Klebe G, Abraham U 1999 J. Comp.-Aided Mol. Design. 13, 1.
14. Klebe G, Abraham U, Mietzner T 1994 J. Med. Chem. 37, 4130.
15. Liu M, Wilairat P, Go PM 2001 J. Med. Chem. 44, 4443.
16. Carbó-Dorca R, Arnau M, Leyda L 1980 Int. J. Quant. Chem. 17, 1185.
17. Amat L, Carbó-Dorca R 2002 Int. J. Quant. Chem. 87, 59.
18. Gironés X, Carbó-Dorca R 2006 QSAR Comb. Sci. 25, 579.
19. Carbó-Dorca R, Gironés X 2005 Int. J. Quant. Chem. 101,8.
20. Bultinck P, Gironés X, Carbó-Dorca R (2005) Rev. Comput. Chem. 21,127.
21. Hirshfeld F L 1977 Theor. Chim. Acta. 44, 129.
22. De Proft F, Van Alsenoy C, Peeters A, Langenaeker W, Geerlings P 2002 J. Comput. Chem. 23, 1198.
23. Randic M, Johnson M A, Maggiora G M 1990. In Concepts and Applications of Molecular Similarity, Design of Compounds with Desired Properties. Eds., Wiley-Interscience. New York. 77.
24. (a) Boon G, Van Alsenoy C, De Proft F, Bultinck P, Geerlings P 2005 J. Mol. Struct. 727, 49. (b) Morales-Bayuelo A, Caballero, J. 2015 J. Mol. Mod. 21, 45.
25. Ayers P W, Anderson J S M, Bartolotti L J 2005 Int. J Quant. Chem. 101, 520.
26. Harbola MK, Chattaraj PK, Parr RG 1991 Isr. J. Chem. 31, 395.
27. Pearson R G 1997 Chemical Hardness; Applications from Compounds to Solids; Wiley-VHC, Verlag GMBH: Weinheim, Germany.
28. Yang W T, Parr R G 1985 Proc. Natl. Acad. Sci. 82, 6723.
29. Chattaraj PK, Sarkar U, Roy DR 2006 Chem. rev. 106, 2065.
30. Ayers P, Parr R G 2000 J. Am. Chem. Soc. 122, 2010.

31. Galván M, Pérez P, Contreras R, Fuentealba P 1999 Chem. Phys. Lett. 30, 405.
32. Mortier W J, Yang W 1986 J. Am. Chem. Soc. 108, 5708.
33. Fuentealba P, Pérez P, Contreras R 2000 J. Chem. Phys. 113, 2544.
34. Girones X, Robert D, Carbó-Dorca R 2001 J. Comput. Chem. 22, 255.
35. Carbó-Dorca R, Besalú E, Amat L, Fradera X 1995 J. Math. Chem. 18, 237.
36. Besalú E, Girones X, Amat L, Carbó-Dorca R 2002 Acc. Chem. Res. 35, 289.
37. Boon G, Langenaeker W, De Proft F, De Winter H, Tollenaere JP, Geerlings 2001 P J. Phys. Chem. A. 105, 8805.
38. (a) Becke A. D 1988 Phys. Rev. A. 38, 3098, (b) Lee C, Yang W, Parr R G 1988 Phys. Rev. B. 37, 785.
39. GAUSSIAN 09, Revision C.01, Frisch M J, G. Trucks W, Schlegel H B, Scuseria G E, Robb M A, Cheeseman J R, Scalmani G, Barone V, Mennucci B, Petersson G A, Nakatsuji H, Caricato M, Li X, Hratchian H P, Izmaylov A F, Bloino J, Zheng G, Sonnenberg J L, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery J A Jr., Peralta J E, Ogliaro F, Bearpark M, Heyd J J, Brothers E, Kudin K N, Staroverov V N, Keith T, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant J C, Iyengar S S, Tomasi J, Cossi M, Rega N, Millam J M, Klene M, Knox J E, Cross J B, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann R E, Yazyev O, Austin A J, Cammi R, Pomelli C, Ochterski J W, Martin R L, Morokuma, K. Zakrzewski V G, Voth G A, Salvador P, Dannenberg J J, Dapprich S, Daniels A D, Farkas O, Foresman J B, Ortiz J V, Cioslowski J, and Fox D J, Gaussian, Inc., Wallingford CT. 2010.
40. Arfken GB, Weber HJ (2000) Mathematical methods for physicists, 5th edn. Academic, Boston.
41. Carbó-Dorca R, Besalu E, Amat L, Fradera X 1996 J. Math. Chem. 19, 47.
42. Fradera X, Amat L, Besalú E, Carbó-Dorca R 1997 Quant. Struct.-Act. Relat. 16, 25.
43. Oliveira M, Cenzi G, Nunes RR, Andrighetti, Valadão D MS, Reis C, Oliveira Simões CM, Nunes RJ, Júnior MC, Taranto AG, Sanchez BAM, Viana GHR Varotti FP 2013 Molecules. 18, 15276.
44. Rongshi L, Kenyon G, Cohen FE, Chen X, Gong B, Dominguez JN, Davidson E, Kurzban G, Miller RE, Nuzum EO, Rosenthal PJ, McKerrow JH 1995 J. Med. Chem. 38, 5031.
45. Domínguez JN, León C, Rodrigues J, Domínguez DG, Gut J, Rosenthal PJ. 2005 Farmaco I. 60, 307.
46. Tomar V, Bhattacharjee G, Kamaluddin S, Rajakumar KS, Puri SK 2010 Eur. J. Med. Chem. 45, 2745.

SciForum
Mol2Net

SISTEMAT X - A Web Tool to Manage Databases of Secondary Metabolites

Marcus Tullius Scotti[1*], **Roberto Oliveira Da Silva Junior**[1], **Silas Yudi Konno De Oliveira Santos**[1], **Luciana Scotti**[1,2]

[1] Laboratory of Cheminformatics, IPEFARM, Address; E-Mail: mtscotti@gmail.com; r2@email

* Author to whom correspondence should be addressed; E-Mail: mtscotti@gmail.com;
Tel.: 55-83-99869-0415.

Published: 4 December 2015

Abstract: : The internet aids to promote a new process of data/information transmission that two decades ago simply did not exist. A simple search on the internet provides an answer; new discussion forums often provide answers that would take days or even weeks of research. Nevertheless, some information is still obtained indirectly and relatively time consuming, hence techniques of bank architecture chemical data, query and visualization have been developed constantly. We can find several internet applications to search and predict spectroscopic data, biological activity of ligands, or to predict toxicity of new compounds or pesticides. Our research group is developing SISTEMAT X web, [2] a tool that manages databases of natural products. Currently, our database has more than 1,100 sesquiterpene lactones and 800 flavonoids with more than four thousand botanical occurrences of the Asteraceae family and approximately 400 alkaloids which represents more than 750 botanical occurrences of the Apocynaceae family and several terpenes of Annonaceae that correspond more than 800 botanical occurrences. SISTEMAT X is a set of integrated programs and tools that perform cheminformatics tasks that include database management, chemical structure editor, visualization of chemical structures, and prediction of physicochemical properties among others. Most of the components are intuitive and friendly using a graphical interface. In the last year we have migrated applications that use Java interface for JavaScript, since the last is geared to web pages. This software has already been registered by our research group, through UFPB, in “Instituto Nacional de Propriedade Industrial” with number BR 51 2015 000073. The site is running at the address: www.sistematx.ufpb.br. We are developing constantly it, improving existing features such as adding new. All tools are available to the scientific community.

Keywords: web tools, secondary metabolites, databank, cheminformatics.

Mol2Net YouTube channel: *http://bit.do/mol2net-tube*

1. Introduction

The internet aids to promote a new process of data/information transmission that two decades ago simply did not exist. A simple search on the internet provides an answer; new discussion forums often provide answers that would take days or even weeks of research. Nevertheless, some information is still obtained indirectly and relatively time consuming, hence techniques of bank architecture chemical data, query and visualization have been developed constantly. We can find several internet applications to search and predict spectroscopic data, biological activity of ligands, or to predict toxicity of new compounds or pesticides [1].

Studies of publications on modern chemistry dating from the eighteenth century and its volume has increased steeply since the First World War. Being the huge volume of current data and highly complex nature, efforts were directed to contribute to the successful organization and accessibility in the last 25 years [1].

2. Results and Discussion

The SISTEMATX WEB has the function of producing and designing Web pages, objects such as images, headings, tables, among others. It was developed the system of management of chemicals that is in operation at: http://www.sistematx.ufpb.br [2]. The contact modules were created by the chemical name for SMILES (Simplified Molecular Input Line Entry Specification) and substructure. Also created by the query returns all species compounds (secondary metabolites) that have been isolated in this (Figure 1), and all the species in the taxonomic classification family. SISTEMAT X added to the web registration system that was developed in the last year, data from the molecular structure and its occurrence botany from the literature review. To draw a molecule associate with a class, skeleton, both previously registered, and finally a name. It can be in the query screens besides the registered data, information generated automatically by the system using Application Programming Interface (API) Chemaxon (www.chemmaxon.com): IUPAC name, SMILES code, compound oxidation number, INCHIKEY and compound ID in the database (figure 2).

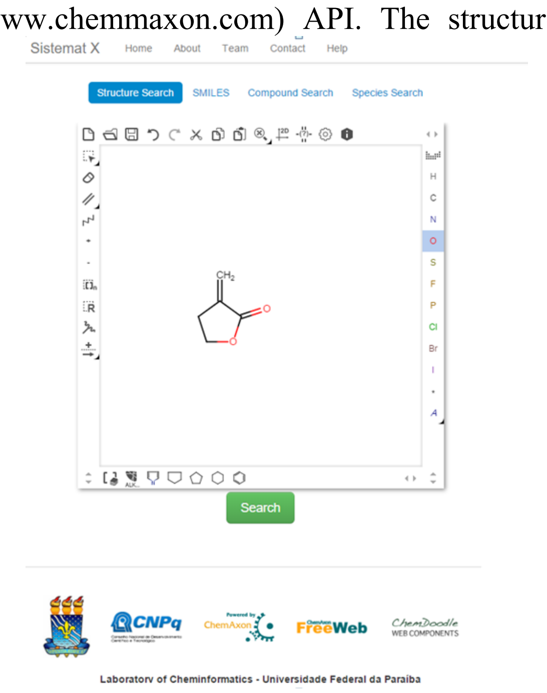
SISTEMATX WEB has functions to produce and design Web pages, objects such as images, headings, tables, among others. It was developed to manage secondary metabolites an is online at: http://www.sistematx.ufpb.br. Modules in order to perform searches using a chemical name (IUPAC or common), SMILES (Simplified Molecular Input Line Entry Specification) and substructure (Figure 1). It is possible to perform selecting a specific genus and species (Figure 1). The query returns all compounds (secondary metabolites) that correspond the performed query and selecting a structure, it is possible to visualize all species where the compound has been isolated (Figure 2), and all the species and their taxonomic classification: family, subfamily, tribe, subtribe, genus and specie. We added to SISTEMATX web registration modules that were developed in the last year, data from the molecular structure and its occurrence botany (species where a secondary metabolite has been isolated), data that is being collected from the literature review.

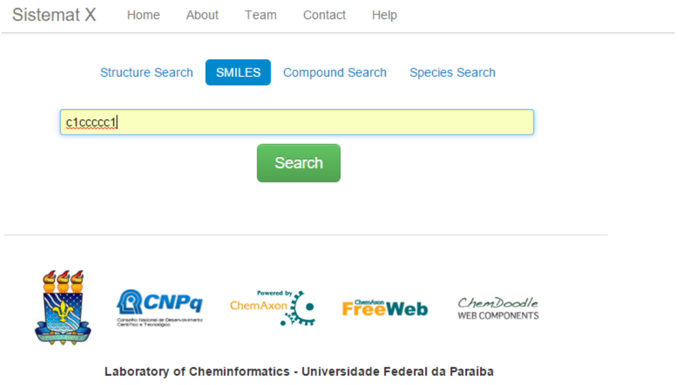
To input a compound in SISTEMATX web, it is necessary to draw a structure and state respective class, skeleton (both should be previously registered), and finally a common name (optional). IUPAC name, number of oxidation, SMILES code, INCHIKEY and ID (identification number) are generated automatically as soon as a structure is registered in the SISTEMATX web databank by the system using API Chemaxon (www.chemmaxon.com): IUPAC name SMILES code, compound oxidation number, INCHIKEY and compound ID in the database (Figure 2).

It is also generated three-dimensional data structure from two dimensions using Chemaxon (www.chemmaxon.com) API. The structure in 3D (three dimensions) is displayed in figure 3 using the API developed by ChemDoodle (web.chemdoodle.com). Both kind of structures 2D (two dimensions) as 3D can be downloaded.

The SISTEMATX web interface is very light and friendly, suitable to use in several kinds of devices (including the mobile), and works on browsers as Chrome and Mozilla and it does not require JAVA installed on the device.

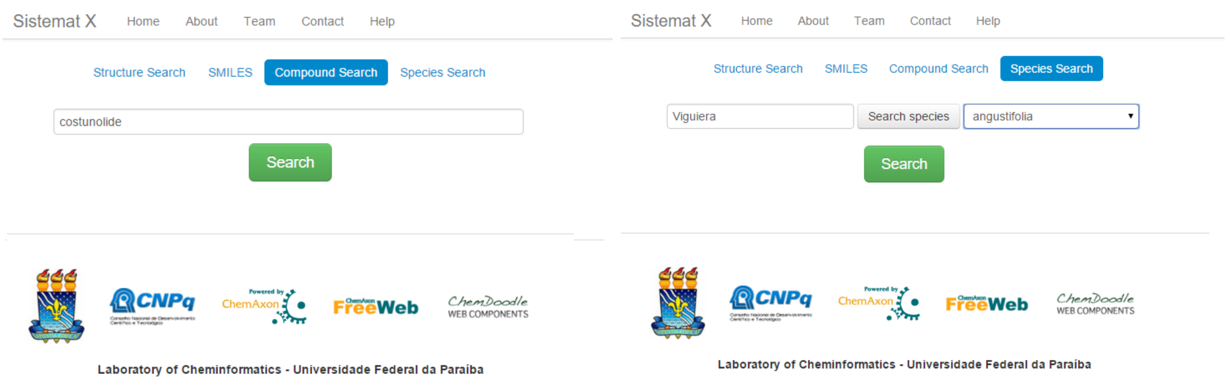
Currently, our database has more than 1,100 sesquiterpene lactones and 800 flavonoids with more than four thousand botanical occurrences of the Asteraceae family and approximately 400 alkaloids which represents more than 750 botanical occurrences of the Apocynaceae family and several terpenes of Annonaceae that correspond more than 800 botanical occurrences





c1ccccc1

Search
CNPq
Powered by ChemAxon
ChemAxon FreeWeb
ChemDoodle WEB COMPONENTS
Laboratory of Cheminformatics - Universidade Federal da Paraiba

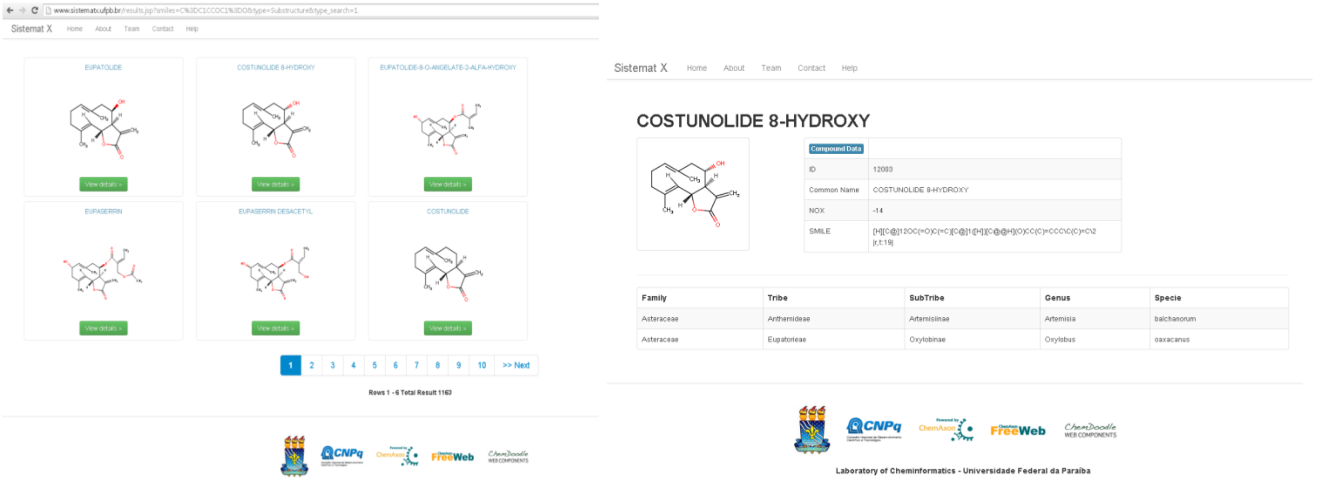


costunolide

Viguiera

angustifolia

Figure 1. Screens of search for substructure, for SMILES code, for compound name, and for species name.


Sistemat X
Home
About
Team
Contact
Help
EUPATOLIDE
COSTUNOLIDE 8-HYDROXY
EUPATOLIDE-8-O-ANGELATE-2-ALFA-HYDROXY
EUPASERRIN
EUPASERRIN DESACETYL
COSTUNOLIDE
View details »
1 2 3 4 5 6 7 8 9 10 >> Next
Laboratory of Cheminformatics - Universidade Federal da Paraiba

www.sistematx.ufpb.br/results.jsp?smiles=C%3DC1CCOC1%3DO&type=Substructure&type_search=1

Rows 1 - 6 Total Result 1163

COSTUNOLIDE 8-HYDROXY

Compound Data	
ID	12003
Common Name	COSTUNOLIDE 8-HYDROXY
NOX	-14
SMILE	[H][C@]12OC(=O)C(=C)[C@]1([H])[C@@H](O)CC(C)=CCC\C(C)=C\2 |r,t:19|

Family	Tribe	SubTribe	Genus	Specie
Asteraceae	Anthemideae	Artemisiinae	Artemisia	balchanorum
Asteraceae	Eupatorieae	Oxylobinae	Oxylobus	oaxacanus

Figure 2. Screen with the result of a search for lactone substructure and some data of costunolide 8-hydroxy available.

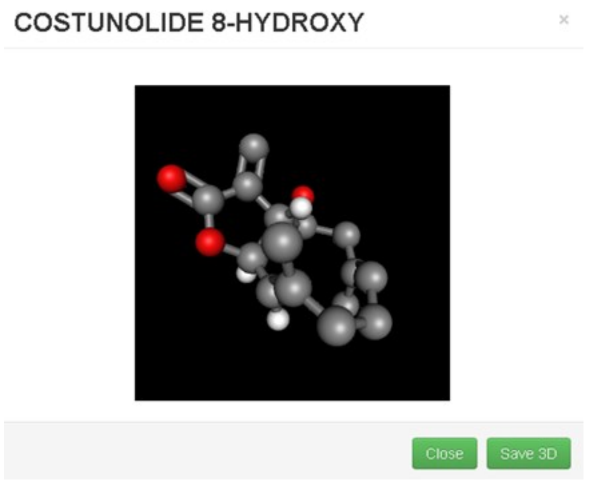


Figure 3. Screen of a structure in 3D of SISTEMATX web.

3. Materials and Methods

To build all tools and develop interfaces that is able to be used in several kinds of mobile devices and computers without the need for Java installed in the machine, we use technologies such as HTML 5, CSS3, JavaScript, PHP, AJAX (Asynchronous JavaScript XML together), JSF (JavaServer Faces).

The database was developed using MySQL and various security tools based on OWASP (Open Web Application Security Project - https://www.owasp.org/index.php/Main_Page) were implemented. APIs the CHEMAXON were used for the generation of auxiliary data related to the structure and to make the search tool. The API ChemDoodle was used for visualizing the 3D structures.

4. Conclusions

The site of the web tool is: www.sistematx.ufpb.br. We are developing constantly it, improving existing features such as adding new. All tools are available to the scientific community.

Acknowledgments

National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico) CNPq

.

Author Contributions

Marcus Tullius Scotti is the coordinator of this work, and idealize the web tool. Selected the IDE and API that are used. **Roberto Oliveira da Silva Junior aid to develop the database and the integration APIs, working mainly in the background of the web tool configuring the server and improving its security. Silas Yudi Konno De Oliveira Santos has worked mainly in the front end of the web tool, aid to planning the screen and building it. He aids in the integration of some APIs too. Luciana Scotti aids to design some functionalities of the web tool and testing some implemented functionalities.**

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Gasteiger, J. Handbook of Chemoinformatics: From Data to Knowledge, 4th ed,; Wiley-VCH. Weinheim, 2003.
2. SISTEMAT X. Avaliable online: http://www.sistematx.ufpb.br

SciForum
Mol2Net

ASD Module: A Software to Support the Personal Autonomy in the Daily Life of Children with Autism Spectrum Disorder

Betania Groba[1,*], Javier Pereira[1], Laura Nieto[1], Thais Pousada[1], Susana Falcón[1], Cristian R. Munteanu[2] and Alejandro Pazos[2]

[1] Centre of Medical Informatics and Radiological Diagnosis (IMEDIR), Faculty of Health Science, University of A Coruña, As Xubias s/n, 15006, Spain

[2] Centre of Medical Informatics and Radiological Diagnosis (IMEDIR), Faculty of Computer Science, University of A Coruña, Campus de Elviña, s/n, 15071 A Coruña, Spain

* E-Mail: bgroba@udc.es; Tel.: +34 981-167-000 (ext. 5870)

Published: 4 December 2015

Abstract: Introduction: It was observed that technology developers expressed a clear interest to design programs that meet the needs of individuals with Autism Spectrum Disorder (ASD). Several authors indicate that any software designed for people with ASD has to include special requirements in the design. Methodology: This research study describes a software for children with ASD named Module ASD and describes the interactive process of design that it was followed. This research focuses on a software development and the design process, based on scientific evidence study, consultation and tests done by specialists, children with ASD and their families. The techniques used to formalize the collection of information from different groups of participants were: observation, interview, group discussions and field book. Results: The ASD Module is the result of the study and it is a free technological application that is made up of a set of virtual keyboards (or adapted interfaces), digital schedules and activities, especially designed and tested by and for children with ASD. The application is included in the In-TIC PC software. The software is available for the Windows operating system and was implemented with the Visual Studio development tool, .NET environment, C# programming language and Windows Form technology. Other materials used for content development were the Interactive Books Multimedia (LIM according to the Spanish acronym) software and the ARASAAC pictograms. Discussion and conclusions: The results show that the digital content can be oriented to promote independence in several daily activities (ADL, education, leisure and social participation). This technology design provides useful information for researchers, developers, social and healthcare professionals and families, with the aim of offering alternatives for children with ASD and facilitating the understanding of daily life.

Keywords: Autism Spectrum Disorders; technology; software; design; daily live activities.

1. Introduction

Nowadays, the virtual context and Information and Communications Technology (ICT) have an effect on our lives, in many cases facilitating the performance of daily activities (1). Likewise, technology is beginning to change the lives of many people with Autism Spectrum Disorder (ASD) to the extent that it is increasingly used in the intervention and research related to people with ASD (2).

The type of technology used in interventions with people with ASD is varied (computers, mobile devices, video recordings, robots and virtual reality). There has been an increase in the use of technology in this field for several reasons:

1. Reports and studies stating that people with ASD show interest and motivation for using visual technology devices;
2. Studies supporting the effectiveness of technology as an intervention tool;
3. Increased software development in this field.

Therefore, at present, the technology-based intervention for persons with ASD faces many challenges. The main challenges are related to the development of software which take into account the specific learning styles, abilities and needs of children with ASD (3, 4).

2. Results and Discussion

Results

The ASD Module is the result of the study and it is a free technological application that is made up of a set of virtual keyboards (or adapted interfaces), digital schedules and activities, especially designed and tested by and for this study group and included in the In-TIC PC software (5). In-TIC PC is the baseline software with which the specific block for people with ASD was created and whose purpose is to adapt the Windows environment by means of virtual keyboards that allow easy computer access and use and/or to facilitate social communication and participation. The software is available for the Windows operating system and was implemented with the Visual Studio development tool, .NET environment, C# programming language and Windows Form technology. Other materials used for content development were the Interactive Books Multimedia (LIM according to the Spanish acronym) software (6) and the ARASAAC pictograms (7).

The ASD Module is based on the perspective of considering the person with autism as the central axis, and, around it, the activities of their life interests. To provide an example that can be adapted and customized to other people, let us consider the case of Arancha, a girl with ASD. When analyzing the main screen (Figure 1), Arancha is observed in the central part and, around her, the activities in which TIC can support her on a daily basis. There are 5 sections:

1. Schedule. This section refers to the basic daily activities. It was designed to make access to timetables, schedule and activity sequences easier, to help organizing a person's day and activities.
2. Education. This section includes activities necessary for learning and participation in the school environment (8). It is made up of contents about five basic categories: colors, numbers,

letters, parts of the body and feelings. Moreover, dynamic activities have been designed in the same line, with the LIM system, which are integrated with the keyboards of this module (see Figure 2).

3. Leisure. This section refers to leisure activities, that is, time spent away from compulsory activities (8). It includes resources classified as games, stories and documents about people with ASD (see Figure 3).

4. Communication. This section contains basic communication keyboards. Pictograms and/or pictogram writing are used, which, combined with speech synthesis resources, stimulate the communication of basic requests and needs. There are also other keyboards that favor the narrative discourse and the development of questions in order to encourage a person's social participation.

5. Computer access. This section refers to the access to several conventional programs from Windows environment; in this section both access and use are simplified. The programs used are as follows: Wordpad (text editor), Microsoft Paint (drawing program), Windows calculator and Windows Media Player.

Discussion

The results of the software development and design show that the digital content can be oriented to promote independence for several daily activities (ADL, education, leisure and social participation). The literature in this field shows that people with ASD have difficulty in performing these activities (9). Therefore, within this software, the person is seen as the central focus and surrounding them are the activities in which ICT can be a facilitator. However, so far, many of the software tools that have taken into account the perspective of people with ASD, have focused solely on promoting specific abilities (10, 11, 12, 13, 14, 15, 16, 17).

On the other hand, motivation is essential for learning in all areas of daily life. As stated previously, a high percentage of people with ASD have a predilection for technological devices. Therefore, an alternative to traditional interventions is proposed, based on the preference and motivation for the use of ICT.

The decisions made by the interdisciplinary team developing the ASD Module have been described and they have shown the need to involve both people with ASD and those within their immediate surroundings in these processes, as already described in several publications (18, 19)

.

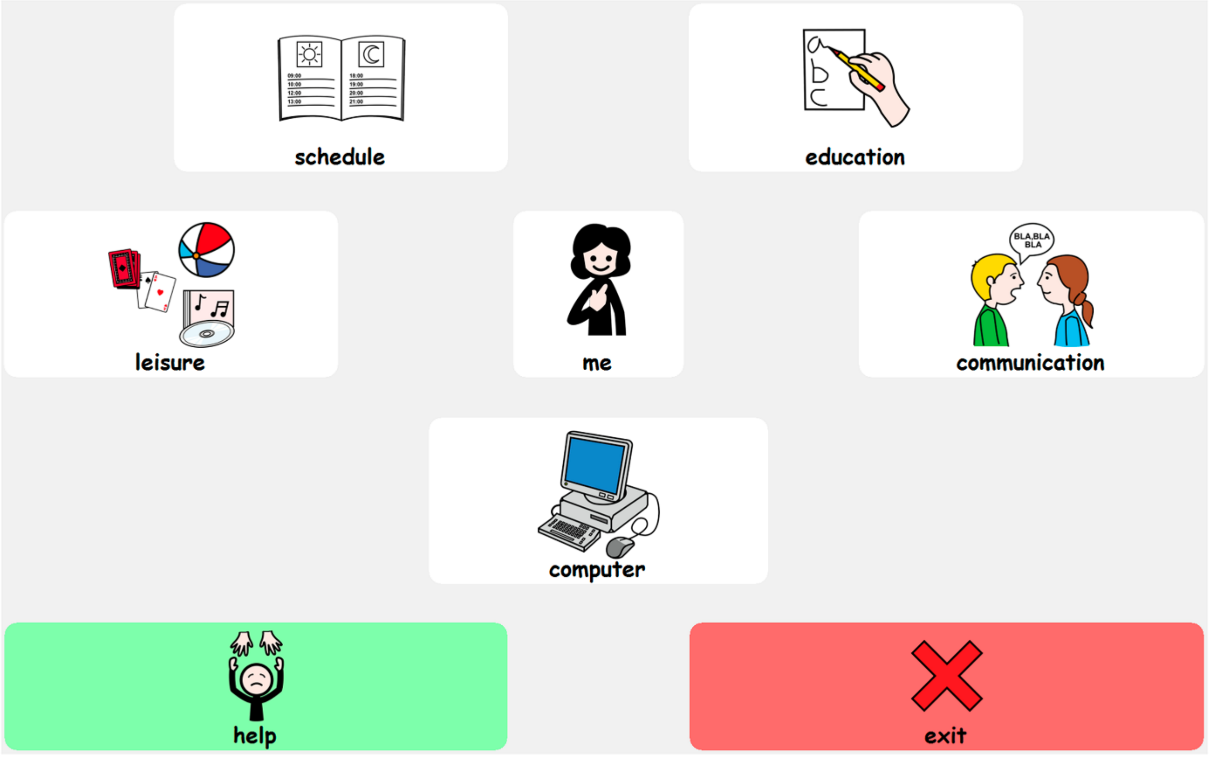


Figure 1. Main screen of the ASD Module

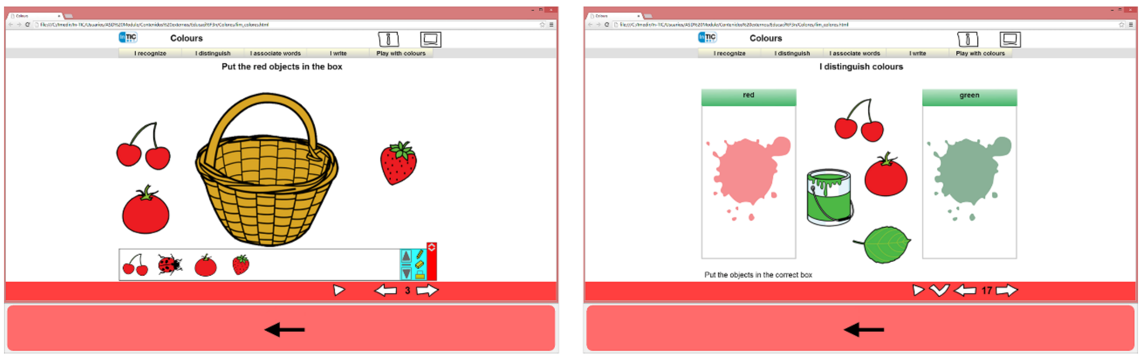


Figure 2. Examples of LIM activities about using colors for the Education section

Figure 3. Example of the application browser to locate the videos of interest

3. Materials and Methods

Participants and settings

This research study involved four groups of respondents, meeting the following criteria:

1st Group: Professionals with experience in the intervention with people with ASD

The study involved 20 professionals who were recruited from centers of direct care for people with ASD, such as two schools of special education, specific for children with ASD diagnosis, an ASD-specialized clinic for psychological intervention, and an association that provides support for families living with an adult with ASD. The selection criteria for these professionals were as follows: (a) having a degree in education sciences or health sciences or representing the interests of organizations that serve people with ASD; (b) being able to prove work experience of at least one year in the intervention with people with ASD. Finally, a heterogeneous group was obtained, made up of 3 representatives of organizations, 3 psychologists, 11 schoolteachers, 1 social worker and 2 speech therapists.

2nd Group: Professionals with experience in the development and design of technology for disabled people

The project involved 13 professionals recruited from two centers that make use of the technology applied to social and/or health context. Participants also met the following criteria: (a) having a degree in the field of technology or health sciences or promoting social projects; (b) being able to prove work experience of at least one year in the design, development and/or testing of accessible technology related to health and quality of life for disabled people. The group involved the collaboration of 4 experts in the implementation of social projects, 1 doctor, 5 engineers in computer science and 3 occupational therapists.

3rd Group: Family members of people with ASD

Participants were 3 direct relatives of people with ASD and who, in turn, were part of the organizations for people with this diagnosis.

4th Group: Children with ASD

In the final phase, 3 children (two boys and a girl) participated, aged between 10 and 13 years old. All of them were selected from a special education center and met the criteria described below: (a) having a diagnosis of autism disorder according to the Diagnostic and Statistical Manual of Mental Disorders (4th ed.; DSM-IV); (b) presenting learning difficulties; and (c) not having used the specific software prior to testing.

The protocol was approved by the Autonomous Ethics Committee of Research in Galicia. The project participants signed the informed consent to participate in the study and, in the case of children with ASD, their guardians were responsible for authorizing their participation. In addition, in the cases of the 1st, 2nd and 3rd Groups, a specific authorization for interview recording was signed.

Procedure

During software development and design, a search for scientific evidence was performed, in addition to consultation and testing by specialists in ASD, and/or technology, and by children with ASD. To this end, research has been based on user-centered design and has followed an iterative procedure. This is a cyclical process, divided into the following phases:

1. Study and analysis of the recent scientific evidence on the design of technology for people with ASD. Participants: 2nd Group.
2. Study and analysis of the recent scientific evidence on the skills and ways of processing information of people with ASD. Participants: 2nd Group.

3. Observation, analysis and discussion on the skills and ways of processing of this population and their influence on the design of technology. Participants: 1st, 2nd, 3rd, and 4th Groups.
4. Design and development of the application. Participants: 2nd Group.
5. Software Testing by professionals and family members of people with ASD. Participants: 1st and 3rd Groups.
6. Software quality improvement. Participants: 2nd Group.

After this iterative process, the resulting application was tested by the 4th Group, that is, children with ASD.

Information collection techniques

The techniques used to formalize the collection of information from different groups of participants were: observation, interview, group discussions and field book.

.

4. Conclusions

The ASD Module has been developed based on the learning styles, abilities and needs of children with ASD and has been integrated into the free In-TIC computer software. This block includes virtual keyboards, schedules and activities to promote independence in everyday activities.

The ASD Module is provided as an example available to families and professionals assisting people with ASD so that, through the customization of the necessary aspects, it could be turned into a functional tool.

The basic ideas used for the development and design of the application have been explained, including scientific evidence and considerations set out during the iterative process of the software.

This list of rules for technology design provides useful information for researchers, developers, social and healthcare professionals and families, with the aim of offering alternatives for children with ASD and facilitating the understanding of daily life. Moreover, the need to research and develop studies for analyzing the responses and opinions of people with ASD and those within their immediate surroundings was demonstrated herein.Main text paragraph.

Acknowledgments

This study was supported by the Orange Foundation in Spain and the Spanish Ministry for Industry, Commerce and Tourism through Avanza Program.

This work was partially supported by the General Directorate of Culture, Education and University Management of Xunta de Galicia (Ref. GRC2014/049 and (CN2012/211); Galician Network for Colorectal Cancer Research (REGICC) (Ref. R2014/039), funded by the Xunta de Galicia and "Collaborative Project on Medical Informatics (CIMED)" PI13/00280 funded by the Carlos III Health Institute from the Spanish National plan for Scientific and Technical Research and Innovation 2013–2016. European Fund for Regional Development (FEDER) in the European Union.

We would like to thank to each participant in the project for their contributions, knowledge and time. We would also like to thank the families, the professionals and the children of ASPANAES (Association of Parents of individuals with Autism Spectrum Disorders of A Coruña) and aÚPa (children´s and baby´s clinic of Psychology and Physiotherapy of A Coruña).

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Groba, B.; Canosa, N.; Nieto, L. Tecnologías de la Información y las Comunicaciones en salud mental. In *Terapia Ocupacional en Salud Mental*, 1st ed.; Moruno, P., Talavera, M., Eds.; Elsevier Masson: Barcelona, 2012, pp.371-391.
2. Bölte, S. Computer-based intervention in autism spectrum disorders. In *Focus on Autism Research*; Ryaskin, O.T., Ed.; Nova Biomedical: New York, 2004, pp. 247-260.
3. Barry, M.; Pitt, I. Interaction design: a multidimensional approach for learners with autism. In *Proceedings of the 5th International Conference for Interacting Design and Children*, Tampere, Finland; ACM Digital Library, 2006; pp. 33–36. doi:http://doi.acm.org/10.1145/1139073.1139086
4. Putnam, C., & Chong, L. Software and technologies designed for people with autism. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility - Assets '08*), New York, EE.UU.; ACM Press, 2008, pp. 3-10. doi: 10.1145/1414471.1414475
5. Fundación Orange; IMEDIR. In-TIC: Integración de las Tecnologías de la Información y las Comunicaciones en los colectivos de personas con diversidad funcional. Avaliable online: http://www.proyectosfundacionorange.es/intic (accessed on 9 November 2015).
6. Macías, F. Libros Interactivos Multimedia. Avaliable online: http://www.educalim.com/cinicio.htm (accessed on 9 november 2015).
7. Gobierno de Aragón. Portal Aragonés de la Comunicación Aumentativa y Alternativa. Avaliable online: http://catedu.es/arasaac (accessed on 9 November 2015).
8. American Occupational Therapy Association. (2014). Occupational Therapy Framework: Domain & Process 3rd Edition. *The American Journal of Occupational Therapy*, **2014**, 68, S1 - S48. doi:10.5014/ajot.2014.682006
9. Watling, R.; Tomchek, S.; LaVesser, P. The scope of occupational therapy services for individuals with Autism Spectrum Disorders across the lifespan. *American Journal of Occupational Therapy*, **2005**, 59, 680-683. doi:10.5014/ajot.59.6.680
10. Bernard-Opitz, V.; Sriram, N.; Nakhoda-Sapuan, S. Enhancing social problem solving in children with autism and normal children through computer-assisted instruction. *Journal of Autism and Developmental Disorders*, **2001**, 31, 377–384. doi:10.1023/A:1010660502130
11. Bosseler, A., & Massaro, D. W. Development and evaluation of a computer-animated tutor for vocabulary and language learning in children with autism. *Journal of Autism and Developmental Disorders*, **2003**, 33, 653–672. doi:10.1023/B:JADD.0000006002.82367.4f
12. Campillo, C.; Herrera, G.; Remírez de Ganuza, C.; Cuesta, J. L.; Abellán, R.; Campos, A.; et al. Using Tic-Tac software to reduce anxiety-related behaviour in adults with autism and learning difficulties during waiting periods: A pilot study. *Autism*, **2014** 18, 264–271. doi:10.1177/1362361312472067

13. Den Brok, W. L. J. E.; Sterkenburg, P. S. Self-controlled technologies to support skill attainment in persons with an autism spectrum disorder and/or an intellectual disability: a systematic literature review. *Disability and Rehabilitation: Assistive Technology*, **2015**, 10, 1-10. doi:10.3109/17483107.2014.921248
14. Goldsmith, T. R.; LeBlanc, L. A. Use of technology in interventions for children with autism. *Journal of Early and Intensive Behavior Intervention*, **2004**, 1, 166–78. doi:10.1037/h0100287
15. Moore, M.; Calvert, S. Brief report: vocabulary acquisition for children with autism: teacher or computer instruction. *Journal of Autism and Developmental Disorders*, **2000**, 30, 359-362. doi:10.1023/A:1005535602064
16. Ramdoss, S.; Lang, R.; Mulloy, A.; Franco, J.; O'Reilly, M.; Didden, R.; Lancioni, G. Use of computer-based interventions to teach communication skills to children with autism spectrum disorders: A systematic review. *Journal of Behavioral Education*, **2011**, 20, 55–76. doi:10.1007/s10864-010-9112-7
17. Ramdoss, S.; Mulloy, A.; Lang, R.; O'Reilly, M.; Sigafoos, J.; Lancioni, G.; et al. Use of computer-based interventions to improve literacy skills in students with autism spectrum disorders: A systematic review. *Research in Autism Spectrum Disorders*, **2011**, 5, 1306–1318. doi:10.1016/j.rasd.2011.03.004
18. Abascal, J.; Nicolle, C. Moving towards inclusive design guidelines for socially and ethically aware HCI. *Interacting with Computers*, **2005**, 17, 484-505. doi:10.1016/j.intcom.2005.03.002
19. Porayska-Pomsta, K.; Frauenberger, C.; Pain, H.; Rajendran, G.; Smith, T.; Menzies, R.; et al. (2012). Developing technology for autism: an interdisciplinary approach. *Personal and Ubiquitous Computing*, **2012**, 16, 117-127. doi:10.1007/s00779-011-0384-2

SciForum
Mol2Net

Asymmetric Mizoroki-Heck Reactions: Generation of Quaternary Stereocenters and Cascade Cyclizations

Ane R. Azcargorta,[1] Iratxe Barbolla,[1] E. Coya,[1] Nuria Sotomayor[1], and Esther Lete[1]*

[1] Departamento de Química Orgánica II, Facultad de Ciencia y Tecnología, Universidad del País Vasco / Euskal Herriko Unibertsitatea UPV/EHU. Apdo. 644. 48080 Bilbao, Spain. http://www.ehu.es/oms

* Author to whom correspondence should be addressed; E-Mail: esther.lete@ehu.es; Tel.: +34 946012576; Fax: +34 946012748.

Published: 4 December 2015

Abstract: Mizoroki-Heck reaction constitutes an effective method for the formation of quaternary stereocenters. This reaction has been applied to functionalized 2-alkenylpyrroles as substrates in which the β-elimination is blocked by a methyl substituent. 10,10-Disubstituted pyrrolo[1,2-*b*]isoquinolines can be obtained using different palladium catalysts and chiral bidentate phosphanes as ligands, although with low enantioselectivities. In some cases, competition between Mizoroki-Heck reaction and C-H direct arylation reaction on the pyrrole nucleus has also been observed for this type of polyfunctionalized substrates. Finally, we have shown that quaternary stereocenters can be generated using chiral phosphane ligands as (*R*)-BINAP, through a cascade polyene cyclization. This procedure has been successfully applied to the construction of tetracyclic framework of Lycorine class of *Amaryllidaceae* alkaloids.

Keywords: Palladium; Heck reaction; cascade reactions; alkaloids; pyrrolo[1,2-*b*]isoquinolines

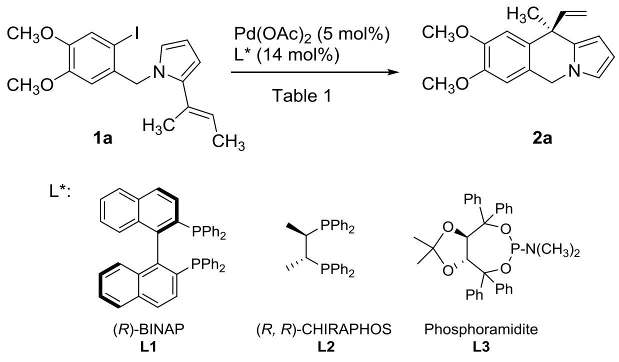
1. Introduction

The Mizoroki-Heck reaction of aryl and vinyl halides with alkenes has developed into one of the most important carbon-carbon bond-forming reactions in organic synthesis.[1] The intramolecular variant represents an extremely powerful method for the construction of small and medium-size rings. Particularly in recent years, this procedure has become an effective method for the formation of quaternary stereocenters, even in an asymmetric fashion.[2] The use of substrates where the β-elimination is blocked by a substituent allows to maintain the sp^3 centre formed after the migratory insertion, driving the elimination to another β' position, and generating a quaternary stereocenter.

In the context of our research program in palladium catalyzed reactions,[3] we have previously reported that pyrrolo[1,2-*b*]isoquinolines can be accessed via the intramolecular palladium-catalyzed Heck reaction of 2-alkenyl-substituted *N*-(o-iodobenzyl)pyrroles, avoiding the direct arylation on the pyrrole nucleus by choosing the appropriate catalytic system.[4] The procedure has also been applied to the selective synthesis of medium-sized rings and to (hetero)fused indolizine systems.[5] Thus, our next goal was to examine the possibility of generating quaternary stereocenters on C-10 of the pyrroloisoquinoline skeleton. For that purpose, we selected 2-alkenylpyrroles as substrates in which the β-elimination is blocked by a methyl substituent.

2. Results and Discussion

To start studying the generation of a quaternary center, we selected as substrate *N*-benzylpyrrole **1a** (Scheme 1).



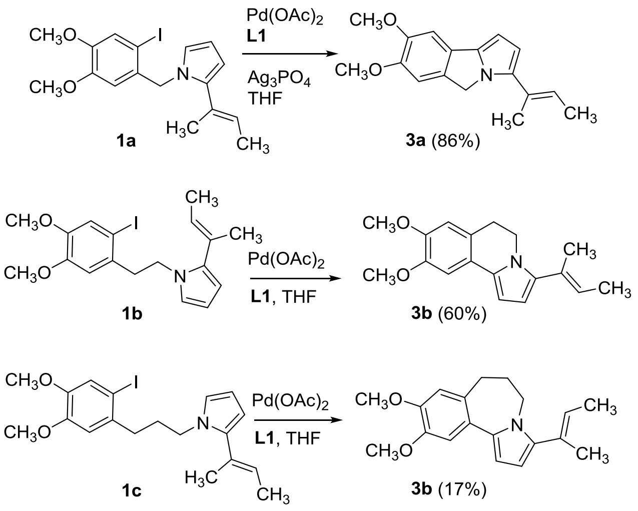
Scheme 1

The reaction was tested first in the racemic version. When **1a** was treated with $Pd(OAc)_2$ (5 mol%) in the presence of PPh_3 (14 mol%) in refluxing THF, racemic **2a** was obtained, although in a low yield (37%). The reaction conditions were optimized for the enantioselective version. Thus, three privileged ligands were selected: (*R*)-BINAP (**L1**), Chiraphos (**L2**) and phosphoramidite **L3** (Scheme 1). As shown on Table 1, when the reaction was carried out using (*R*)-BINAP in THF, pyrroloisoquinoline **2a** was obtained with a promising *ee* (78% ee), although in a very low yield (entry 1). The use of PMP as a base resulted in an increased yield, though with loss of enantioselectivity (entry 2). The use of Chiraphos (**L2**) gave almost no conversion (entries 3, 4), while phosphoramidite **L3** provided the opposite enantiomer, but with low yields and enantioselectivities. The change of the solvent to DMF did not improve the results.

. **Table 1.** Cyclization reactions of **1a**

Entry	**Base**	**Solvent**	**L***	**2a** Yield (%)	***ee*** (%)[a]
1	-	THF[b]	**L1**	5	78
2	PMP	THF[b]	**L1**	12	47
3	-	THF[b]	**L2**	8	2
4	PMP	THF[b]	**L2**	-	-
5	-	THF[b]	**L3**	12	-25
6	PMP	THF[b]	**L3**	16	-26
7	-	DMF[c]	**L1**	47	18
8	-	DMF[c]	**L2**	5	22
9	PMP	DMF[c]	**L3**	24	8

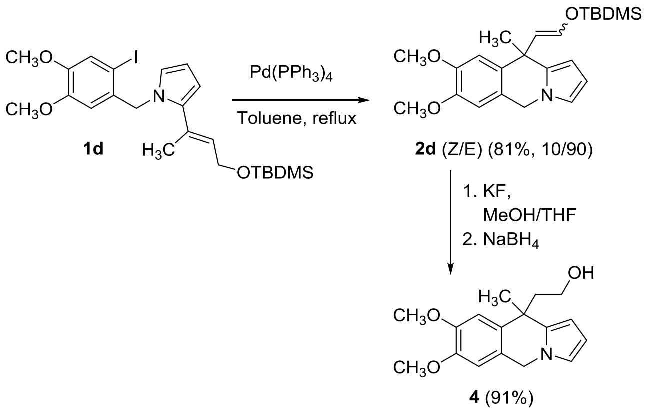
[a] Determined by Chiral Stationary Phase HPLC (Chiralcel OZ3 2% hexane/*i*-PrOH). [b] Reflux [c] 80 ºC.



Scheme 2

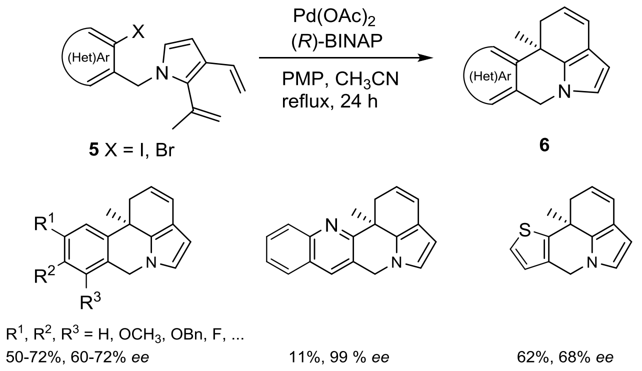
On the other hand, when silver salts were added to try to improve the enantioselectivity, **1a** afforded the pyrroloisoindol **3a**, through a direct C-H arylation reaction (Scheme 2). In a similar fashion, it was observed that the direct C-H arylation reaction predominates under all conditions tested when **1b** and **1c** were treated with Pd(OAc)$_2$, obtaining the pyrrolo-isoquinoline **3b** and pyrrolobenzazepine **3c**, not observing the formation of the Heck products, as shown in Scheme 2.

In view of these results, we decided to introduce a protected allylic alcohol moiety in the alkene, to favor the cyclization, as a result of the formation of an enol ether. Thus, when **1d** was treated with Pd(PPh$_3$)$_4$ in toluene, an excellent yield of the pyrroloisoquinoline **2d** was obtained, as a mixture of *Z/E* isomers (Scheme 3), forming a quaternary centre. The enol ether could be deprotected and the resulting aldehyde was reduced to afford the 10-hydroxymethyl derivative **4** in high yield. Unfortunately, when the cyclization was attempted using (*R*)-BINAP in different solvents, the yields of **2d** were moderate (38 to 65%) and the enantioselectivities were very poor (up to 11% *ee*).



Scheme 3

Finally, we reasoned that the formation of the quaternary center would be favored if the alkylpalladium is involved in a second reaction, in a cascade process Thus, we have recently shown that that quaternary stereocenters can be generated using chiral phosphane ligands as (*R*)-BINAP (**L1**), through a cascade polyene cyclization.[6] This procedure has been successfully applied to the construction of tetracyclic framework of Lycorine class of *Amaryllidaceae* alkaloids.[7] Thus, *N*-benzyl 2,3-dialkenyl pyrroles **5** undergo sequential 6-*exo*/6-*endo* cyclizations to yield enantiomerically enriched (11b*R*)-substituted pyrrolo-phenanthridines. (*R*)-BINAP (**L1**) has been shown to be the most efficient chiral phosphane ligand. The reaction can be extended to various substitution patterns on the aromatic ring, and also to heteroaromatic rings.



Scheme 4

4. Conclusions

Mizoroki-Heck reaction on 2-substituted *N*-benzylpyrroles in which the β-elimination is blocked by a substituent allows the generation of a quaternary stereocenter on C-10 position of the pyrrolo[1,2-*b*]isoquinoline skeleton. However, the reaction proceeds only with moderate yield and low enantioselectivity when chiral phosphanes (**L1-L3**) are used as ligands. The generation of larger rings is not possible, as the C-H direct arylation reaction predominates under all conditions tested. The generation of the quaternary stereocenter is favored when a protected allylic alcohol moiety is used obtaining high yields of the C-10 disubstituted pyrroloisoquinoline, but the reaction was not efficient when chiral phosphanes were used. Finally, the formation of the quaternary stereocenter is more efficient when the resulting alkylpalladium intermediate is involved in a cascade process. Thus, 2,3-dialkenyl pyrroles undergo sequential 6-*exo*/6-*endo* cyclizations to afford (11b*R*)-substituted pyrrolophenanthridines in good yields and enantioselectivities. This procedure allows a rapid and efficient access to a wide variety of enantiomerically enriched C-11b substituted lycorane analogues.

Acknowledgments

Ministerio de Economía y Competitividad (CTQ2013-41229-P), Gobierno Vasco (IT-623-13) and Universidad del País Vasco / Euskal Herriko Unibertsitatea UPV/EHU (UFI11/22, PPM12/03) are gratefully acknowledged for their financial support. We also thank Gobierno Vasco for grants (ARA, EC). Technical and human support provided by Servicios Generales de Investigación SGIker (UPV/EHU, MINECO, GV/EJ, ERDF and ESF) is also acknowledged.

.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. a) Oestreich M. Ed, *The Mizoroki-Heck reaction*, Wiley: Chichester, 2009; b) M. Larhed, Ed., *Science of Synthesis - Cross Coupling and Heck-Type Reactions,* Vol. 3, Thieme: Stuttgart, 2013.
2. a) Tietze, L. F.; Ila, H; Bell, H. P. *Chem. Rev.* **2004**, *104*, 3453-3516. b) McCartney, D.; Guiry, P. J. *Chem. Soc. Rev.* **2011***, 40*, 5122-5150.
3. a) Martínez-Estíbalez, U.; García-Calvo, O.; Ortiz-de-Elguea, V.; Sotomayor, N.; Lete E.; *Eur. J. Org. Chem.* **2013**, 3013-3022;; b) Ortiz-de-Elguea, V.; Sotomayor, N.; Lete, E. *Adv. Synth. Catal.* **2015**, *356*, 463-473
4. Lage, S.; Martínez-Estibalez, U.; Sotomayor, N.; Lete, E. *Adv. Synth. Catal.* **2009**, *351*, 2460-2468
5. Coya, E.; Sotomayor, N.; Lete E. *Adv. Synth. Catal.* **2014**, *356*, 1853-1865
6. Coya, E.; Sotomayor, N.; Lete E. *Adv. Synth. Catal.* **2015**, *357*, 3206-3214.

7. Jin, Z.; *Nat. Prod. Rep*. **2013**, *30*, 849-868, and previous reports on these series.

SciForum
Mol2Net

Diastereoselective Formation of Tertiary Stereocenters *via* Mizoroki-Heck Reaction

Ane R. Azcargorta,[1] Esther Lete,[1] and Nuria Sotomayor[1]*

[1] Departamento de Química Orgánica II, Facultad de Ciencia y Tecnología, Universidad del País Vasco / Euskal Herriko Unibertsitatea UPV/EHU. Apdo. 644. 48080 Bilbao, Spain. http://www.ehu.es/oms

* Author to whom correspondence should be addressed; E-Mail: nuria.sotomayor@ehu.eus; Tel.: +34 946015389; Fax: +34 946012748.

Published: 4 December 2015

Abstract: The diastereoselective Mizoroki-Heck reaction of *N*-benzylpyrrolidines that incorporate a protected allylic alcohol moiety allows the synthesis of enantiomerically pure pyrrolo[1,2-*b*]isoquinolines, generating a tertiary stereocenter. The best results were obtained with the use of bulky phosphanes, as P(*o*-Tol)$_3$. When a good leaving group, such as pivaloyl is used as a protecting group, the *trans*-10-vinyl substituted pyrroloisoquinoline (10S,10aS)-**2a** is obtained as the major diastereoisomer in moderate yield. On the other hand, when the allylic alcohol is protected as a silyl ether, the protected alcohol is retained, obtaining an enol ether, whichafter deprotection and reduction leads to the *trans*-10-hydroxymethyl substituted pyrrolisoquinoline (10*S*,10a*S*)-**5,** in enantiomerically pure form, with complete diastereoselectivity.

Keywords: Palladium; diastereoselective Heck reaction; alkaloids; pyrrolo[1,2-*b*]isoquinolines

Mol2Net YouTube channel*: http://bit.do/mol2net-tube*

1. Introduction

The Mizoroki-Heck reaction (M-H), has found wide application in the preparation of complex organic molecules, from simple substrates including heterocycles.[1] Particularly, the enantioselective intramolecular M-H reaction has emerged as an excellent tool for the construction of polycyclic frameworks.[2]

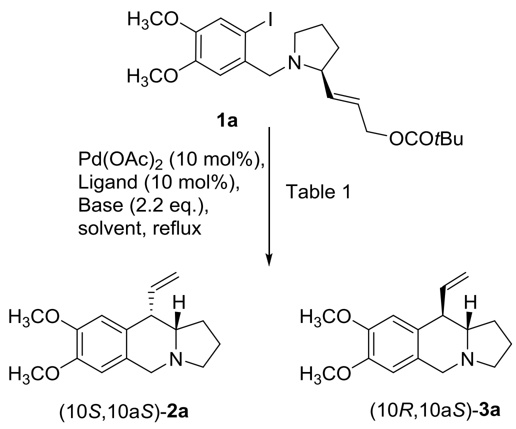
In connection with our interest our interest in palladium catalyzed reactions[3] we recently showed that quaternary stereocenters can be generated using chiral phosphane ligands as (*R*)-BINAP, through a cascade polyene cyclization.[4]

The pyrrolo[1,2-*b*]isoquinoline core is also the characteristic structural unit present in numerous biologically active compounds, as the phenanthroindolizidine alkaloids.[5] In this context, we have shown that the 6-*exo* carbolithiation of 2-alkenylpyrrolidines takes place with complete diastereoselectivity, allowing the synthesis of enantiomerically pure hexahydropyrrolo[1,2-*b*]isoquinolines in high yields.[6] On the other hand, the Mizoroki-Heck reaction of this type of pyrrolidines leads to enantiomerically pure 10-alkyidene substituted hexahydropyrrolo[1,2-*b*]isoquinolines7 Therefore, we decided to investigate further the scope of Mizoroki-Heck intramolecular reaction towards the stereo controlled synthesis of pyrrolo[1,2-*b*]isoquinolines, generating a tertiary stereocenter, using a diastereoselective approach.

.

2. Results and Discussion

To start studying the generation of a tertiary centre, we selected as substrate an enantiomerically pure *N*-benzylpyrrolidine that incorporates a protected allylic alcohol, as pivalate **1a**, which was prepared in enantiomerically pure form from commercially available *N*-Boc L-prolinal. (Scheme 1).



Scheme 1.

In fact, under classical Mizoroki-Heck conditions [Pd(PPh$_3$)$_4$ (10 mol%), NaHCO$_3$, Bu$_4$NCl, CH$_3$CN, reflux 48 h], pivalate elimination took place, generating a tertiary stereocenter. However, only a low yield (16%) of a diastereomeric mixture of pyrroloisoquinolines **2a** and **3a** was obtained, with moderate diastereoselectivity in favor of the *trans*-isomer **2a** (66:34 ratio). After some experimentation, we found that palladium acetate with a bulky phosphane, as tri-*ortho*-tolylphosphane (Scheme 1, Table 1) was required to obtain moderate to good yields of the diastereomeric mixture of pyrroloisoquinolines **2a** and **3a** (Table 1). The use of a mixture of CH$_3$CN/H$_2$O (10:1) as solvent resulted in reduced reaction times (72 h *vs* 5 h, entry 2 *vs* entry 1), obtaining a comparable yield with no loss of diastereoselectivity. Other phosphanes (entries 3-7) and bases (entry 8) were also used, but the diastereoselectivity was not improved. Both diastereomers could be separated and characterized. Their stereochemistry was established by NMR and confirmed by X-ray analysis (Figure 1)

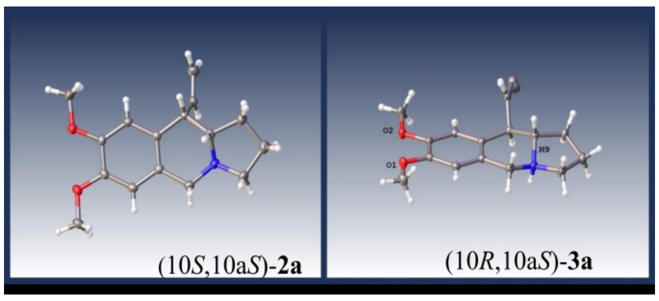
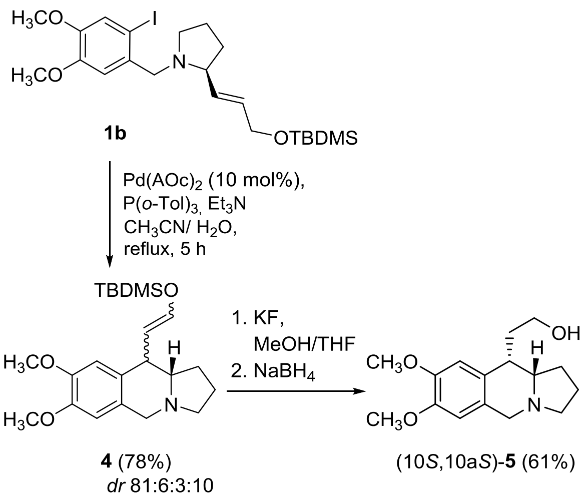


Figure 1. ORTEP plots of compounds **2a** and **3a**



Scheme 2.

Interestingly, when the allylic alcohol is protected with a TBDMS group (**1b**, Scheme 2), the protected alcohol is retained, generating a tertiary centre and obtaining an enol ether **4** in good yield, as a mixture of diastereomers with the *trans* major diastereomer as a *E*/*Z* mixture. The enol ether could be deprotected and the resulting aldehyde was reduced, to obtain alcohol **5** as a single diastereomer, enantiomerically pure.

Table 1. Pd(0)-catalyzed cyclization reactions of **1a**.

Entry	**Ligand**	**Base**	**Solvent**	**Time (h)**	**Yield (%)**	**Ratio 2a/3a**
1	P(*o*-Tol)$_3$	Et$_3$N	CH$_3$CN	72	51	83:17
2	P(*o*-Tol)$_3$	Et$_3$N	CH$_3$CN:H$_2$O	5	53	78:22
3	P*t*Bu$_3$	Et$_3$N	CH$_3$CN:H$_2$O	5	32	78:22
4	PCy$_3$	Et$_3$N	CH$_3$CN:H$_2$O	5	45	78:22
5	DavePhos	Et$_3$N	CH$_3$CN:H$_2$O	5	51	76:24
6	PPh$_3$	Et$_3$N	CH$_3$CN:H$_2$O	5	32	66:34
7	dppp	Et$_3$N	CH$_3$CN:H$_2$O	5	51	50:50
8	P(*o*-Tol)$_3$	BuNMe$_2$	CH$_3$CN:H$_2$O	5	35	72:28
9	P(*o*-Tol)$_3$[a]	Et$_3$N	CH$_3$CN:H$_2$O	22	46	79:21

[a] 5 mol% of Pd(AcO)$_2$ was used

4. Conclusions

Tertiary stereocenterss can be efficiently generated *via* Mizoroki-Heck reaction using protected allylic alcohol moieties. The β'-elimination can be controlled by selecting the protecting group (Piv or TBDMS). The best results in terms of yield and diastereoselectivity were obtained by using bulky phosphanes. Thus, *trans*-10-vinylpyrroloisoquinoline (10*S*,10a*S*)-**2a** and *trans*-10-hydroxymethylpyrroloisoquinoline (10*S*,10a*S*)-**5,** have been obtained in enantiomerically pure form.

Acknowledgments

Ministerio de Economía y Competitividad (CTQ2013-41229-P), Gobierno Vasco (IT-623-

13) and Universidad del País Vasco / Euskal Herriko Unibertsitatea UPV/EHU (UFI11/22, PPM12/03) are gratefully acknowledged for their financial support. We also thank Gobierno Vasco for a grant (A. R. A). Technical and human support provided by Servicios Generales de Investigación SGIker (UPV/EHU, MINECO, GV/EJ, ERDF and ESF) is also acknowledged.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Oestreich M. Ed, *The Mizoroki-Heck reaction*, Wiley: Chichester, 2009; b) Larhed, M., Ed., *Science of Synthesis - Cross Coupling and Heck-Type Reactions,* Vol. 3, Thieme: Stuttgart, 2013.
2. a) Tietze, L. F.; Ila, H; Bell, H. P. *Chem. Rev.* **2004**, *104*, 3453-3516. b) McCartney, D.; Guiry, P. J. *Chem. Soc. Rev.* **2011**, *40*, 5122-5150.
3. a) Lage, S.; Martínez-Estibalez, U.; Sotomayor, N.; Lete, E. *Adv. Synth. Catal.* **2009**, *351*, 2460-2468; b) Martínez-Estíbalez, U.; García-Calvo, O.; Ortiz-de-Elguea, V.; Sotomayor, N.; Lete E.; *Eur. J. Org. Chem.* **2013**, 3013-3022; c) Coya, E.; Sotomayor, N.; Lete E. *Adv. Synth. Catal.* **2014**, *356*, 1853-1865; d) Ortiz-de-Elguea, V.; Sotomayor, N.; Lete, E. *Adv. Synth. Catal.* **2015**, *356*, 463-473.
4. Coya, E.; Sotomayor, N.; Lete E. *Adv. Synth. Catal.* **2015**, *357*, 3206-3214.
5. For reviews, see: a) Burtoloso, A. C. B.; Bertonha, A. F.; Rosset, I. G. *Curr. Top. Med. Chem.* **2014**, *14*, 191-199; b) Chemler, S. R. *Curr. Bioact. Comp.* **2009**, *5*, 2-19
6. García-Calvo, O.; Coya, E.; Lage, S.; Coldham, I.; Sotomayor, N.; Lete E. *Eur. J. Org. Chem.* **2013**, 1460-1470.
7. García-Calvo, O.; Sotomayor, N.; Lete E.; Coldham, I. *Arkivoc* **2011**, 57-66.

SciForum
Mol2Net

Microwave Activated Synthesis of Benzalacetones and Study of Their Potential Antioxidant Activity Using Artificial Neural Networks Method

Anita Maria Rayar[1], Elizabeth Goya Jorge[2] Stephen Jones Barigye[3], María Elisa Jorge Rodríguez[2], Clotilde Ferroud[1] and Maite Sylla Iyarreta Veitía [1]*

[1] Equipe de Chimie Moléculaire du Laboratoire CMGPCE, EA 7341, Conservatoire national des arts et métiers, 2 rue Conté, 75003, Paris; anitarayar@hotmail.fr (A.M.R), clotilde.ferroud@cnam.fr (C.F.), maite.sylla@cnam.fr (M.S.-I.V.)

[2] Pharmacy Department, Faculty of Chemistry, Central University "Marta Abreu" of Las Villas, C-54830 Santa Clara, Cuba; egoyaj@gmail.com, elisajorge@yahoo.es

[3] Department of Chemistry, Federal University of Lavras, P.O. Box 3037, 37200-000 Lavras, MG, Brazil; sjbarigye@gmail.com

* Author to whom correspondence should be addressed; E-Mail: maite.sylla@cnam.fr;
Tel.: +33-1-58 80 84 82; Fax: +33-1-40 27 25 84.

Published: 4 December 2015

Abstract: The α,β-unsaturated ketones known as benzalacetones are an interesting class of compounds frequently used as key intermediates in organic synthesis. Due to their conjugated system, benzalacetone and derivatives have been described as radical scavengers with potential antioxidant properties. We report here a simple and direct method to prepare functionalized α,β-unsaturated ketones via a microwaves activated Claisen-Schmidt reaction. The experimental protocol developed selectively produces benzalacetones without self-condensation product in very short reaction times and good yields. Interested in the biological properties of benzalacetones, we also studied the antioxidant potential of these compounds using an *in silico study* based on the DPPH• radical scavenging ability. The built mathematical model was based on the 0-3D DRAGON molecular descriptors and the artificial neural networks technique showing a correlation coefficient for the training set $(R^2) = 0.71$, an external correlation coefficient $(Q^2_{ext}) = 0.65$. Unfortunately, the results obtained in the *in-silico* study revealed that synthesized benzalacetones have no antioxidant activity. The predicted results have been confirmed experimentally by an *in vitro* assay of DPPH• scavenging capacity.

Keywords: benzalacetones; artificial neural networks; antioxidant; QSAR; DPPH; free radical scavengers, microwaves-assisted synthesis

1. Introduction

As part of our studies involving the synthesis of bioactive compounds structurally related with the coumarin skeleton we have recently focused our attention on the synthesis of α,β-unsaturated ketones known as benzalacetones, which possess interesting properties for organic synthesis. Due to their conjugated system, benzalacetone and derivatives have been described as radical scavengers with potential antioxidant properties [1]. Various methods of synthesis for this type of compounds have been described in the literature. The Claisen-Schmidt is one of the simplest condensation methods. This reaction is typically catalyzed by acids ($AlCl_3$ or HCl) and more often by bases with or without solvent at room temperature or under conventional heating [2-4]. In order to increase the yield and to avoid the formation of by-products, several protocols relative to Claisen-Schmidt condensation have also been reported using different catalysts, sonochemical activation or microwaves irradiation. However, in all these conditions, side reactions start decreasing the yield of the desired product and entail further purification steps [5-20]. Consequently, we were particularly interested in developing an efficient preparation of benzalacetones from acetone and aromatic substituted aldehydes in basic conditions under microwave-activation.

Secondly we were interested in the study of the antioxidant activity of synthesized benzalacetones. To achieve our goal we developed an *in silico* study using the OD-3D DRAGON molecular descriptors (MDs) and the artificial neural networks (ANN) method. The ANN is one of the artificial intelligence techniques applied to Quantitative Structure Activity Relationships (QSAR) evaluations. In this study we develop an ANN in order to relate scavenging ability of the DPPH• radical and molecule features defined by established MDs. Finally the theoretical results were confirmed experimentally by the *in vitro* assay of DPPH• scavenging capacity.

2. Results and Discussion

2.1 Synthesis

Microwave activation for the synthesis of benzalacetones has not been widely described in the literature. Kappe et *al* reported the aldol condensation of *p*-methoxybenzaldehyde with acetone using microwave activation but could not prevent self-condensation. [21]

As a wide variety of aryl aldehydes is commercially available, microwave activation would provide a higher degree of flexibility with respects to functional groups which may be introduced in the benzalacetone skeleton. The details of the synthesis were previously described by our group **Figure 1**[22].

Following our interest in establishing an efficient, rapid and selective access to benzalacetones and considering our results previously obtained under conventional heating (dibenzalacetone formation in yields between 4-

39%), the Claisen-Schmidt condensations were carried out under controlled microwave activation. The reactions were performed with 1.5 equiv of NaOH, in a Discover^{TM} microwave synthesizer. The compounds were mixed in a sealed microwave reaction tube and irradiated for 10 to 30 minutes (5 W) with stirring at 50 °C or 40°C. After irradiation, reactions were controlled by GC-MS analysis, and the purity of the desired products was evaluated by NMR spectroscopy. All synthetic details of microwave activation procedure and the extension of these conditions to the synthesis of various benzalacetones have been previously described by our research group [22].

The use of microwave activation resulted in a dramatic decrease of reaction times. The reactions were generally achieved within 10-15 min. The desired compounds were isolated with excellent yields (typically higher than 79% and often quantitative) and clean enough to be further used without any purification. These microwave-assisted condensation reactions could be "directly scalable". Identical yields were obtained on 50 mg and 500 mg scale [22].

2.1 Modeling

The 0D-3D DRAGON MDs were computed for an in-house dataset of 1329 compounds whose DPPH• scavenging capacity has been experimentally determined and reported in the literature. Using a wrapper based variable selection procedure; a subset of 14 variables was obtained and posteriorly used as the ANN input. The mathematical model constructed showed a correlation coefficient (R^2) for the training set of 0.71. The predictive ability of the models was assessed using the external validation procedure yielding a correlation coefficient obtain (Q^2_{ext}) of 0.65. Both values are above the limits established for model acceptance, which is an indicator of the robustness and predictive power of the obtained MLP model.

2.2. Prediction:

Virtual screening allows for prior assessment of the potential bioactivity of chemical compounds, and thus providing key guidelines in posterior experimental work. In this study the MLP model previously obtained was used to predict the DPPH• scavenging capacity of a series of functionalized benzalacetones (Bz-der). The results of the predictions are shown in **Table 1**.

As it can be noticed the Bz-der seem to be less effective in DPPH• radical capturing, since their values of pIC_{50} are much higher than the reference compound, butylated hydroxytoluene (BHT, experimental pIC_{50} of 2.10). The exception is 1-naphtalene, in which the substituent is benzene fused to the benzalacetone moiety (pIC_{50} of 2.97).

According to conventional understanding of antioxidant activity, increasing of the number of phenolic hydroxyl groups enhances the compounds' antioxidant effectiveness. That's why many efforts have been carried out to synthesize antioxidants containing phenolic hydroxyl groups [23].

Consequently, we suggest that low antioxidant activity of the benzalacetones studied here could be due to the absence of hydroxyl groups linked to the aromatic system. On the other hand, several compounds used as references in the evaluation of antioxidant activity contain the phenolic group, for instance, Trolox, Gallic Acid and BHT.

2.3. <u>In vitro Assay</u>: The result obtained with the *in silico* modeling was corroborated using the experimental study of DPPH• scavenging capacity for the non-substituted compound, *(E)*-4-phenylbut-3-en-2-one whose pIC_{50} was previously predicted (4, 21), and experimentally obtained (4.97). These values of IC_{50} were indicative of very low scavenging activity, compared with the value obtained for BHT (2.10). The results demonstrate the predictive power of the designed ANN model and its possible applicability in the study of benzalacetones as antioxidant compounds. These results also suggest the need to synthesize benzalacetone derivatives bearing hydroxyl groups which would improve the antioxidant activity of these compounds

Figure 1: Preparation of benzalacetones under microwaves conditions

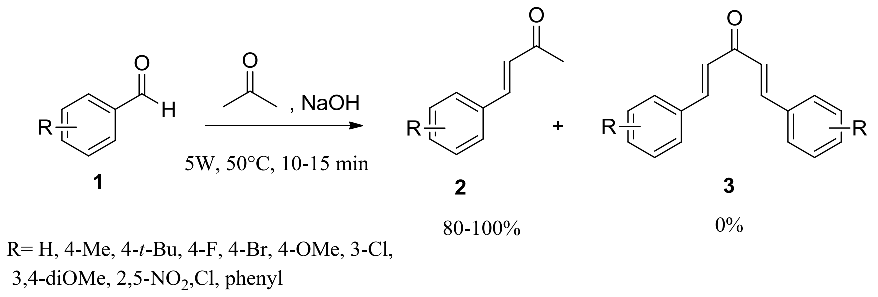
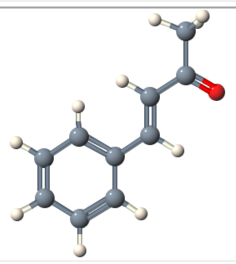
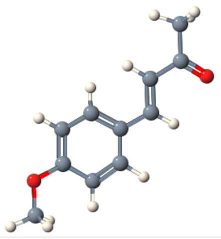
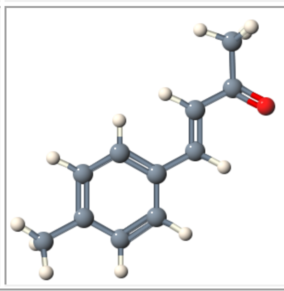
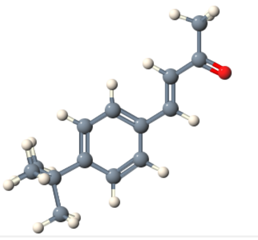
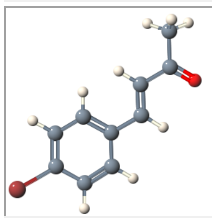
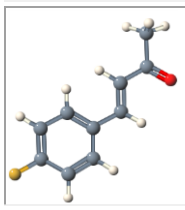
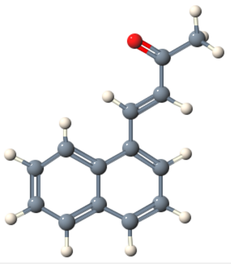

O
H
R
1
O
, NaOH
5W, 50°C, 10-15 min
O
R
2
80-100%
+
O
R
R
3
0%
R= H, 4-Me, 4-*t*-Bu, 4-F, 4-Br, 4-OMe, 3-Cl,
3,4-diOMe, 2,5-NO_2,Cl, phenyl

Table 1: Predictions of the pIC_{50} values for Bz-der

Origin	IUPAC name	3D Structures	Predict pIC50
Bz-der		Benzalacetone	4,216405
Bz-der		4-OMeBenzalacetone	3,997319
Bz-der		4-MeBenzalacetone	4,228023

Bz-der		4-tBuBenzalacetone	3,931096
Bz-der		4-BrBenzalacetone	4,138268
Bz-der		4-FBenzalacetone	4,276980
Bz-der		1-naphtalenebenzalacetone	2,973071

3. Materials and Methods

3.1 General procedure for the microwave-assisted syntheses

In a capped 10 mL MW-vessel, the aldehyde (50 mg, 1 equiv) and acetone (13.6 equiv) were mixed and then an aqueous solution of NaOH (0.6 g /cm^3 of water) was added. The tube was positioned in the irradiation cavity and the mixture was stirred and heated at the temperature 40/50 °C (measured with an IR temperature sensor), in the monomode microwave oven (5 W) for 10/15 min. After completion, upon cooling to room temperature, the conversion was directly controlled by GC-MS analysis. The product was extracted with AcOEt. The organic layers were dried over anhydrous $MgSO_4$, filtered and concentrated under reduced pressure to obtain the corresponding benzalacetone. The purity of the final products was controlled by NMR [22].

3.2 In silico study

Data: Experimental results of the scavenging ability of the DPPH• radical (expressed as IC_{50}) for 1329 molecules were extracted from over 170 scientific reports in the literature; and thus yielding a comprehensive and diverse database of compounds for the mathematical analysis. All the structures were optimized using the CORINA software. The response variable (IC_{50}) were transformed to their corresponding pIC_{50} values.

Molecular Descriptors: The parameterization of the structures was performed using 3224 molecular descriptors implemented in the DRAGON software. A wrapper based variable selection procedure was used to obtain a subset of variables for the ANN building.

Development of ANN model: The QSAR model was develop using as chemometric tool a Multilayer Perceptron Neural Network implemented in STATISTICA 8.0 software. For the modeling a Broyden-Fletcher-Goldfarb-Shanno algorithm was used as the optimization method; and the following network architecture was considered: fourteen inputs; eight neurons in hidden layer and one output.

Predictions of antioxidant activity: The benzalacetone derivatives were optimized following the same configurations previously used and the corresponding MDs computed.

3.3 In vitro DPPH• assay: The free radical scavenging activity of benzalacetone was measured using the stable DPPH radical, according to Blois' method [24]. Briefly, 0.1 mM solution of DPPH• in methanol was prepared and this solution (1 mL) was added to sample solution in methanol (3 mL) at different concentrations (250–1250 µg/mL). The mixture was shaken vigorously and left to stand for 30 min in the dark, and the absorbance was then measure at 517 nm. Butylated hydroxytoluene (BHT) was used for comparison. Both determinations were due in triplicate. The capability to scavenge the DPPH• radical was expressed as IC_{50} (concentration of antioxidant that produces 50% of absorbance inhibition)

.

4. Conclusions

In conclusion, an efficient and selective general method has been developed for the synthesis of benzalacetones via a Claisen-Schmidt reaction using microwaves activation. The desired compounds were obtained in shorter times and in almost all cases in quantitative yields. No further purification was required. Under microwaves activation, no dibenzalacetone formation has been observed except in the case of electron-withdrawing substituted aldehydes. Moreover, an ANN based model was developed to predict the antioxidant activity of the synthesized benzalacetones. Unfortunately, the *in-silico* study showed that benzalacetones do not have antioxidant activity. Finally, the predicted results have been experimentally confirmed by the *in vitro* assay of DPPH• scavenging capacity.

Acknowledgments

MENRT is gratefully acknowledged for the graduate fellowship awarded to Anita Rayar. The authors gratefully acknowledge support for part of this work through the program PYTHAGORAS II of EPEAEK II (MIS: 97436/073). We would like to thank Dr. C. Hansch and Biobyte Corp. 201 West 4th Street, Suite 204, Claremont, CA 91711, USA for free access to the C-QSAR program. The authors also thank the "Service de Coopération et d'Action Culturelle" of Embassy of France in Cuba for the financial support.

Author Contributions

The French team, A.M.R., M.S.-I.V. and C.F., is responsible for the synthesis and characterization of compounds, the Cuban team (E.G.R. and E.J.R.D.) and S.J.B. are responsible for the *in silico* study and the *in vitro* evaluation. All authors contributed to the drafting and revision of the article and approved the final version.

Conflicts of Interest

The authors declare no conflict of interest

References and Notes

1. Handayani S.; Arty I.S. Synthesis of Hydroxyl Radical Scavengers from Benzalacetone and its Derivatives. *J. Phys. Sci.,* **2008**, *19*; 61–68.
2. Drake N.L.; Allen P. Benzalacetone. *J.Org. Synth.* **1923**, 3:17.
3. Rahman A.F.M.M.; Ali R.;Jahng Y.; Kadi A.A.A Facile Solvent Free Claisen-Schmidt Reaction: Synthesis of α,α′-*bis*-(Substituted-benzylidene)cycloalkanones and α,α′-*bis*-(Substituted-alkylidene) cycloalkanones. *Molecules* **2012**, *17*, 571-583.
4. Jayapal, M.R.; Prasad, K.S. and Sreedhar N.Y., Synthesis and characterization of 2,6-dihydroxy substituted chalcones using PEG-400 as a recyclable solvent. *J. Pharm. Sci. Res*. **2010**, 2, 450-458.
5. Aguilera, A.; Alcantara, A. R.; Marinas, J. M.; Sinisterra, J. V. $Ba(OH)_2$ as the catalyst in organic reactions, part XIV: Mechanism of Claisen–Schmidt condensation in solid–liquid conditions. *Can. J. Chem.* **1987**, *65*, 1165–1171.
6. Narender, T.; Papi Reddy, K. A simple and highly efficient method for the synthesis of chalcones by using borontrifluoride-etherate. Tetrahedron Lett. **2007**, 48, 3177–3180.
7. Pal R. Ammonium chloride catalyzed microwave - assisted Claisen - Schmidt reaction between ketones and aldehydes under solvent - free conditions. IOSR J. App. Chem. (IOSR-JAC) **2013**, 3, 74-80.
8. Kumar D.S.; Sandhu J.S. An efficient green protocol for the synthesis of chalcones by a Claisen – Schmidt reaction using bismuth(III)chloride as a catalyst under solvent-free condition. *Green Chem. Lett. Rev.* **2010**, 3, 283-286.
9. Deng G.; Ren T. Indium Trichloride Catalyzes Aldol-Condensations of Aldehydes and Ketones. Synth. Commun. **2003**, *33*, 2995–3001.
10. Iranpoor N.; Zeynizadeh B.; Aghapour A. Aldol Condensation of Cycloalkanones with Aromatic Aldehydes Catalysed with $TiCl_3(SO_3CF_3)$. *J. Chem. Res.* **1999**, S, 554–555.
11. Wang L.; Sheng J.; Tian H.; Han J.; Fan Z.; Qian C. A convenient synthesis of α, α'-bis (substituted benzylidene) cycloalkanones catalyzed by Yb $(OTf)_3$ under solvent-free conditions. *Synthesis* **2004**, 3060-3064.
12. Zhang X.; Fan X.; Niu H.; Wang J. An ionic liquid as a recyclable medium for the green preparation of a, a'-bis (substituted benzylidene)cycloalkanones catalyzed by $FeCl_3$ $6H_2O$. *Green Chem.* **2003**, *5*, 267–269.
13. Sashidhara K.V.; Rosaiah J.N.; Kumar A. Iodine catalyzed mild and efficient method for the synthesis of chalcones. *Synth. Commun.* **2009**, *39*, 2288–2296.
14. Irie K.; Watanabe K-I. Catalysis of metal (II) acetate-2,2'-bipyridine complexes in the aldol condensations. *Bull. Chem. Soc. Jpn.* **1981**, *54*, 1195–1198.
15. Esmaeili AA; Tabas MS; Nasseri MA; Kazemi F. Solvent-Free Crossed Aldol Condensation of Cyclic Ketones with Aromatic Aldehydes Assisted by Microwave Irradiation. *Monatsh. Chem.* **2005**, *136*, 571–576.
16. Bogdal D.; Loupy A. Application of Microwave Irradiation to Phase-Transfer Catalyzed Reaction. *Org. Process. Res. Dev.* **2008**, *12*, 710–722.

17. Zheng M.; Wang L.; Shao J.; Zhong Q. A Facile Synthesis of α, α'-bis(Substituted Benzylidene)cycloalkanones Catalyzed by bis(p-ethoxyphenyl)telluroxide(bmpto) Under Microwave Irradiation. *Synth. Commun.* **1997**, *27*, 351–354.
18. Nakano T.; Irifune S.; Umano S.; Inada A.; Ishii Y.; Ogawa M. Cross-Condensation Reactions of Cycloalkanones with Aldehydes and Primary Alcohols under the Influence of Zirconocene Complexes. *J. Org. Chem.* **1987**, *52*, 2239–2244.
19. Pal R.; Sarkar T.; Khasnobis S (2012) Amberlyst-15 in organic synthesis *Arkivoc* i: 570–609.
20. Gladkowski W.; Skrobiszewski A.; Mazur M.; Siepka M.; Pawlak A.; Obminska-Mrukowicz B.; Bialonska A.;Poradowski D.; Drynda A.; Urbaniak, M. Synthesis and anticancer activity of novel halolactones with β-aryl substituents from simple aromatic aldehydes. *Tetrahedron* **2013**, *69*, 10414–10423.
21. Viviano M; Glasnov TN; Reichart B.; Tekautz G.; Kappe C.O. A Scalable Two-Step Continuous Flow Synthesis of Nabumetone and Related 4-Aryl-2-butanones. *Org. Process. Res. Dev*. **2011**, 15, 858–870.
22. Rayar A.; Sylla-Iyarreta Veitía M.; Ferroud, C. An efficient and selective microwave-assisted Claisen-Schmidt reaction for the synthesis of functionalized benzalacetones. *Springer Plus* **2015**, *4*, 221-226.
23. Gao-Lei X.; Zai-Qun L. Antioxidant effectiveness generated by one or two phenolic hydroxyl groups in coumarin-substituted dihydropyrazoles. *Eur. J. Med. Chem.* **2013**, *68*, 385-393.
24. Blois M.S. Antioxidant determinations by the use of a stable free radical. *Nature*. **1958**, 181, 1199–1200.

SciForum
Mol2Net

Computational Models of the Brain

L.A. Pastur-Romay[1], F. Cedrón[1], A. Pazos[12] & A. B. Porto-Pazos[12]*

[1] Department of Information and Communications Technologies, University of A Coruña, A Coruña, 15071, Spain; E-Mails: pastur90@gmail.com, flanciskinho@gmail.com
[2] Instituto de Investigación Biomédica de A Coruña (INIBIC), Complexo Hospitalario Universitario de A Coruña (CHUAC), A Coruña, 15006, Spain; E-Mail: alejandro.pazos@udc.es;
* Author to whom correspondence should be addressed; E-Mail: ana.portop@udc.es;
Tel.: +34-881-011-380; Fax: +34-981-167160.

Published: 4 December 2015

Abstract: Different research projects around the world are trying to emulate the human brain. They employ diverse types of computational models: digital models, analog models and hybrid models. This communication includes a summary of some main projects, as well as future trends in this subject. It is focused on various works that look for advanced progress in Neuroscience and still others which seek new discoveries in Computer Science (neuromorphic hardware, machine learning techniques). In addition, given the proven importance of glial cells in information processing, the importance of considering astrocytes into the brain computational models is pointed out.

Keywords: brain emulation, neuromorphic chip, neuron-astrocyte computational models, brain computational models

1. Introduction

The first computational brain models were created with the goal of reproducing this extraordinary organ, in order to understand and mimic the way the information is processed, as well as its energy efficiency [1-9]. From these works, basically two scientific disciplines emerge: the connectionism branch of Artificial Intelligence, which is aimed at developing algorithms based on neural networks to process the information, and Computational Neuroscience which seeks to create realistic models of the brain. In the seventies the field of Brain Machine Interface (BMI) also emerged, whose purpose was to create systems that connected the brain directly to an external device. At the same time, a branch of

Neuroscience, known as Neuroprosthetics, was formed, which sought to build artificial devices to replace the functions of nervous systems which are damaged in patients. At the end of the eighties, Carver Mead [10, 11] proposed the concept of Neuromorphic Engineer to describe the use of Very Large Scale Integration (VLSI) systems which contained analog circuits to mimic the neurons.

All these scientific disciplines have tried to model the brain in one way or another. Over the past century, many experts in these fields have predicted that in 10 or 20 years a computational system comparable to the human brain would be built. But all these predictions had failed because of the technological limitations and the underestimation of the brain capacity. Although IBM ran the first simulation with approximately the same number of neurons as the human brain, the neuron models were very simple and the simulation was x1542 times slower than in real time [12]. However, it should be pointed out that until now in most computational brain models the capacity to process the information from the other half of the brain, containing 84 billion glial cells [13], has not been taken in consideration. According to the Neural Doctrine, neurons are the only cells in the nervous system involved in information processing, and the glial cells only play a support role. But over the past two decades this theory has started to be seriously debated. Some discoveries have demonstrated the capacity of the glial cells to participate in information processing [14-17]. In this communication, some works focused on implementing artificial astrocytes in the brain models are referenced.

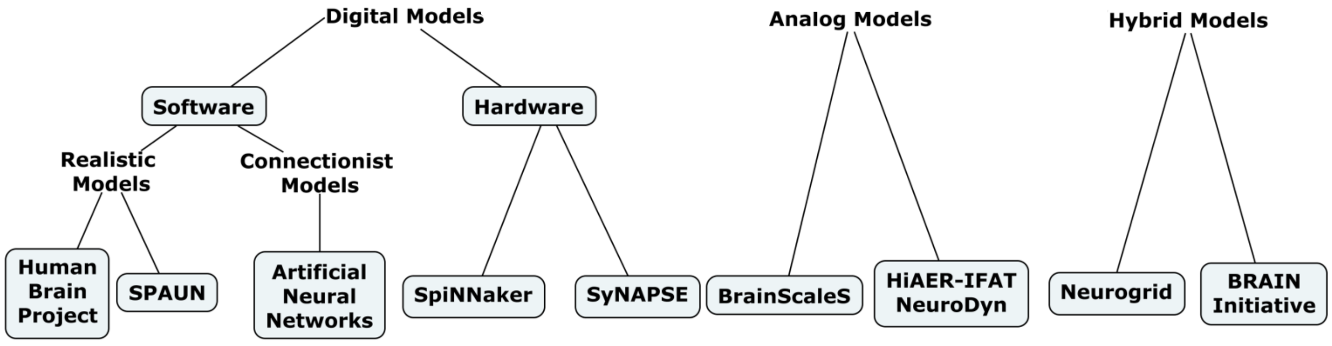


Figure 1. Brain Model classification.

2. Models Classification

Classifications of brain models can be performed from different perspectives. In this communication, different computer models that have been classified from the point of view of signal processing by hardware are currently under development, such as: digital models, analog models and hybrid models.

This classification is shown in Figure 1.

- Digital models: they compute information using the binary system to simulate and parallelize the behavior of the brain cells. From the software models, the realistic computer models are first considered, which are those shaping the internal structure of the cells (ion channels, organelles, etc.) allowing the study of their functions/operations. The generation of action potentials, activation of neurons, and synapse creation are simulated by mathematical equations implemented in the software, with specifically-designed tools. In addition, the connectionist models are taken into account, which, given a known behavior is expected to be achieved, such as a classification, object recognition in images, regression, etc., allow searching for a structure of artificial process elements (neurons and/or astrocytes) that give sufficient rise to such behavior. With regard to digital hardware models, they propose new computer architectures based on brain functioning.

- Analog models: they consist of neuromorphic hardware elements where information is processed with analog signals, that is, they do not operate with binary values, as information is processed with continuous values. This allows computation to be more efficient, so that analog computation could be used in applications where energy efficiency is very important.
- Hybrid models: they have been classified as such those assembled using hardware composed of both analog and digital components. These models seek to make the most of each type of computer.

Projects		Name	Institution	Num. neurons	Type of neurons	Simulated synapses	Objectives	Duration	Refs.
Digital Models	**Software**	Human Brain Project	European Union	10^6	Hogdking & Huxley	$5x10^8$	1, 2, 3, 4	2005	18
		SPAUN	Univ. Waterloo	$2.5x10^6$	Leaky integrate-and-fire	10^{12}	1	2012	19
	Hardware	SpiNNaker	Univ. Manchester	$2.5x10^5$	Point neuron models, leaky integrate-and-fire, Izhikevich's models	$8x10^7$	1, 2, 3, 4	2005	20
		SyNAPSE	IBM	10^{11}	Improved leaky integrate-and-fire.	10^{14}	2, 3	2008	21
Analog Models		BrainScales	European Union	$4x10^6$	Adaptive exponential integrate and fire neurons	10^9	1, 2, 3	2011-2015	22
		HiAER-IFAT	Univ. California at San Diego	250.000	Integrate-and-fire with two compartments for neuron	$5x10^6$	1, 2, 3, 4	2004	23-27
		NeuroDyn	Univ. California at	4	Hogdking & Huxley. 384	12	1	2004	23-27

		San Diego		parameters and 24 channels.				
Hybrid Models	Neurogrid	Stanford University	10^6	Quadratic integrate-and-fire somatic compartment + Dendritic compartment model with 4 Hogdking & Huxley channels	10^9	1, 3, 4	2007	28
	BRAIN Initiative	Qualcomm	not public	not public	not public	1, 2, 3	2013	29

Table 1. Overview of key features of relevant projects. Objectives (1. Computational Neuroscience; 2. Artificial Intelligence; 3. Neuromorphic chips; 4. Build devices to help disable people).

3. Characteristics of the models

Table 1 shows an overview of key features of main projects around the world that model the brain. In this table they are grouped according to the classification referred to:

Project name: it usually contains words like 'neuron', 'spike' or 'brain'.

Institution: it is observed that most institutions are universities, but there are some projects developed in companies, such as IBM (SyNAPSE) or QUALCOMM (BRAIN Initiative). Most modeling works are coordinated by groups of the prestigious American universities, like Stanford (Neurogrid). There are also projects coordinated in prestigious European universities like University of Lausanne (HBP) or University of Manchester (SpiNNaker). The projects developed in European universities are mainly supported through the European Union, while in the case of US projects, funding comes from DARPA and NIH (National Institutes of Health). The most important difference between European and American projects is that Europeans try to increase scientific knowledge about the brain. However, the major American projects are rather focused on carrying out a revolution in the computer industry, laying the foundation for future computer systems.

Number of neurons: the simulation with the largest number of neurons was made by the SyNAPSE project in 2012 with 5.4×10^{11} neurons, a quantity even higher than a human brain, which is around 8.6×10^{10} [13]. It should be noted that this simulation is not expected to be realistic and uses very simplified neuronal models. Furthermore, the simulation runs 1542 times slower than real time and 1.5 million BlueGene/Q cores [30] were necessary.

Types of Neurons: there are many types of neuronal models with different levels of realism and complexity. These implementations can be either software or hardware-based. When it comes to software connectionist models, artificial neurons are simple processing elements which operate following sigmoid or threshold mathematical functions [31], although there are progressively more software models using built-in spiking neurons [32] that simulate action potentials. In the case of realistic models, the latter usually present ion channels responsible for the spike generation. The Hodgkin-Huxley model [9] requires more computational resources because it simulates Ca+2, K+ and Na+ currents. It is used in the HBP [18], Neurogrid [28] and NeuroDyn [26]. When the 3D arrangement of axons and dendrites is considered, the simulation becomes significantly more complicated, as a space-time integration is necessary. For the sake of simplicity, Rall's Cable Theory [33] and compartment models [34] are used. For more information about these models, please refer to [35]. The simplest model is "Integrate-and-fire point neuron", which adds the inputs to the associated weights and compares the sum to a threshold, resulting in a binary decision of either generating a spike output or not. There is an extension of this model that uses a charge decay, known as "leaky integrate and fire". It is used for example in SPAUN [19], SpiNNaker [20] and SyNAPSE [21]. Other ways to improve the models are: non-linear sum, time-dependent threshold, programmable delay in the release of the spikes and other variations.

Simulated synapses: in 2012 the SyNAPSE project achieves $1.37x10^{14}$ simulated synapses [12], roughly the same number as in the human brain. A problem encountered by the models is the synaptic connectivity because of the large number of existing connections in the brain. In addition, the connections between neurons are formed during development, but they change daily to allow learning. To date, the most common solution involves using networks with AER architecture [36, 37] that make neurons communicate only when they need to send a spike. The information is sent in a package that contains only the address of the neuron that fires the spike. The synaptic connectivity is stored in tables that are used by the network routers [38]. In analog models, the nearby connections between neurons are usually done through a direct cable. However, for long-distance connections AER is necessary, for which Analog/Digital and Digital/Analog converters are employed. This is a problem because the circuit that the neuron needs for conversion and routing is much larger than the neuron circuit itself. The brain modeling projects use supercomputer and CPU [39] or GPU clusters [40]. Moreover, others use neuromorphic chips specifically designed to process information emulating the brain, both digital (SpiNNaker [20], TrueNorth [21]) and analog (HICANN [22]), and even hybrid (Neurogrid [28], Zeroth [41]). One of the advantages of the neuromorphic systems is that, as they are implemented within the hardware, they eliminate the overhead of the simulation software, providing a more accurate output in a shorter space of time. Furthermore, the emulation speed and communication in neuromorphic solutions can be run faster than the biological equivalent. Another advantage of the neuromorphic solutions is that they have a lower consumption per emulated neuron. Although the analog model is faster, it has not been shown that its fixed neural structure adequately captures biological neural behavior.

Project duration: these are very complex modeling projects and works and, therefore, their time span is long. The case of Blue Brain Project should be pointed out, which began in 2005 and later became part of the Human Brain Project which is still underway. The older projects (started 10 or more years ago) include: Spinnaker, HiAER-IFAT or NeuroDyn. As seen in Table 1, the most recent is SPAUN. All of them are still under development, except for FACETS and BrainScales.

Objectives: most brain models are not completed, although some projects have already built parts of them that have been applied to certain fields or specific studies. On the one hand, the projects which are mainly focused on understanding some aspects of the brain were divided as follows: HBP is trying to simulate the effect of new drugs for brain diseases; SPAUN is testing neuroscientific hypotheses related to behavior studies; and the Neurogrid project is aimed at figuring out how cognition arises. On the other hand, there are models which allow automatic processing of large amounts of data using intelligent software (SyNAPSE, SpiNNaker). There are also projects that develop new processing hardware architectures, such as BrainScales, SpiNNaker, SyNAPSE. Finally, there are also some which allow even building devices to help disabled people, as in the case of the SpiNNaker project.

References: fundamental webs or papers, where the projects were presented.

4. Brain Computational Models with Glia

So far there were no projects including astrocytes in a neuromorphic chip. There are only realistic computational models [42-49] and connectionist ones [50-54] which have taken glial cells into account. Currently, there are two projects aimed at implementing astrocytes in neuromorphic chips, one is BioRC [55-57] developed by the University of Southern California and the other project is carried out by the University of Tehran and University of Kermanshah (Iran) [58-60]. Moreover, there is a project under development at the University of A Coruña, which extends classical ANN by incorporating recent findings and suppositions regarding the way information is processed via neural and astrocytic networks in the most evolved living organisms. Considering the works published over the past two decades on the multiple modes of interaction between neurons and glial cells [14-17], it would be a very interesting approach if most of these groups tried to implement these behaviors in computer models. In addition, it is worth noting that glial cells have evolved more than neurons. For example, in mammals there are no major differences between neurons of different species. However, a rodent’s astrocytes may include between 20,000 and 120,000 synapses, while a human’s may include up to 2 million synapses [61, 62]. Furthermore, the ratio between neurons and glial cells varies in different brain regions. In the cerebellum, for instance, there are almost 5 times more neurons than astrocytes. However, in the cortex, there are 4 times more glial cells than neurons [13]. All these data suggest that the more complex the task, performed by either an animal or a brain region, the greater number of glial cells is involved.

4. Conclusions

There are a great variety of projects and models of the brain around the world. The development of digital, analog and hybrid models is expedient and allows for advances in Neuroscience and Computer Science.

With regard to the cerebral phenomena emulated by computer models, the importance of considering the glial system should be stressed. Such system is crucial for the development of complex cognitive capacities of human beings. Therefore, it should be part of brain models to be truly realistic.

In the short and medium term, the modeling of the brain and neuromorphic chips will advance the development of prosthetic devices and Brain-Machine Interface. However, all the brain simulations that will be performed within this period will use very simplified models. It is therefore questionable that the whole brain could be analyzed through realistic simulations.

In the long term, it is more difficult to make predictions about the brain simulations, as their approach is rather philosophical than scientific. The question of creating an artificial brain is old, but today there is a clear division between scientists who believe it is possible, and could even be accomplished within the next two decades, and those who believe it will never be possible.

Acknowledgments

This work is supported by the General Directorate of Culture, Education and University Management of Xunta de Galicia (Ref. GRC2014/049) and the European Fund for Regional Development (FEDER) in the European Union, the Galician Network for Colorectal Cancer Research (REGICC) funded by the Xunta de Galicia (Ref. R2014/039) and by the "Collaborative Project on Medical Informatics (CIMED)" PI13/00280 funded by the Carlos III Health Institute from the Spanish National plan for Scientific and Technical Research and Innovation 2013–2016 and the European Regional Development Funds (FEDER).

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. McCulloch, W. S.; Pitts, W. A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bull. Math. Biophys.* **1943**, 5 (4), 115–133.
2. Turing, A. Computing Machinery and Intelligence. *Mind* **1950**, 49, 433–460.
3. Minsky, M. Steps toward Artificial Intelligence. Proc. IRE **1961**, 49 (1), 8–30.
4. Minsky, M.; Papert, S. Perceptrons**. 1969.**
5. Minsky, M.; Papert, S. Artificial Intelligence Progress Report. **1972.**
6. Von Neumann, J. The General and Logical Theory of Automata. *Cereb. Mech. Behav*. **1951**, 1–41.
7. Hebb, D. The Organization of Behaviour: A Neuropsychological Theory. **1949.**
8. Rosenblatt, F. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psichol. Rev.* **1958**, 65 (6), 386.
9. Hodgkin, A. L.; Huxley, A. F. A Quantitative Description of Membrane Current and Its Application to Conduction and Excitation in Nerve*. J. Physiol.* **1952**, 117 (4), 500–544.
10. Sivilotti, M. A.; Emerling, M. R.; Mead, C. A. VLSI Architectures for Implementation of Neural Networks, **1986.**
11. Mead, C. Analog VLSI and Neural Systems; Addison-Wesley, **1989.**
12. Wong, T. M.; Preissl, R.; Datta, P.; Flickner, M.; Singh, R.; Esser, S. K.; Mcquinn, E.; Appuswamy, R.; Risk, W. P.; Simon, H. D.; Modha, D. S.; Jose, S.; Berkeley, L. 10^{14} *IBM Journal Report.* **2012**, 10502, 13–15.

13. Azevedo, F. A. C.; Carvalho, L. R. B.; Grinberg, L. T.; Farfel, J. M.; Ferretti, R. E. L.; Leite, R. E. P.; Jacob Filho, W.; Lent, R.; Herculano-Houzel, S. Equal Numbers of Neuronal and Nonneuronal Cells Make the Human Brain an Isometrically Scaled-up Primate Brain. *J. Comp. Neurol*. **2009**, 513 (5), 532–541.
14. Fields, R. D. The Other Brain: From Dementia to Schizophrenia, How New Discoveries about the Brain Are Revolutionizing Medicine and Science; Simon and Schuster, **2009.**
15. Koob, A. The Root of Thought: Unlocking Glia--the Brain Cell That Will Help Us Sharpen Our Wits, Heal Injury, and Treat Brain Disease; FT Press, **2009**.
16. Fields, R. D.; Araque, A.; Johansen-Berg, H.; Lim, S.-S.; Lynch, G.; Nave, K.-A.; Nedergaard, M.; Perez, R.; Sejnowski, T.; Wake, H. Glial Biology in Learning and Cognition. *Neuroscientist* **2014**, 20 (5), 426–431.
17. Perea, G.; Sur, M.; Araque, A. Neuron-Glia Networks: Integral Gear of Brain Function. *Front. Cell. Neurosci*. **2014**, 8, 378
18. Human Brain Project: https://www.humanbrainproject.eu/ (Accessed August 10, **2015**)
19. Spaun: http://models.nengo.ca/spaun (Accessed August 23, **2015**)
20. SpiNNker Wiki: https://spinnaker.cs.manchester.ac.uk/tiki-index.php (Accessed August 10, **2015)**
21. IBM Research. Cognitive Computing. Neurosynaptic chips. http://www.research.ibm.com/cognitive-computing/neurosynaptic-chips.shtml#fbid=IcooSM9Q3jV (Accessed August 10, **2015**)
22. BrainScaleS: http://brainscales.kip.uni-heidelberg.de/ (Accessed August 10, **2015**)
23. Park, J.; Yu, T.; Maier, C.; Joshi, S.; Cauwenberghs, G. Live Demonstration: Hierarchical Address-Event Routing Architecture for Reconfigurable Large Scale Neuromorphic Systems. In Circuits and Systems (ISCAS), 2012 IEEE International Symposium on; IEEE, 2012; pp 707–711.
24. Joshi, S.; Deiss, S.; Arnold, M.; Park, J.; Yu, T.; Cauwenberghs, G. Scalable Event Routing in Hierarchical Neural Array Architecture with Global Synaptic Connectivity. In Cellular Nanoscale Networks and Their Applications (CNNA), 2010 12th International Workshop on; IEEE, 2010; pp 1–6.
25. Yu, T.; Park, J.; Joshi, S.; Maier, C.; Cauwenberghs, G. 65k-Neuron Integrate-and-Fire Array Transceiver with Address-Event Reconfigurable Synaptic Routing. In Biomedical Circuits and Systems Conference (BioCAS), 2012 IEEE; IEEE, **2012**; pp 21–24.
26. Yu, T.; Cauwenberghs, G. Analog VLSI Biophysical Neurons and Synapses with Programmable Membrane Channel Kinetics. *IEEE Trans. Biomed. Circuits Syst*. **2010**, 4 (3), 139–148.
27. Al-Shedivat, M.; Naous, R.; Cauwenberghs, G.; Salama, K. N. Memristors Empower Spiking Neurons With Stochasticity. *IEEE J. Emerg. Sel. Top. Circuits Syst*. 2015, 5 (2), 242–253.
28. Stanford University. Brains in silicon: http://web.stanford.edu/group/brainsinsilicon/neurogrid.html (Accessed August 10, **2015**)
29. BRAIN Initiative: http://www.braininitiative.nih.gov/ (Accessed August 10, **2015**)
30. IBM Blue Gene/Q: http://www-03.ibm.com/systems/technicalcomputing/solutions/bluegene/ (Accessed September 2, **2015**)

31. Basheer, I. A.; Hajmeer, M. Artificial Neural Networks: Fundamentals, Computing, Design, and Application. *J. Microbiol. Methods* **2000**, 43 (1), 3–31.
32. Brette, R.; Rudolph, M.; Carnevale, T.; Hines, M.; Beeman, D.; Bower, J. M.; Diesmann, M.; Morrison, A.; Goodman, P. H.; Harris Jr, F. C. Simulation of Networks of Spiking Neurons: A Review of Tools and Strategies. *J. Comput. Neurosci*. **2007**, 23 (3), 349–398.
33. Rall, W. Branching Dendritic Trefaes and Motoneuron Membrane Resistivity. *Exp. Neurol*. **1959**, 1 (5), 491–527.
34. Herz, A. V. M.; Gollisch, T.; Machens, C. K.; Jaeger, D. Modeling Single-Neuron Dynamics and Computations: A Balance of Detail and Abstraction. *Science*. **2006**, 314 (5796), 80–85.
35. Koch, C.; Segev, I. Methods in Neuronal Modeling: From Ions to Networks; MIT press, **1998**.
36. Sivilotti, M. A. Wiring Considerations in Analog VLSI Systems, with Application to Field-Programmable Networks, 1990.
37. Computational Intelligence and Bioinspired Systems; Cabestany, J., Prieto, A., Sandoval, F., Eds.; *Lecture Notes in Computer Science*; Springer Berlin Heidelberg: Berlin, Heidelberg, **2005**; Vol. 3512.
38. Cattell, R.; Parker, A. Challenges for Brain Emulation: Why Is Building a Brain so Difficult? **2012**, 1–28.
39. The Blue Brain Project EPFL, in silico experiments: http://bluebrain.epfl.ch/page-58125-en.html (Accessed August 6, 2015)
40. Vasilache, N.; Johnson, J.; Mathieu, M.; Chintala, S.; Piantino, S.; LeCun, Y. Fast Convolutional Nets With Fbfft: A GPU Performance Evaluation. **2014.**
41. Qualcomm Zeroth Platform, cognitive technologies: https://www.qualcomm.com/invention/cognitive-technologies/zeroth (Accessed August 30, **2015**)
42. Linne, M.-L.; Havela, R.; Saudargienė, A.; McDaid, L. Modeling Astrocyte-Neuron Interactions in a Tripartite Synapse. *BMC Neurosci*. **2014**, 15 (Suppl 1), P98.
43. Tewari, S. G.; Majumdar, K. K. A Mathematical Model of the Tripartite Synapse: Astrocyte-Induced Synaptic Plasticity. *J. Biol. Phys*. **2012**, 38 (3), 465–496.
44. De Pitta, M.; Volman, V.; Berry, H.; Parpura, V.; Volterra, A.; Ben-Jacob, E. Computational Quest for Understanding the Role of Astrocyte Signaling in Synaptic Transmission and Plasticity. *Front. Comput. Neurosci.* **2012***, 6.*
45. Min, R.; Santello, M.; Nevian, T. The Computational Power of Astrocyte Mediated Synaptic Plasticity. *Front. Comput. Neurosci*. **2012**, 6.
46. De Pittà, M.; Volman, V.; Berry, H.; Ben-Jacob, E. A Tale of Two Stories: Astrocyte Regulation of Synaptic Depression and Facilitation. *PLoS Comput. Biol* 2011, 7 (12), e1002293.
47. Wade, J. J.; McDaid, L. J.; Harkin, J.; Crunelli, V.; Kelso, J. A. Bidirectional Coupling between Astrocytes and Neurons Mediates Learning and Dynamic Coordination in the Brain: A Multiple Modeling Approach. *PLoS One* **2011**, 6 (12), e29445.
48. Wade, J.; McDaid, L.; Harkin, J.; Crunelli, V.; Kelso, S. Self-Repair in a Bidirectionally Coupled Astrocyte-Neuron (AN) System Based on Retrograde Signaling. *Front. Comput. Neurosci*. **2012**, 6.

49. Wallach, G.; Lallouette, J.; Herzog, N.; De Pittà, M.; Jacob, E. Ben; Berry, H.; Hanein, Y. Glutamate Mediated Astrocytic Filtering of Neuronal Activity. *PLoS Comput. Biol*. **2014**, 10 (12), e1003964.
50. Porto, A.; Pazos, A.; Araque, A. Artificial Neural Networks Based on Brain Circuits Behaviour and Genetic Algorithms. In Computational Intelligence and Bioinspired Systems; Springer, 2005; pp 99–106.
51. Porto, A.; Araque, A.; Rabuñal, J.; Dorado, J.; Pazos, A. A New Hybrid Evolutionary Mechanism Based on Unsupervised Learning for Connectionist Systems. *Neurocomputing* **2007**, 70 (16), 2799–2808.
52. Porto-Pazos, A. B.; Veiguela, N.; Mesejo, P.; Navarrete, M.; Alvarellos, A.; Ibáñez, O.; Pazos, A.; Araque, A. Artificial Astrocytes Improve Neural Network Performance. *PLoS One* **2011**, 6 (4).
53. Alvarellos-González, A.; Pazos, A.; Porto-Pazos, A. B. Computational Models of Neuron-Astrocyte Interactions Lead to Improved Efficacy in the Performance of Neural Networks. *Comput. Math. Methods Med.* **2012.**
54. Mesejo, P.; Ibáñez, O.; Fernández-Blanco, E.; Cedrón, F.; Pazos, A.; Porto-Pazos, A. B. Artificial Neuron–Glia Networks Learning Approach Based on Cooperative Coevolution. *Int. J. Neural Syst.* **2015**.
55. Irizarry-valle, Y.; Parker, A. C.; Joshi, J. A CMOS Neuromorphic Approach to Emulate Neuro-Astrocyte Interactions. **2013**.
56. Irizarry-Valle, Y.; Parker, A. C. Astrocyte on Neuronal Phase Synchrony in CMOS. Circuits and Systems (ISCAS), 2014 IEEE International Symposium on, **2014,** 261–264.
57. Irizarry-valle, Y.; Parker, A. C. An Astrocyte Neuromorphic Circuit That Influences Neuronal Phase Synchrony. *IEEE Trans. Biomed. Circuits Syst*. **2015**, 9 (2), 175–187.
58. Nazari, S.; Amiri, M.; Faez, K.; Amiri, M. Multiplier-Less Digital Implementation of Neuron-Astrocyte Signalling on FPGA. *Neurocomputing* **2015**, 164, 281–292.
59. Nazari, S.; Faez, K.; Amiri, M.; Karami, E. A Digital Implementation of Neuron–astrocyte Interaction for Neuromorphic Applications. *Neural Networks* **2015**, 66, 79–90.
60. Nazari, S.; Faez, K.; Karami, E.; Amiri, M. A Digital Neurmorphic Circuit for a Simplified Model of Astrocyte Dynamics. *Neurosci. Lett.* **2014**, 582, 21–26
61. Oberheim, N. A.; Goldman, S. A.; Nedergaard, M. Heterogeneity of Astrocytic Form and Function. *Methods Mol. Biol.* **2012**, 814, 23–45.
62. Oberheim, N. A.; Takano, T.; Han, X.; He, W.; Lin, J. H. C.; Wang, F.; Xu, Q.; Wyatt, J. D.; Pilcher, W.; Ojemann, J. G.; Ransom, B. R.; Goldman, S. A.; Nedergaard, M. Uniquely Hominid Features of Adult Human Astrocytes. *J. Neurosci.* **2009**, 29 (10), 3276–3287.

SciForum
Mol2Net

Fragment-Based Approach for Affinity and Selectivity of *Pf*dUTPase inhibitors: Insights for Design of New Anti-Malarial Agents

Marília N. Nascimento, Marina R. Martins, Rodolpho C. Braga, Bruno J. Neves, Vinícius M. Alves and Carolina H. Andrade *

Labmol – Laboratory for Molecular Modeling and Drug Design, Faculty of Pharmacy, Federal University of Goias, Goiania, Goiás, 74605-170, Brazil.

*Author to whom correspondence should be addressed; E-Mail: carolina@ufg.br.
Tel: + 55 62 3209-6451; Fax: + 55 62 3209-6037.

Published: 4 December 2015

Abstract: Malaria is one of the leading causes of death by infectious disease worldwide. The widespread of resistance of *Plasmodium falciparum* to the current antimalarial drugs makes urgent the search and discovery of new targets and new drugs. A potential target for the development of new antimalarial drugs is deoxyuridine triphosphatase (dUTPase), and it has been validated for other organisms such as *Escherichia coli*, *Saccharomyces cerevisiae* and *Mycobacterium smegmatis*. This enzyme plays an important role in maintaining the balance between 2'-deoxyuridine 5'-triphosphate (dUTP) and 2'-deoxythymidine 5'-triphosphate (dTTP), in order to avoid the erroneous incorporation uracil in the DNA tape. In this study, we developed robust conformation-independent fragment-based quantitative structure–activity (QSAR) and structure–selectivity relationship (QSSR) models for a series of β-branched acyclic nucleotides inhibitors of *Plasmodium* and human dUTPase, aiming to design new antimalarial agents. The Hologram QSAR and QSSR models generated showed good robustness and external predictability, and is capable of predict affinity and selectivity of untested compounds inside the applicability domain. Therefore, the generated models can be used in virtual screening campaigns in the search of new potent and selective *Pf*dUTPase inhibitors.

Keywords: malaria; *Plasmodium falciparum*; drug design; dUTPase; QSAR; QSSR

1. Introduction

About 3.2 billion of people in 97 countries and territories are under the risk to contract and develop malaria. It is estimated that 198 million cases of malaria occurred in 2013, and the disease caused 584,000 deaths. 90% of cases of malaria deaths occur in Africa, and of those, 78% in children below 5 years-old[1]. Recently, notable reduction in cases of malaria have been achieved by vector control, proper case management, and combination of drugs, but the emergence of drug resistant *Plasmodium falciparum* creates a frequent need to continue the search for new antimalarial drugs [2].

The enzyme 2'-deoxyuridine 5'-triphosphate nucleotide hydrolase (dUTPase) is responsible for the hydrolytic cleavage of dUTP (deoxyuridine triphosphate) in dUMP (deoxyuridine monophosphate) and pyrophosphate [3]. Although the enzyme DNA glycosylase can excise the uracil base of DNA, many repairs can destabilize the DNA strand resulting in the breaking of the tape, that is fatal to the cell[4]. Because of that, dUTPase is a potential target for development of new drugs, and there are experimental findings that dUTPase is essential for some organisms such as *Escherichia coli*, *Saccharomyces cerevisiae* and *Mycobacterium smegmatis* [5–7]. Moreover, the *P. falciparum* enzyme (*Pf*dUTPase) has relatively low sequence similarity with human ortholog (*Hs*dUTPase) (28,4% identity) [4], making it an attractive target for the development of selective inhibitors as potential new antimalarial drugs.

The aim of this work was to develop robust conformation-independent fragment-based quantitative structure–activity (QSAR) and structure–selectivity relationship (QSSR) models for a series of β-branched acyclic nucleotides inhibitors of *Plasmodium* and human dUTPase, in order to obtain structural information on the requirements for affinity and selectivity of the enzyme *Pf*dUTPase and to design new antimalarial agents.

2. Results and Discussion

The fragment-based QSAR and QSSR models were derived for a series 127 inhibitors of *Pf*dUTPase and 47 inhibitors with biological data for both enzymes *Pf*dUTPase and *Hs*dUTPase (as measured by K_i) were determined under the same experimental conditions [4,8–13]. The K_i values were converted to the corresponding pK_i (−log K_i) and used as dependent variables in the QSAR investigations. The selectivity parameter (S) was defined as the ratio of the binding affinities of the *Plasmodium* and human enzyme (*Pf*dUTPase K_i / *Hs*dUTPase K_i), whose values were then converted to the corresponding log S. The generation of reliable QSAR models is dependent on the creation of appropriate modeling and evaluation sets. Therefore, we used a modified Kennard-Stone (KS-MD) algorithm to guide an appropriate compound selection in such a way that structurally diverse molecules, possessing activities of a wide range, were included in both sets.

Table 1 shows the results for the best HQSAR and HQSSR models developed for affinity and selectivity for *Pf*dUTPase, with different combinations of atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor and acceptor (DA) as fragment distinctions for obtaining the molecular fragments. According to Table 1, the best statistical results for HQSAR were obtained for the combination A/B fragment distinction (q^2 = 0.80; r^2 = 0.89). For HQSSR, the best model was obtained using C/H/Ch/DA as fragment distinction (q^2 = 0.79; r^2 = 0.98). Moreover, the best HQSAR and HQSSR models showed substantial predictive power with values of Q^2_{ext} equal to 0.79 and 0.83, respectively. These values indicate the reliability of the models in predicting the affinity and selectivity of untested compounds. We have also performed an analysis of the modified correlation coefficient (r_m^2). This method is considered a parameter for an additional external validation, because it is based on the actual difference between the predicted and experimental values, and is not influenced by the division of modeling and evaluation sets [14–16]. All r_m^2 metrics for our models were within the recommended range values, confirming the robustness of the models.

Besides predicting the biological property of untested molecules, HQSAR models should also provide important hints as to what molecular fragments are directly related to the biological property. This can be achieved through a careful interpretation of the structural fragments incorporated to the hologram-based QSAR models. The contribution maps show color scales that indicate the magnitude of the contribution of each atom/fragment. The colors next to red end of the spectrum (red, orange and reddish orange) indicate unfavorable or negative contributions, while colors in the green region (yellow, green-blue and green) indicate favorable or positive contributions. Atoms with intermediate contributions are colored white.

The most important fragments for the most potent and less potent inhibitors of *Pf*dUTPase and (*Hs/Pf*)dUTPase can be viewed in their individual contributions maps (Figure 1). It can be noted that the trityl ring has a favorable for the biological properties (affinity and selectivity) (colored in green). This result suggests that the trytil group could be important for the biological activity. Previous molecular modeling and crystallography studies indicated that two of the three trityl rings have significant interaction with hydrophobic site of the dUTPase enzyme[13].

Table 1. Fragment-based QSAR and QSSR models for affinity and selectivity of *Pf*dUTPase.

	q^2	r^2	Q^2_{ext}	$\bar{r}_m^2$	Δr_m^2
Model	**Affinity *Pf*dUTPase**				
A	0.72	0.85	0.85	0.81	0.09
A/B	0.80	0.89	0.79	0.76	0.11
B/C	0.79	0.90	0.77	0.75	0.15
Model	**Selectivity**				
C/H/Ch/DA	0.79	0.98	0.83	0.50	0.25
H/Ch/DA	0.75	0.96	0.88	0.55	0.22
Ch/DA	0.75	0.96	0.88	0.55	0.22

q^2: Cross-validated correlation coefficient; r^2: non cross-validated correlation coefficient; Q^2_{ext}: determination coefficient of the evaluation set; $\bar{r}_m^2$: average regression coefficient; Δr_m^2: difference between regression coefficient; Fragment distinction: A, atoms; B, bonds; C, connections; H, hydrogen atoms; Ch, chirality, DA, donor and acceptor.

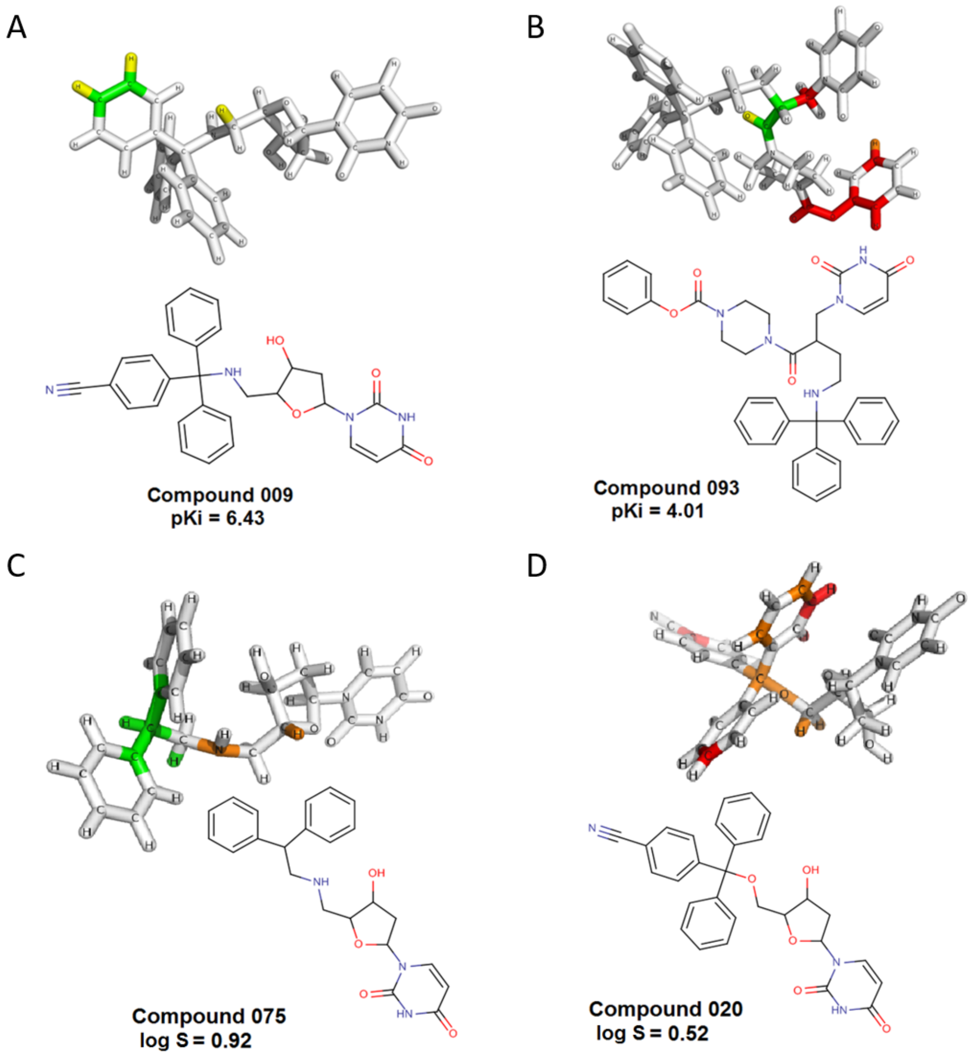


Figure 1. Structural features required for affinity, as highlighted by the fragment-based HQSAR models. Contribution maps for the most potent inhibitor of *Pf*dUTPase (A), and one of the least potent inhibitor of *Pf*dUTPase (B). Structural requirements for *Pf*dUTPase selectivity, as highlighted by the HQSSR models. Contribution maps for the most selective compound (C) and one less selective compound (D).

3. Materials and Methods

The dataset was compiled and integrated from a series of publications of Prof. Ian Gilbert (University of Dundee, UK) [4,8–13], and contained of 127 compounds along with *Ki* values against *Pf*dUTPase. From these, only 47 compounds had data reported against both enzymes from *Plasmodium* and human. This data was used to calculate the selectivity (S) (Eq. 1), which was later used to generate QSSR models. The *Ki* (inhibition constant) values were converted to the corresponding p*Ki* (-log*Ki*) and used as dependent variables in fragment-based QSAR and QSSR modeling.

Equation 1.

$$S = log\frac{HsdUTPase\ Ki}{PfdUTPase\ Ki}$$

The dataset was curated using the protocol described by Fourches and co-workers [17]. Briefly, counterions were removed, whereas specific chemotypes such as aromatic and nitro groups were normalized using the ChemAxon Standardizer (v.5.3, ChemAxon, Budapest, Hungary, http://www.chemaxon.com). The presence of duplicates, *i.e.,* identical compounds reported several times in the same dataset, is known to lead to over-optimistic estimations of the predictivity for developed QSAR models. Thus, after structural standardization, the duplicates were identified using ISIDA Duplicates [18] and HiT QSAR [19] software and carefully analyzed. If the experimental properties associated with two duplicated structures were identical, then one compound was deleted. However, if their experimental properties were significantly different, we deleted both records from the dataset. The dataset was then divided into modeling and evaluation sets, based on modified Kennard-Stone (KS-MD) algorithm. The applicability domain (AD) was assessed using the qsaR in R package.

Hologram QSAR (HQSAR) and QSSR (HQSSR) models were generated using SYBYL-X v.1.2 software (Tripos, Inc., St. Louis, MO, USA). QSAR and QSSR models were constructed using the HQSAR approach as previously described in several drug design studies [20–22]. Briefly, molecular holograms were generated for each molecule of the dataset, using different combinations of parameters concerning hologram generation. Holograms were generated using 6 distinct fragment sizes over the 12 default series of hologram lengths (53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401 bins). Several combinations of fragment distinction parameters were considered, such as atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor and acceptor (DA). The patterns of fragment counts were then related to the dependent variables using the partial least squares (PLS) regression analyses to derive the HQSAR and HQSSR models. Cross-validation procedures (q^2 leave-one-out, q^2_{LOO}) were used to determine the number of components that yielded optimally predictive models and to assess the stability and statistical significance of the models.

4. Conclusions

The HQSAR and HQSSR models presented good internal consistency, with values of q^2>0.6 and r^2>0.7. The best models were used to predict the affinity of external sets and were rigorously evaluated using several metrics, demonstrating high prediction accuracy. Therefore, the best models should be useful in predicting the affinity and selectivity of untested compounds inside the applicability domain in virtual screening campaigns, in the search of new potent and selective *Pf*dUTPase inhibitors.

Acknowledgments

We are grateful to Brazilian funding agencies, Coordination for the Improvement of Higher Education Personnel (CAPES), the National Counsel of Technological and Scientific Development (CNPq), and the State of Goiás Research Foundation (FAPEG) for their financial support and fellowships. We are also grateful to ChemAxon (Budapest, Hungary) for providing us with the academic license for their software.

Author Contributions

Conceived and designed the experiments: MNN, BJN, VMA, RCB, CHA. Performed the experiments: MNN, MRM. Analyzed the data: MNN, MRM, BJN, VMA, RCB, CHA. Contributed analysis tools: MNN, BJN, VMA, RCB, CHA. Wrote the paper: MNN, BJN, VMA, RCB, CHA.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. WHO. World malaria Report 2014. Available online: http://www.who.int/malaria/publications/world_malaria_report_2014/report/en/ (accessed Oct 19, 2015).

2. Njogu, P. M.; Guantai, E. M.; Pavadai, E.; Chibale, K. Computer-Aided Drug Discovery Approaches against the Tropical Infectious Diseases Malaria, Tuberculosis, Trypanosomiasis, and Leishmaniasis. *ACS Infectious Diseases* **2015**, 1–24.

3. Nyman, P. O. Introduction. *Current Protein and Peptide Science* **2001**, *2*, 277–285.

4. Whittingham, J. L.; Leal, I.; Nguyen, C.; Kasinathan, G.; Bell, E.; Jones, A. F.; Berry, C.; Benito, A.; Turkenburg, J. P.; Dodson, E. J.; Ruiz Perez, L. M.; Wilkinson, A. J.; Johansson, N. G.; Brun, R.; Gilbert, I. H.; Gonzalez Pacanowska, D.; Wilson, K. S. dUTPase as a platform for antimalarial drug design: structural basis for the selectivity of a class of nucleoside inhibitors. *Structure* **2005**, *13*, 329–38.

5. El-hajj, H. H.; Zhang, H. U. I.; Weiss, B. Lethality of a dut (Deoxyuridine Triphosphatase) Mutation in Escherichia coli. *Journal of Bacteriology* **1988**, *170*, 1069–1075.

6. Gadsden, M. H.; Mclntosh, E. M.; Game, J. C.; Wilson, P. J.; Haynes, R. H. dUTP pyrophosphatase is an essential enzyme in Saccharomyces cerevisiae. *The EMBO Journal* **1993**, *12*, 4425–4431.

7. Pecsi, I.; Hirmondo, R.; Brown, A. C.; Lopata, A.; Parish, T.; Vertessy, B. G.; Tóth, J. The dUTPase enzyme is essential in Mycobacterium smegmatis. *PloS one* **2012**, *7*, e37461.

8. Nguyen, C.; Kasinathan, G.; Leal-Cortijo, I.; Musso-Buendia, A.; Kaiser, M.; Brun, R.; Ruiz-Pérez, L. M.; Johansson, N. G.; González-Pacanowska, D.; Gilbert, I. H. Deoxyuridine triphosphate nucleotidohydrolase as a potential antiparasitic drug target. *Journal of Medicinal Chemistry* **2005**, *48*, 5942–54.

9. Nguyen, C.; Ruda, G. F.; Schipani, A.; Kasinathan, G.; Leal, I.; Musso-Buendia, A.; Kaiser, M.; Brun, R.; Ruiz-Pérez, L. M.; Sahlberg, B.-L.; Johansson, N. G.; Gonzalez-Pacanowska, D.; Gilbert, I. H. Acyclic nucleoside analogues as inhibitors of Plasmodium falciparum dUTPase. *Journal of Medicinal Chemistry* **2006**, *49*, 4183–95.

10. McCarthy, O.; Musso-Buendia, A.; Kaiser, M.; Brun, R.; Ruiz-Perez, L. M.; Johansson, N. G.; Pacanowska, D. G.; Gilbert, I. H. Design, synthesis and evaluation of novel uracil acetamide derivatives as potential inhibitors of Plasmodium falciparum dUTP nucleotidohydrolase. *European Journal of Medicinal Chemistry* **2009**, *44*, 678–88.

11. Baragaña, B.; McCarthy, O.; Sánchez, P.; Bosch-Navarrete, C.; Kaiser, M.; Brun, R.; Whittingham,

J. L.; Roberts, S. M.; Zhou, X.-X.; Wilson, K. S.; Johansson, N. G.; González-Pacanowska, D.; Gilbert, I. H. β-Branched acyclic nucleoside analogues as inhibitors of Plasmodium falciparum dUTPase. *Bioorganic & Medicinal Chemistry* **2011**, *19*, 2378–91.

12. Ruda, G. F.; Nguyen, C.; Ziemkowski, P.; Felczak, K.; Kasinathan, G.; Musso-Buendia, A.; Sund, C.; Zhou, X. X.; Kaiser, M.; Ruiz-Pérez, L. M.; Brun, R.; Kulikowski, T.; Johansson, N. G.; González-Pacanowska, D.; Gilbert, I. H. Modified 5'-trityl nucleosides as inhibitors of Plasmodium falciparum dUTPase. *ChemMedChem* **2011**, *6*, 309–20.

13. Hampton, S. E.; Baragaña, B.; Schipani, A.; Bosch-Navarrete, C.; Musso-Buendía, J. A.; Recio, E.; Kaiser, M.; Whittingham, J. L.; Roberts, S. M.; Shevtsov, M.; Brannigan, J. a.; Kahnberg, P.; Brun, R.; Wilson, K. S.; González-Pacanowska, D.; Johansson, N. G.; Gilbert, I. H. Design, synthesis, and evaluation of 5'-diphenyl nucleoside analogues as inhibitors of the Plasmodium falciparum dUTPase. *ChemMedChem* **2011**, *6*, 1816–31.

14. Mitra, I.; Roy, P. P.; Kar, S.; Ojha, P. K.; Roy, K. On further aplication of rm2 as a metric for validation of QSAR Models. *Journal of Chemometrics* **2010**, *24*, 22–33.

15. Roy, K.; Chakraborty, P.; Mitra, I.; Ojha, P. K.; Kar, S.; Das, R. N. Some case studies on application of "r(m)2" metrics for judging quality of quantitative structure-activity relationship predictions: emphasis on scaling of response data. *Journal of Computational Chemistry* **2013**, *34*, 1071–82.

16. Roy, K.; Mitra, I. On the use of the metric rm^2 as an effective tool for validation of QSAR models in computational drug design and predictive toxicology. *Mini Reviews in Medicinal Chemistry* **2012**, *12*, 491–504.

17. Fourches, D.; Muratov, E.; Tropsha, A. Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. *Journal of Chemical Information and Modeling* **2010**, *50*, 1189–204.

18. Varnek, A.; Fourches, D.; Horvath, D.; Klimchuk, O.; Gaudin, C.; Vayer, P.; Solov'ev, V.; Hoonakker, F.; Tetko, I.; Marcou, G. ISIDA - Platform for Virtual Screening Based on Fragment and Pharmacophoric Descriptors. *Current Computer Aided-Drug Design* **2008**, *4*, 191–198.

19. Kuz'min, V. E.; Artemenko, a G.; Muratov, E. N. Hierarchical QSAR technology based on the Simplex representation of molecular structure. *Journal of Computer-Aided Molecular Design* **2014**, *22*, 403–21.

20. Andrade, C. H.; Salum, L. D. B.; Castilho, M. S.; Pasqualoto, K. F. M.; Ferreira, E. I.; Andricopulo, A. D. Fragment-based and classical quantitative structure-activity relationships for a series of hydrazides as antituberculosis agents. *Molecular diversity* **2008**, *12*, 47–59.

21. Salum, L. B.; Andricopulo, A. D.; Honório, K. M. A fragment-based approach for ligand binding affinity and selectivity for the liver X receptor beta. *Journal of molecular graphics & modelling* **2012**, *32*, 19–31.

22. Salum, L. B.; Dias, L. C.; Andricopulo, A. D. Fragment-Based QSAR and molecular modeling studies on a series of discodermolide analogs as microtubule-stabilizing anticancer agents. *QSAR & combinatorial science* **2009**, *28*, 325–337.

SciForum
Mol2Net

Iterative Kernel K-Means for Metagenomic Sequences

Isis Bonet [1,*], Andrea Mesa-Múnera [1], Adriana Escobar [1] and Juan Fernando Alzate [2]

[1] Escuela de Ingeniería de Antioquia, Envigado, Antioquia, Colombia; E-Mails: amesamu@gmail.com; adriana.escobarv@gmail.com

[2] Centro Nacional de Secuenciación Genómica, Facultad de Medicina, Universidad de Antioquia, Colombia; E-Mail: jfernando.alzate@udea.edu.co;

* E-Mail: ibonetc@gmail.com; Tel.: +57-4-3549090 (ext. 330)

Published: 4 December 2015

Abstract: This paper shows an iterative clustering method based on kernel *k*-means, which changes the parameter *k* automatically in each iteration of the algorithm. In addition, a way to initialize the centroids is proposed. The method is applied to a binning process in metagenomics using a complex database with different organisms. The aim of this method is to reduce the sensitivity of clusters based on strength measures. The results demonstrate that the proposed method is better than the simple kernel *k*-means for metagenome databases.

Keywords: Metagenomics; k-means; clustering; bioinformatics

1. Introduction

Metagenomics is the science that studies microbial DNA of many organisms recovered from environmental samples. Ever since the studies of DNA in a single organism the use of computational resources was an important need. Now this science has stirred the rise of new computational challenges. Next-generation sequencing technologies can sequence up to billions of bases in a single day at low cost, producing a huge amount of short fragments of DNA called reads. The next process and new challenge is to assemble these reads into longer sequences called contigs and scaffolds by a process of overlapping [1]. Assembling this huge amount of short reads was difficult in the classic genomic study for a single microorganism. Now assembling imposes great computational challenge because in metagenomics the data we are dealing with contains different microorganisms at uneven abundances.

Binning process for assignment of genomic fragments into taxonomic groups is one of the most important steps in the analysis of metagenomic data, but in spite of several developed tools it is still a challenge for scientists. Similarity-based and supervised methods are more accurate than unsupervised methods because they are based on reference sequences, but for the same reason they are more time consuming and have limitations when they are dealing with unknown organisms or these are not present in their databases. The huge amount of reads or contigs to align with known sequences coupled with the big size of the known sequences databases are the cause of the high time consumption. Therefore if we reduce the amount of reads or contigs to align with, the time to find the sequence that match should considerably decrease. A previous clustering process can be an efficient way to provide different taxonomic groups in order to ease the analysis of a few fragments of sequences that probably belong to the same organism. This process can be used as a previous step in some processes in the study of metagenomic samples such as before the assembly or in the process of functional assignment. Some researchers have used variants of *k*-means [2,3], variants of Self-Organizing Maps [4], [5] and others clustering techniques [1,6]. In [7] a comparison of some different clustering methods is done.

The selection of an appropriate clustering method to represent the taxonomic groups is yet a challenge. The complexity and the high dimensional of the data are two of the problems to keep in mind in clustering metagenomics.

K-means is one of the most popular clustering algorithms, but it has some limitations. One of the most important disadvantages is the number of clusters needs to be specified by the user. A key limitation of *k*-means is the way to build the clusters, typically spherical clusters with similar size, which are linearly separable.

Taking into account the potential of k-means without forgetting its limitations this paper focuses on a kernel *k*-means method with a variant consisting of iterations in order to select an appropriate number *k* of clusters. Also a random way to select the centroids based on the distance between them is used improving the convergence of the method.

2. Data and Methods

2.1 Data

The database used in this paper was previously used by Bonet et al. [3]. It consists of assembled genomic sequences of different organisms: viruses, bacteria and eukaryotes from the FTP site of the Sanger institute.

Selected viral sequences include Influenza and Dengue virus genomes. Bacterial sequences come from Bacteroides dorei and Bifidobacterium longum. The selected eukaryotes included two fungi, one nematode and one insect.

The database contains 165014 contigs that ranged between 50 and 2962289 bases because the enormous difference in the size of contigs is needed to represent the sequences using biological or mathematical features.

2.2 Features

For the experiment some features were selected:

- GC: G + C content, that means the ratio between the number of G+C and the total of nucleotides of the sequence (A+T+G+C).
- Nucleotide frequencies: Number of occurrences of A, T, G and C in the sequence. It was normalized by the size of the sequence.
- Codon frequencies: Number of each possible codon in the sequence. It was normalized by the total of codons (64 codons)
- *k*-mer (*k*=4): are represented for the 256 possible tetranucleotides. It was compute as

the number of each tetranucleotide and normalized with the total of tetranucleotides in the sequence.

Features were used in all combinations, producing 15 databases.

2.3 Methods

K-means is one of the most popular clustering methods, despite the problem to estimate the parameter *k* (number of cluster). This algorithm finds a set of *k* centroids, and associates each instance in the data to the nearest centroid, based on a distance function [8].

Some researchers have focused on the initialization part of the method, based on the selection of better centroids in order to improve the convergence of the algorithm. One of the most known is *k*-means++ [9] and variants of it, including scalable *k*-means++ [10]. Most of these algorithms need to analyze the entire database, which requires a lot of time in large databases. Here we propose a simple and fast way to select a set of optimal centroids.

Appling *k*-means to massive data is easy because of its nature. Given a set of centroids, the assignment of each point to clusters can be done independently.

Kernel *k*-means works as k-means but applied in kernel space [11].

Here we proposed a clustering method based on kernel *k*-means.

Polynomial and cosine distance kernel were used to compare the sequences.

For the implementation of the clustering method, we used Weka [12], which is a free machine learning package that has implemented *k*-means.

3. Iterative kernel *k*-means

3.1 Selecting centroids

The process to select the centroids consists on:

1. Select *k* random points (*k* cases of the database).
2. Select a *k*+1 point. Compute the distance matrix of theses *k*+1 point. For each point, compute the average distance. Delete the point with lowest average.
3. Repeat step 2 until obtain an average greater than a threshold or a number of iterations.

Using this simple idea, we obtain a set of centroids more distant from each other, what is one of the objectives of the final clusters.

In this paper, we use *k* as the number of iterations for the step 2.

3.2 Iterative kernel *k*-means

The proposed process of clustering is based on the algorithm suggested by Bonet et al. [3] with the addition of a distance kernel. The distance kernel is based on a cosine transformation with a lineal kernel as is shown in equation 1.

$$CosineKernel(x_1, x_2) = \frac{k(x_1,x_2)}{\sqrt{k(x_1,x_1)*k(x_2,x_2)}} \quad (1)$$

where k is a kernel. In our case the linear kernel was use, i.e. $k(x_1, x_2)$ is a dot product of x_1 and x_2.

The process is following these steps:

Step 1: Select a tentative *k* (this *k* varies in the rest of the process), preferably a higher value than expected. Run *k*-means with the data using the initialization process and the cosine distance kernel described above.

Step 2: After getting the first set of clusters, they are evaluated based on measures of strengths of clusters. Clusters with low compactness, that is low distance inter-cluster, are used to build the new database to repeat the clustering process returning to step 2.

Step 3: Once the strengths measures are lower than a threshold, the last step is to minimize, if possible, the number of clusters. Clusters evaluation is repeated, for all clusters resulted of each iteration of *k*-means. Clusters with low separation between theirs centroids are merged into one.

In metagenomics the aim to assign the sequences to a phylum is associated with the sensitivity taking into account the phylum that best represents each cluster. That means the sensitivity is measured centered around the percentage that represents each organism in each cluster. For this problem, we use the sensitivity of clusters to evaluate them.

4. Results and Discussion

A metagenome database composed of eight different organisms is used to evaluate the method.

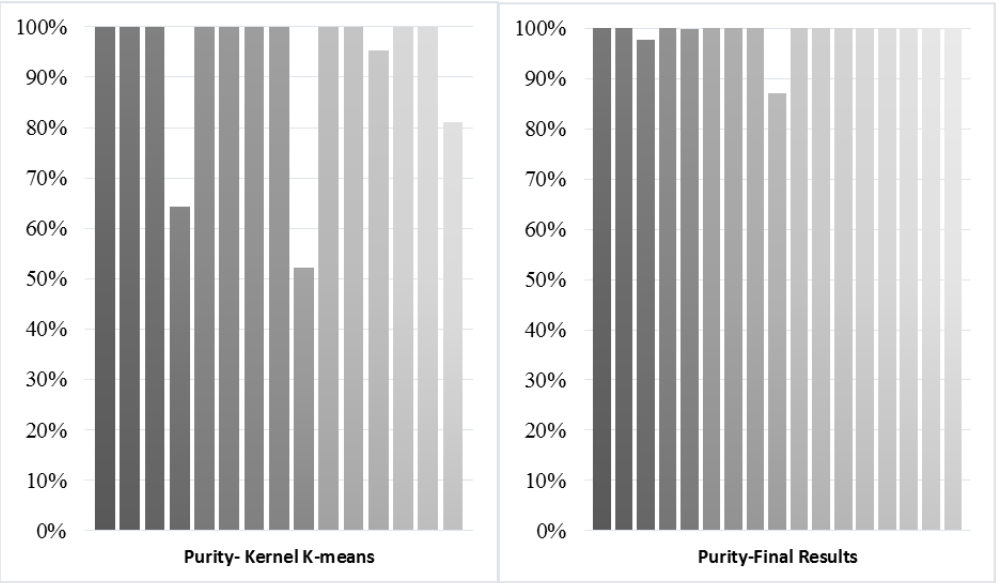


Figure 1. *K*-means vs. Iterative kernel *k*-means

Some different attributes are used to describe the sequences: GC content, nucleotides frequencies, codon frequencies and tetranucleotides. All combinations of features were tested, but the best performance was obtained using tetranucleotides.

Polynomial and cosine kernel were used for the kernel k-means algorithm. The best result was obtained with cosine kernel. The algorithm was tested with k between 5 and 15 achieving the best performance with k=15.

Figure 1 shows the results with kernel cosine and k=15. The clusters obtained with kernel k-means (left) vs. the clusters obtained using the proposed algorithm with five iterations (right). The figure represents the percent of purity of the clusters that means, the percent of the genomic fragments that belongs to the predominant organism in the cluster.

The results of the last step of the model yielding a 99.1% of sensitivity of the clusters, which results are in the range of 87.14 and 100%. The error of misassigned sequences is 5.516%.

.

4. Conclusions

In this paper we present an algorithm based on kernel *k*-means. The algorithm was tested in a metagenome database. The result achieved by the proposed method in line with the objective of obtaining clusters with high sensitivity outperforms results obtained with a simple *k*-means. Taking into account the sensitivity of the clusters the model yielding a 99.1%.

Conflicts of Interest

The authors declare no conflict of interest

References

1. Reddy, R.M.; Mohammed, M.H.; Mande, S.S. Metacaa: A clustering-aided methodology for efficient assembly of metagenomic datasets. *Genomics* **2014**, *103*, 161-168.
2. Kelley, D.; Salzberg, S. Clustering metagenomic sequences with interpolated markov models. *BMC Bioinformatics* **2010**, *11*, 544.
3. Bonet, I.; Montoya, W.; Mesa-Múnera, A.; Alzate, J. Iterative clustering method for metagenomic sequences. In *Mining intelligence and knowledge exploration*, Prasath, R.; O'Reilly, P.; Kathirvalavakumar, T., Eds. Springer International Publishing: 2014; Vol. 8891, pp 145-154.
4. Weber, M.; Teeling, H.; Huang, S.; Waldmann, J.; Kassabgy, M.; Fuchs, B.M.; Klindworth, A.; Klockow, C.; Wichels, A.; Gerdts, G*., et al.* Practical application of self-organizing maps to interrelate biodiversity and functional data in ngs-based metagenomics. *The ISME journal* **2011**, *5*, 918-928.
5. Abe, T.; Kanaya, S.; Kinouchi, M.; Ichiba, Y.; Kozuki, T.; Ikemura, T. Informatics for unveiling hidden genome signatures. *Genome Research* **2003**, *13*, 693-702.
6. Kislyuk, A.; Bhatnagar, S.; Dushoff, J.; Weitz, J. Unsupervised statistical clustering of environmental shotgun sequences. *BMC Bioinformatics* **2009**, *10*, 316.
7. Li, W.; Fu, L.; Niu, B.; Wu, S.; Wooley, J. Ultrafast clustering algorithms for metagenomic sequence analysis. *Briefings in Bioinformatics* **2012**, *13*, 656-668.
8. MacQueen, J. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, University of California Press: Berkeley, Calif., 1967; pp 281-297.
9. Arthur, D.; Vassilvitskii, S. In *K-means ++: The advantages of careful seeding*, 8th Annual ACM-SIAM Symposium on Discrete Algorithms, New Orleans, 7-9 January 2007, 2007; New Orleans, pp 1027-1035.
10. Bahmani, B.; Moseley, B.; Vattani, A.; Kumar, R.; Vassilvitskii, S. Scalable k-means++. *Proc. VLDB Endow.* **2012**, *5*, 622-633.
11. Scholkopf, B.; Smola, A.; Muller, K.-R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation* **1998**, *10*, 1299-1319.

12. Witten, I.; Frank, E. *Data mining: Practical machine learning tools and techniques*. 2nd ed.; Morgan Kaufmann: San Francisco, 2005; p 525.

SciForum
Mol2Net

Combination of Microscopic and Spectroscopic Techniques to Study the Presence and the Effects of Microplastics in Mussels

Mireia Irazola[1]*, Larraitz Garmendia[2], Beñat Zaldibar[2], Urtzi Izagirre[2], Eider Bilbao[2], Sara Danielsson[3], Anders Bignert[3], Kepa Castro[1], Nestor Etxebarria[1], Manu Soto[2], and Ionan Marigomez[2]

[1] IBeA Research Group, Research Centre for Experimental Marine Biology and Biotechnology (PiE-UPV/EHU) and Department of Analytical Chemistry, University of the Basque Country, P.O. Box 644, E-48080 Bilbo, Basque Country, Spain.

[2] CBET Research Group, Research Centre for Experimental Marine Biology and Biotechnology (PiE-UPV/EHU), University of the Basque Country, P.O. Box 644, E-48080 Bilbo, Basque Country, Spain;

[3] Swedish Museum of Natural History, University of Stockholm, P.O. Box 50007, SE-104 05 Stockholm, Sweden C

* Correspondence to: Mireia Irazola, Department of Analytical Chemistry, EHU/UPV, P.O. Box 644 E-48080, Bilbao, Basque Country, Spain. E-mail: mireia.irazola@gmail.com

Published: 4 December 2015

Abstract: The growing concern due to the presence of plastics, especially micro and nanoplastics, in environmental aquatic media requires the development of new methodologies to study the distribution of these particles and the effects that might cause in many organisms. In this work we have performed experiments using synthetic polystyrene microplastics (6-90 µm diameter) and mussels (*Mytilus galloprovincialis*) and we have studied the distribution of these particles by different techniques including FTIR and Raman spectroscopy, light and polarized light microscopy after being exposed for different periods of time (1-72 h). As a result of this work we were able to fine tune the preparation of the samples, from conservation to image and spectra analysis, and it was concluded that it was better to freeze the samples and to prepare the cryosections instead of embedding in paraffin. Regarding the light microscopy darkfield illumination offered less background signals than polarized one and therefore it was more suitable for small size particles. Finally, Raman spectroscopy allowed the characterization of the polystyrene particles better than FTIR allowing the development of image analysis techniques.

Keywords: microplastics; mussels; uptake, accumulation and distribution; method development; microscopy; Raman spectroscopy, FTIR spectroscopy

1. Introduction

There are few features comparable to the evolution and deep impact of polymeric materials and plastics in modern way of life. According to the information provided by the manufacturers, the global production is above 300 million of tons (1). Part of this production is discarded in uncontrolled plastic debris that are physically fragmented in smaller pieces and end up in river and oceans. In addition to the coarse plastic materials, the production and use of microplastics (MPs) in cosmetics and personal care products shows a specific interest due to the higher distribution rate in many environmental compartments. In any case, the impact of the presence of these MPs in aquatic organisms and in food is still being studied (2-4).

The analysis of the microplastics in the physical aquatic media has been described in the most recent literature (2, 5) but the analysis in organisms still requires a deeper methodological development. In any case, the methodological and instrumental approaches depend among others on the size ranges of the particles that are analysed and the sort of effects that are looking for (6).

In this context, the main aims of this work were to develop the methodology to study the accumulation of polystyrene microplastics in controlled exposure experiments to support further studies and to get some insights about the accumulation and distribution of these microplastics in the different tissues.

.

2. Results and Discussion

Differences in MPs distribution were observed in mussels according to the size of the MPs. Big MPs were mainly detected in the lumen of the stomach and also in the digestive conducts, but not in the digestive tubule epithelium. Additionally, smaller MPs (6 and 10 µm diameter) were observed in the connective tissue surrounding the stomach and digestive gland and also in a lesser extent in the lumen of the digestive tubules and inside the digestive epithelium.

Significant differences were observed according to the technique for the MP visualization. Both polarized light and darkfield illumination were able to detect MPs more relevantly than brightfiels illumination. However, darkfield illumination presented more signal to noise ration (Figure 1). Finally, it is noteworthy that in paraffin embedded samples no microplastic were detected but mechanical damage related to big microplastics was observed (Figure 1). In fact, during the sample preparation of paraffin embedded samples the MPs were dissolved.

The Raman results acquired at the same time that we made the histological study helped us to assure that we looked at the microscope was undoubtedly polystyrene MPs. In the Raman spectra of the figure 1 we can observe the main band of the polystyrene at 1000 cm^{-1}. We found the Raman measurements necessary since we got some noise problems or even false positives with the microscopic techniques. It is true that in this study there are few confounding factors since the shape of the MPs that we used for this experiments were identical. But, in the real world, with real samples, the MPs present a huge variety of shapes and colours. Therefore we found useful to combine microscopy results with Raman results.

The FTIR images provided us biochemical information as well as the distribution of the MPs. Anyway, to get the chemical information we were looking for it was necessary to use chemometric tools such as MCR and PCA.

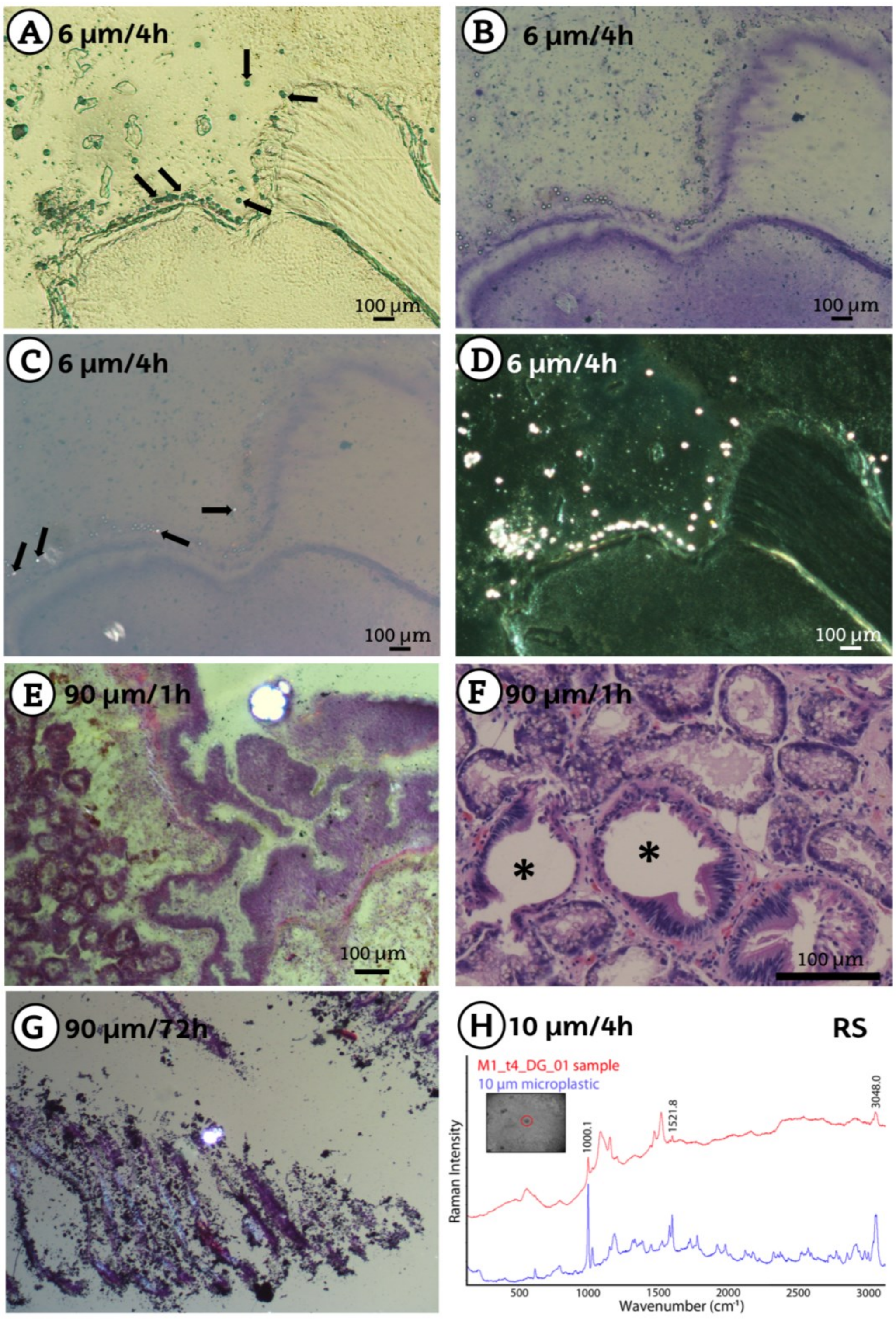


Figure 1. Microplastics (MPs) in the digestive gland of mussels with different illumination techniques (A-F). A-D; 6 µm diameter MPs after 4 h of exposure after brightfield (A and B), polarized (C) and darkfield illumination (D). E-F; 90 µm diameter MPs after 1 h of

exposure after polarized (E) illumination and brightfield illumination in paraffin embedded tissue (F). And asterisks indicate the mechanical damage induced by 90µm diameter polystyrene MPs in the secondary ducts. In the G picture can be observed 90 µm MPs in the gills after 72 h of exposition. Raman Spectroscopy (RS) was employed to check if we really are observing MPs. Arrows indicate the MPs.

3. Materials and Methods

I. Exposure experiments.

Mussels (*Mytilus galloprovincialis*) were collected in the estuary of the Butroe river (Bay of Biscay, Basque Country) and immediately transferred to the laboratory.

Mussels were exposed to three polystyrene (Alfa Aesar) microplastics of different sizes (6, 10 and 90 µm diameter). Mussels were collected after, 1, 4, 8 and 72 h of exposure. Then, they were dissected and some were formalin fixed and routinely processed for histological observation after paraffin embedding. Other mussels, were snap frozen in liquid nitrogen and stored at -80ºC for further cryostat sectioning.

II. Microscopy study.

Both paraffin embedded, and cryostat frozen samples were observed under the microscope in different illumination conditions in order to detect the MPs. On the one hand, mussel tissue sections were observed with brightfield and darkfield illumination in a Nikon ECLIPSE TI-S (Nikon, Tokyo, Japan) microscope. On the other hand, samples were also observed under polarized light in a Olympus BH2 microscope (Olympus Corporation, Tokyo, Japan).

III. Spectroscopy study.

Raman spectra were acquired at the same time that we were performing the microscopy study to verify the presence of the MPs and to avoid false positives. InnoRaman^{TM} portable spectrometer (B&WTEK$_{INC}$, Newark, USA) coupled to a microscope (20x and 50x magnification) and provided with 532 nm laser and CCD detector (Peltier cooled) was used for this issue.

The region of interest, previously selected in the microscopy study, was imaged on a Jasco IMV4000 FTIR imaging system equipped with a liquid nitrogen cooled 16 element linear array MCT detector. The cryosections were placed on a ZnSe sample holder and imaged directly. For each sample image, background was recorded using the same experimental parameters and on an empty region of the ZnSe sample holder. Once the FTIR images were acquired, principal component analysis (PCA) and multivariate curve resolution (MCR) analysis (MATLAB) were performed to get the spatial distribution of MPs .

4. Conclusions

Tissue distribution of MPs varies with their size. 90 µm diameter MPs were mainly limited to the stomach and ducts while 6 and 10 µm diameter MPs were found mainly in the connective and sometimes in the digestive epithelium as well.

One of the keys to obtain good results has been the sample preparation. Frozen tissue allows a better detection of MPs than paraffin embedded tissue in mussels.

Regarding to the microscopy study, polarized light and darkfield microscopy are useful tools to detect microsplastics in mussel's tissue but they are not 100% reliable due to the fact that we found some false positives. We recommend verify the results with molecular spectroscopy, such as Raman and FTIR.

Acknowledgments

This work was supported by the Swedish Museum of Natural History, University of Stockholm, Sweden and by the University of the Basque Country/Basque Government through the grants to CBET and IBeA Consolidated Research Groups. Technical support in the FTIR imaging analysis provided by SGIker (UPV/EHU) is gratefully acknowledged. Mireia Irazola acknowledges to the UPV/EHU for the PhD grant.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. PlasticsEurope, 2015. Plastics. The Facts 2015 (http://www.plasticseurope.org/cust/documentrequest.aspx?DocID=65435) (accessed on 29/11/2015)
2. Claessens M., Van Cauwenberghe L., Vandegehuchte M.B., Jannsen C.R. *New techniques for the detection of microplastics in sediments and field collected organisms*, Marine Pollution Bulletin, 2013, 70, 227-233
3. Wright S.L., Thompson R.C., Galloway T.S. *The physical impacts of microplastics on marine organisms: A review*. Environmental Pollution 2013, 178, 483-492
4. Ivar do Sul, J. Costa M.F. *The present and future of microplastic pollution in the marine environment*, Environmental Pollution 2014, 185, 352-364
5. Masura J., Baker J., Foster G., Arthur C. Analysis of Microplastics in the Marine Environment. Recommendations for quantifying synthetic particles in waters and sediments. 2015. NOAA Technical Memorandum NOS-OR&R-48
6. Von Moos N., Burkhardt-Holm P., Köhler A. *Uptake and Effects of Microplastics on Cells and Tissue of the Blue Mussel Mytilus edulis L. after an Experimental Exposure*, Environmental Science & Technology 2012, 46, 11327-11335.

SciForum
Mol2Net

2-Nitromethylacrylates as Useful Dinucleophiles for the Enantioselective Organocatalytic Michael/Henry Cascade Reaction

Naiara Fernández [1], Iker Riaño [1] , Uxue Uria [1] , Efraim Reyes [1] , Luisa Carrillo [1] and Jose L. Vicario [1,*]

[1] Departamento de Química Orgánica II, Facultad de Ciencia y Tecnología, Universidad del País Vasco/Euskal Herriko Unibertsitatea UPV/EHU, P.O. Box 644, E-48080 Bilbao, Spain; E-Mail: naiara.fh@gmail.com (N.F.), iker.riano@ehu.eus (I.R.); uxue.uria@ehu.eus (U.U.); efraim.reyes@ehu.eus (E.R.); marisa.carrillo@ehu.eus (L.C.)

* Author to whom correspondence should be addressed; E-Mail: joseluis.vicario@ehu.eus; Tel.: (34)-94-601-5454; Fax: (+34)-94-601-2748

Published: 4 December 2015

Abstract:
2-Nitromethylacrylates have proved to be suitable 1,3-dinucleophiles reacting with α,β-unsaturated aldehydes in the presence of a secondary-amine catalyst to furnish Michael/Henry cascade products in moderate yields and with high enantioselectivities although with moderate diastereoselectivities. The reaction proceeds by iminium ion activation of the enal, which reacts regioselectively with the γ-carbon of the nitronate anion formed *in situ*, furnishing the desired cyclohexenes with three new stereocenters. Furthermore, and trying to avoid the diastereoselectivity issue, an efficient sequential Michael/Henry/dehydration reaction has been developed leading to enantiopure cyclohexadienes in moderate yields and excellent enantioselectivities.

Keywords: Michael/Henry reaction; Enantioselective cascade reaction; Iminium catalysis

1. Introduction

The Henry reaction has been associated to the Michael addition in domino type transformations, since it is an useful tool to form C-C bonds and also allows the preparation of a wide range of structurally different compounds due to the ability of the nitro group to be transformed into other nitrogen and oxygen containing functionalities.[1] The application of this cascade reaction to the synthesis of optically active complex products is unquestionable and the examples described in the literature show its remarkable applicability making it a highly efficient tool for obtaining functionalized products with total diastereo- and enantiocontrol employing different types of catalysts. In this sense, it should be highlighted the area of organocatalysis, where enamine and iminium activation together with H-bonding catalysis strategy enable to carry out Michael/Henry cascade reactions with high efficiency and an exceptional level of control of the stereocenters formed in most cases.[2]

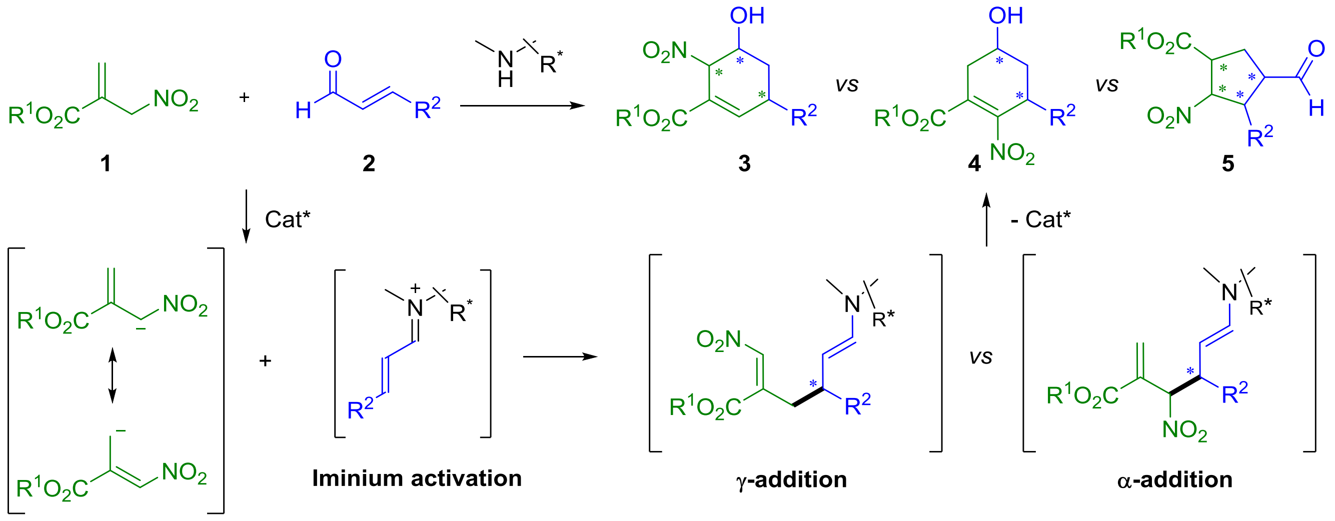
In this context, and taking into account the experience acquired in our research group in the field of organocatalytic Michael/Henry reactions,[3] *we decided to focus our studies on developing a* ***Michael/Henry cascade process employing 2-nitromethylacrylates*** *as suitable functionalized Michael donors with α,β-*

Nevertheless, we have to be aware of the fact that if the catalyst is not released from the resulting Michael addition product, another intramolecular conjugate addition can take place

unsaturated aldehydes under iminium activation (Scheme 1).

2-Nitromethylacrylates are expected to be active as carbon pronucleophiles in an initial Michael reaction because of the presence of two acidic protons in α position to the nitro group.[4] Moreover, the deprotonation of this compound **1** would lead to the generation of a resonance-stabilized allylic anion, with potential to react either at C-α or at C-γ position (Scheme 1). Assuming that the selected pronucleophile can undergo a conjugate addition to the iminium ion resulting from the condensation of the Michael acceptor –an α,β-unsaturated aldehyde **2**– and the aminocatalyst, the release of the catalyst after the initial Michael reaction would form a nitroaldehyde intermediate which is expected to undergo intramolecular Henry reaction in order to yield the desired product (compound **3** *vs* **4**).

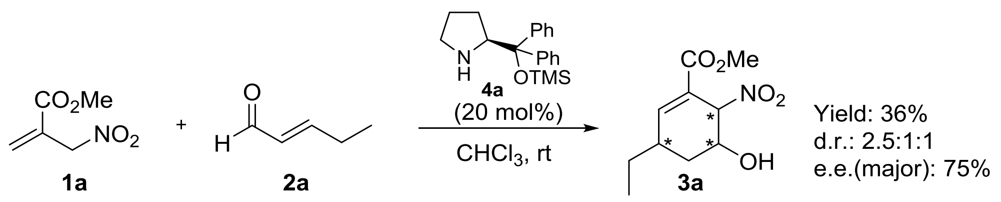
between the enamine and the remaining α,β-unsaturated ester moiety, leading to the formation of cyclopentanes **5** as shown in Scheme 1.



Scheme 1. Catalytic enantioselective Michael/Henry cascade reaction employing 2-nitromethylacrylates

2. Results and Discussion

With this goal in mind, we surveyed the behavior of compound **1a** with *trans*-pentenal **2a** as model system in the presence of a chiral secondary amine catalyst. In a first attempt, we carried out the reaction using diphenylprolinoltrimethyl silyl ether **4a** as catalyst, in chloroform and at room temperature. When the compound **1a** was completely consumed, the ^{1}H-NMR of the crude reaction mixture revealed that the cyclohexene **3a** had been formed without observing other possible byproducts previously predicted (**4** or **5**). We concluded that the 2-nitromethylacrylate **1a** was behaving as a 1,3-dinucleophile, leading to the polysubstituted cyclohexene after performing the projected cascade reaction through initial selective γ-addition to the intermediate iminium ion derived from pentenal, followed by catalyst releasing and intramolecular Henry reaction. The compound **3a** was isolated by flash column chromatography in 36% yield as a mixture of three diastereoisomers (2.5:1:1), and with a 75% e.e. for the major diastereoisomer.

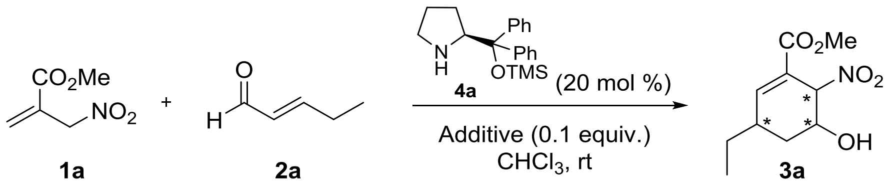


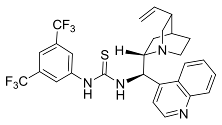
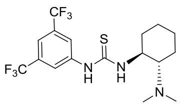
Yield: 36%
d.r.: 2.5:1:1
e.e.(major): 75%

Scheme 2. Viability of the Michael/Henry reaction

These results made us explore the possibilities that this reaction could offer. The first parameter to study was the organocatalyst employed and therefore, we tested some different chiral secondary amines in the model reaction. Taking into account the high enantioselectivity achieved with prolinol derivative **4a**, we decided to use other diphenylprolinol derivatives with a bulkier *O*-protecting group and also diarylprolinol derivatives containing bulkier aryl substituents, but the reaction did not progress in any case and the starting materials were recovered in all cases. As the other catalysts tested did not show any activity, we decided to study the influence of the additive (Table 1). Some examples in the literature,[1a] made us think that Brønsted bases could be beneficial for the reaction, assuming that a base could help in the nucleophile formation by promoting the deprotonation of the substrate **1a**, and later on it could also assist the intramolecular Henry reaction. Fortunately, when bases such as DBU, 1,1,3,3-tetramethylguanidine, triethylamine and DMAP were used cyclohexene **3a** was selectively obtained (entries 2-5). The highest value regarding the yield was achieved with triethylamine but it showed poor diastereoselectivity (entry 4) and the best enantiocontrol was obtained employing DMAP although the yield was not good enough (entry 5).

Table 1. Influence of the additive in the cascade reaction.[a]



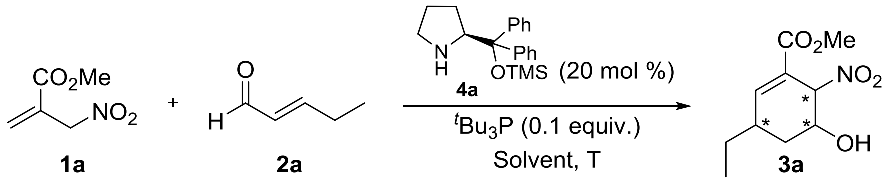
Entry	Additive	d.r.[b]	Yield (%)[c]	e.e. (%)[d]
1	-	2.5:1:1	36	36
2	DBU	2.5:1:1	34	86
3	1,1,3,3-Tetramethylguanidine	2.5:1:1	53	88
4	Et$_3$N	2:1.4:1	76	86
5	DMAP	2.4:1.2:1	43	90
6		5:2:1	40	88
7		3.3:1.7:1	29	79
8	Ph$_3$P	3.3:1:1.6	52	86
9	Bu$_3$P	2.5:1.7:1	33	90
10	^{t}Bu$_3$P	2.5:1:1.5	74	88

[a] One equivalent of **1a** and two equivalents of aldehyde **2a** were used. [b] Determined by ^{1}H-NMR analysis for the mixture of diastereoisomers after flash column chromatography purification. [c] Referred to the mixture of diastereoisomers after flash column chromatography purification. [d] Calculated by HPLC for the major diastereoisomer.

Bifunctional Brønsted bases containing acidic H-donor sites were also tried as additives (entries 6-7) in order to evaluate their potential positive contribution to the reaction, but without giving better results than the ones obtained with triethylamine (entry 6). Finally, we decided to evaluate the use of phosphines as cocatalyst (entries 8- 10) and in this sense, the use of a bulky phosphine like ^{t}Bu$_3$P gave the compound **3a** with good yield and an acceptable value of diastereomeric excess and the best enantioselectivity (entry 10). Moreover, it has to be pointed out that, in all cases the formation of other regioisomers was not detected by NMR analysis of the crude reaction mixture.

Selecting ^{t}Bu$_3$P as the best additive, the optimization of the model reaction went on with the election of the most suitable solvent for the Michael/Henry transformation (Table 2). A battery of tests were run to evaluate the influence of different solvents, using polar as well as non-polar ones, and in a general view, the diastereo- and enantioselection of the cascade reaction was not influenced by the nature of the solvent (entries 1-7). The tests run revealed that chloroform (entry 1) remained being the most suitable solvent for the reaction in terms of overall yield and enantioselectivity. Finally, we tried to improve the diastereo- and enantiomeric ratio by lowering the temperature to -30ºC but the results previously obtained at rt could not be improved (entry 8).

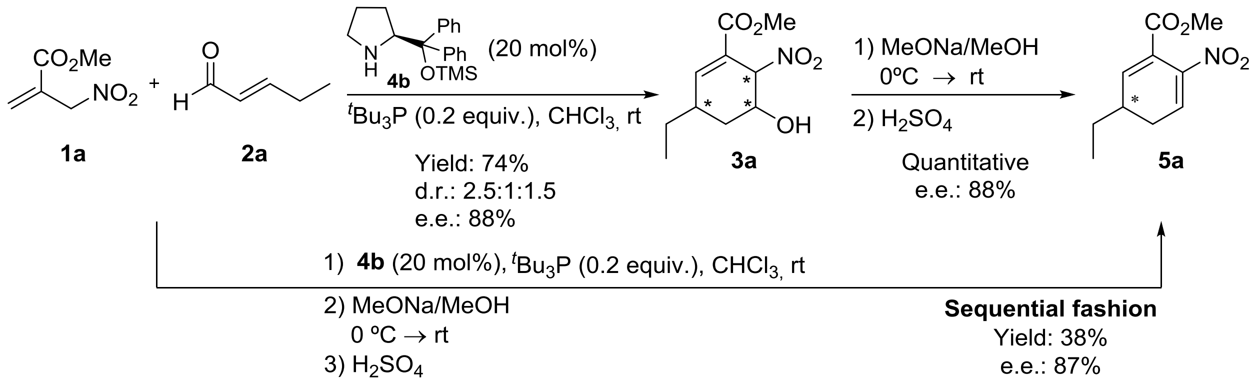
Table 2. Influence of the solvent and temperature in the cascade reaction.[a]



Entry	Solvent	T (ºC)	d.r.[b]	Yield (%)[c]	e.e. (%)[d]
1	$CHCl_3$	rt	2.5:1:1.5	74	88
2	Hexane	rt	2.5:1:1.5	58	86
3	Toluene	rt	2:1.8:1	36	88
4	THF	rt	10:0.7:1	40	87
5	AcOEt	rt	10:7:1	28	88
6	MeOH	rt	1.7:1.3:1	57	83
7	DMF	rt	2.5:1:1	40	87
8	$CHCl_3$	-30	2.5:1.5:1	30	86

[a] One equivalent of **1a** and two equivalents of aldehyde **2a** were used. [b] Determined by ^{1}H-NMR analysis for the mixture of diastereoisomers after flash column chromatography purification. [c] Referred to the mixture of diastereoisomers after flash column chromatography purification. [d] Calculated by HPLC for the major diastereoisomer.

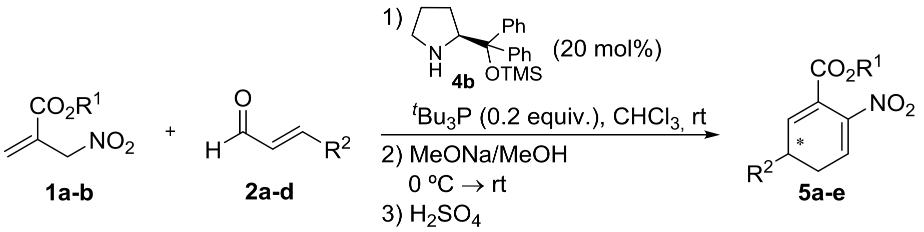
The poor diastereoselectivity achieved in all reactions might lay in the second step of the cascade reaction. Taking into account that the C-3 stereocenter that comes from the Michael addition of the nucleophile to the iminium ion is expected to be efficiently controlled by the catalyst, it seems sensible to think that the lack of stereogenic control derives from the intramolecular Henry reaction in which the stereocenters C-5 and C-6 are created, after the release of the organocatalyst. Aware of this diastereoselection problem, we thought about possible transformations that could lead to a convenient process in which all diastereoisomers would converge into a single product. The dehydration of the cyclohexene **3a**, obtaining cyclohexadiene **5a** seemed a promising way to eliminate two sterocenters and to simplify the structure of the compound, maintaining the stereocenter created at the initial Michael reaction step. Therefore, a dehydration reaction was carried out, employing sodium methoxyde in methanol observing that the reaction took place in a quantitative way, and fortunately, being possible to carry out in a sequential fashion in an efficient manner (Scheme 3).



Scheme 3. Chemical manipulation of adduct **3a**

With those results in our hands we decided to evaluate the importance of the substitution of the enals used as Michael acceptors and the substitution of the acrylate **1**, in order to see the influence of this substitution, in the sequential Michael/Henry/dehydration reaction. As it can be seen in Table 3, overall yields obtained were around 30-40% in all cases in which linear aliphatic substituents were present at the β-position of the enal regardless the length of the chain. On the other hand, the enantiomeric excesses increased when longer chains were introduced (entries 1-3). In contrast, when a bulkier element, such as an ^{i}Pr group, was placed at the β position of the α,β-unsaturated aldehyde (entry 4) the yield decreased drastically although the enantioselection was very high. Similarly when the bulkier acrylate **1b** was employed a poor yield of compound **5e** was achieved although in very good enantiomeric excess.

Table 3. Sequential Michael/Henry/dehydration reaction.[a]



Entry	R^1	Acrylate	R^2	Enal	Product	Yield (%)	e.e. (%)[b]
1	Me	**1a**	Et	**2a**	**5a**	38	87
2	Me	**1a**	Me	**2b**	**5b**	32	85
3	Me	**1a**	^{n}Bu	**2c**	**5c**	33	92
4	Me	**1a**	^{i}Pr	**2d**	**5d**	17	93
5	^{t}Bu	**1b**	Et	**2e**	**5e**	12	88

[a] One equivalent of **1a-b** and two equivalents of aldehyde **2a-d** were used. [b] Calculated by HPLC after flash column chromatography purification.

In a plausible mechanism proposed, a Michael addition of the deprotonated 2-nitromethylacrylate **1** to the iminium ion resulting from the fusion between the aminocatalyst **4a** and the α,β-unsaturated aldehyde **2**, would take place. After the catalyst is released, an intramolecular Henry cyclization occurs, achieving the cyclohexenes **3** which would undergo a sequential dehydration step to provide the corresponding cyclohexadienes **5**.

3. Materials and Methods

General procedure for the Michael/Henry cascade reaction: The α,β-unsaturated aldehyde **2a** (0.50 mmol) was added to a solution of (*S*)-α,α-diphenyl-2-pyrrolidinemethanol trimethylsilyl ether **4a** (0.05 mmol), tri-tertbutylphosphine (0.05 mmol) and acrylate **1a** (0.25 mmol) in chloroform (2 mL). The reaction was stirred at room temperature until full conversion. Afterwards the reaction mixture was diluted in diethyl ether (10 mL) and washed with $NaHCO_3$ (1 × 8 mL), brine (2 × 8 mL) and H_2O (2 × 8 mL). The organic extracts were dried over anhydrous Na_2SO_4 and the solvent was evaporated under vacuum. The crude was charged onto silica gel and subjected to flash column chromatography to obtain the cyclohexenes **3a**.

General procedure for the Sequential Michael/Henry/dehydration reaction: The α,β-unsaturated aldehyde **2a-d** (0.50 mmol) was added to a solution of (*S*)-α,α-diphenyl-2-pyrrolidinemethanol trimethylsilyl ether **4a** (0.05 mmol), tri-*tert*butylphosphine (0.05 mmol) and the acrylate **1a-b** (0.25 mmol) in chloroform (2 mL). The reaction was stirred at room temperature until full conversion. The solvent was evaporated under reduced pressure and after dissolving the reaction crude in dry MeOH (2 mL) it was added *via* canula to a solution of metallic sodium (2 mmol) in MeOH (4 mL) at 0 ºC. The reaction mixture was allowed to reach room temperature and stirred it for 1 hour. Then a solution of H_2SO_4/MeOH (1:5) was added until pH≈5 was achieved and the solvent was evaporated under vacuum. H_2O (20 mL) and CH_2Cl_2 (20 mL) were added and the aqueous layer was extracted with CH_2Cl_2 (3 × 20 mL). The organic extracts were dried over anhydrous Na_2SO_4 and the solvent was evaporated under vacuum. The crude was charged onto silica gel and subjected to flash column chromatography in order to achieve compound **5a-e**.

4. Conclusions

To sum up, we have developed a methodology that involves an organocatalytic Michael/Henry cascade reaction under iminium activation between enals and 2-nitromethylacrylates **1a-c**. The selected Michael donor, 2-nitromethylacrylate, has proved to be a suitable starting material for the reaction, acting as an appropriate 1,3-dinucleophile and has reacted with α,β-unsaturated aldehydes stereocontrolled by α,α-diphenylprolinol *O*-silylated catalyst, but showing poor diastereoselectivity.

In order to avoid this problem, we have designed a methodology involving Michael/Henry cascade reaction followed by a sequential dehydration leading to the enantiopure cyclohexadienes **5a-e** in a moderate yield but good enantioselectivity.

Acknowledgments

The authors thank the Spanish MINECO (CTQ2014-42107-P), the Basque Government (Grupos IT328-10) and UPV/EHU (UFI QOSYC 11/22 and fellowships to I.R. and N.F.) for financial support.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Ono, N. *The Nitro Group in Organic Synthesis*, Wiley-VCH: New York, 2001.
2. ***Enamine activation***: (a) Zhang, B.; Cai, L.; Song, H.; Wang, Z.; He, Z. A Highly Diastereoselective Tertiary Amine-Catalyzed Cascade Michael–Michael–Henry Reaction between Nitromethane, Activated Alkenes and α,β-Unsaturated Carbonyl Compounds. *Adv. Synth. Catal.* **2010**, *352*, 97–102. ***Iminium ion activation***: (b) Reyes, E.; Jiang, A.; Milleli, A.; Elsner, P.; Hazell, R. T.; Jørgensen, K. A. How to Make Five Contiguous Stereocenters in One Reaction: Asymmetric Organocatalytic Synthesis of Pentasubstituted Cyclohexanes. *Angew. Chem. Int. Ed.* **2007**, *46*, 9202-9205. (c) García-Ruano, J. L.; Marcos, V.; Suanzes, J. A.; Marzo, L.; Alemán, J. One-Pot Synthesis of Pentasubstituted Cyclohexanes by a Michael Addition Followed by a Tandem Inter–Intra Double Henry Reaction. *Chem. Eur. J.* **2009**, *15*, 6576-6580. ***H-Bonding activation***: (d) Varga, S.; Jakab, G.; Drahos, L.; Holczbauer, T.; Czugler, M.; Soós, T. Double Diastereocontrol in Bifunctional Thiourea Organocatalysis: Iterative Michael–Michael–Henry Sequence Regulated by the Configuration of Chiral Catalysts. *Org. Lett.* **2011**, *20*, 5416-5419. (e) Tan, B.; Chua, P. J.; Zeng, X.; Lu, M.; Zhong, G. A Highly Diastereo- and Enantioselective Synthesis of Multisubstituted Cyclopentanes with Four Chiral Carbons by the Organocatalytic Domino Michael−Henry Reaction. *Org. Lett.* **2008**, *10*, 3489-3492. (f) Rueping, M.; Kuenkel, A.; Fröhlich, R. Catalytic Asymmetric Domino Michael–Henry Reaction: Enantioselective Access to Bicycles with Consecutive Quaternary Centers by Using Bifunctional Catalysts. *Chem. Eur. J.* **2010**, *16*, 4173-4176. (g) Tan, B.; Lu, Y.; Zeng, X.; Chua, P. J.; Zhong, G. Facile Domino Access to Chiral Bicyclo[3.2.1]octanes and Discovery of a New Catalytic Activation Mode. *Org. Lett.* **2010**, *12*, 2682-2685. (h) Uehara, H.; Imashiro, R.; Hernández-Torres, G.; Barbas III, C. F. Organocatalytic asymmetric assembly reactions for the syntheses of carbohydrate derivatives by intermolecular Michael-Henry reactions. *PNAS*, **2010**, *107*, 20672-20677.
3. Martinez, J. I.; Villar, L.; Uria, U.; Carrillo, L.; Reyes, E.; Vicario, J. L. Bifunctional Squaramide Catalysts with the Same Absolute Chirality for the Diastereodivergent Access to Densely Functionalized Cyclohexanes through Enantioselective Domino Reactions. Synthesis and Mechanistic Studies. *Adv. Synth. Catal.* **2014,** *356*, 3627-3648.
4. Alonso, D. A.; Kitagaki, S.; Utsumi, N.; Barbas III, C. F. Towards Organocatalytic Polyketide Synthases with Diverse Electrophile Scope: Trifluoroethyl Thioesters as Nucleophiles in Organocatalytic Michael Reactions and Beyond. *Angew. Chem. Int. Ed.* **2008**, *47*, 4588-4591.

SciForum
Mol2Net

Improved Virtual Screening Performance through Docking Scoring Fusion in the Discovery of Dual Target Ligands for Parkinson's Disease

Yunierkis Perez-Castillo*,[1,2], Aliuska Morales Helguera[2], M. Natália D. S. Cordeiro[3], Eduardo Tejera[4], Cesar Paz-y-Miño[4], Aminael Sánchez-Rodríguez[5], Fernanda Borges*,[6], Maykel Cruz-Monteagudo[4,6]

[1] Sección Físico Química y Matemáticas, Departamento de Química, Universidad Técnica Particular de Loja, San Cayetano Alto S/N, EC1101608 Loja, Ecuador;

[2] Molecular Simulation and Drug Design Group, Centro de Bioactivos Químicos (CBQ), Central University of Las Villas, Santa Clara, 54830, Cuba;

[3] REQUIMTE, Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal;

[4] Instituto de Investigaciones Biomédicas (IIB), Universidad de Las Américas, 170513 Quito, Ecuador;

[5] Departamento de Ciencias Naturales, Universidad Técnica Particular de Loja, Calle París S/N, EC1101608 Loja, Ecuador;

[6] CIQUP/Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Porto 4169-007, Portugal.

* Author to whom correspondence should be addressed; E-Mail: yperez@utpl.edu.ec (YPC); fborges@fc.up.pt (FB)
Tel.: +351 220402502; Fax: +351 220402659.

Published: 4 December 2015

Abstract: Virtual methodologies have become essential components of the drug discovery pipeline. Specifically, structure-based drug design methodologies exploit the 3D structure of molecular targets to discover new drug candidates through molecular docking. Recently, dual target ligands of the Adenosine A2A Receptor and Monoamine Oxidase B enzyme have been proposed as effective therapies for the treatment of Parkinson's disease. To the best of our knowledge, no theoretical study has been devoted to developing structure-based virtual screening methodologies for the discovery of dual A_{2A}AR antagonists and MAO-B inhibitors. In this communication we propose a structure-based methodology for the discovery this type of molecules

Keywords: Parkinson's disease, Dual target ligands, Molecular docking, Scoring fusion.

1. Introduction

During the last decades, Virtual Screening (VS) methodologies have emerged as efficient alternatives to the expensive, in terms of time and money, High Throughput Screening (HTS) approaches for the discovery of new drug candidates [1]. In terms of efficiency, the hit rates obtained when VS tools are employed to filter large databases of chemical compounds are considerably higher than those obtained with HTS techniques [2]. Literature reports where VS experiments conducted to the identification of hit molecules in a wide range of application can be found elsewhere [3,4]

VS techniques can be divided into two main categories: Structure-Based VS (SBVS) and Ligand-Based VS (LBVS) [5]. The first one includes all the modeling approaches such as Molecular Docking and Molecular Dynamics that depend on the structure of a molecular receptor. This kind of approach uses the three-dimensional structure of the receptor to study its interactions with a set of putative ligands. Depending on the amount of ligands to study and the available computational resources a more or less detailed representation of the receptor-ligands interactions is chosen to estimate the stability of the receptor-ligand complexes. The computed scores serve then to order the investigated compounds from higher to lower probabilities of binding to the receptor [6].

One of the factors that negatively affect the performance of Molecular Docking SBVS studies is the accuracy of the scoring functions. Given that no scoring function can capture all the information relevant for the receptor-ligand binding process, the fusion of different scoring functions has been proposed as an alternative to improve the performance of SBVS methods [7]. These applications range from general ones intended for obtaining the best consensus strategy for any SBVS problem [8,9] to others proposed for specific researches [10,11]. In all these reports the proposed ensemble (fusion) methods outperform the VS performance obtained with a single scoring fusion.

In addition, usually SBVS methodologies are evaluated employing only a small set of decoy molecules. In the case of the standard DUD-E database only 50 decoy molecules can be selected per ligand [12]. This ligands/decoys proportion is far from what is observed in a real screening scenario where the ratio of active molecules is ranges from 0.01 to 0.14% [13]. To address this situation we have previously proposed a home-made algorithm for the generation of larger decoys sets resembling the ligands/decoys ratio of a real screening campaign [14].

On the other side, Parkinson Disease (PD) causes chronic disability and it is the second commonest degenerative condition of the nervous system. The standard treatment for PD is levodopa, which helps to increase the dopamine levels in the brain[15]. However there is a need of finding alternative therapies since levodopa has many side effects and can become ineffective over time. To this end, multicomponent therapies (combination of different drugs) have been used. However the discovery of new multi-target drugs (a single molecule that acts on multiple targets) is attracting more and more attention[16]. Multi-target drugs, compared with the use of combinations of different drugs, have more predictable pharmacokinetic and pharmacodynamic relationships as a consequence of the administration of a single drug [17].

Antagonists of A_{2A} adenosine receptors (A_{2A}AR) with monoamine oxidase B (MAO-B) inhibitory activity are a class of promising dual-target drugs for PD [18]. Thus, there is a need for developing novel and diverse drugs, antagonists of A_{2A}AR with MAO-B inhibitory activity for PD.

In this report we propose a structure-based methodology, which is extensively validated, for the discovery this type of molecules. The proposed methodology involves the molecular docking to both A_{2A}AR and MAO-B of a set of 25744 molecules containing 16 known dual target ligands and decoy molecules. The obtained docking poses are rescored using six different scoring functions for the two molecular targets. Then we investigate several aggregation schemes with the objective of maximizing the enrichment of known ligands at the beginning of the ranked list they produce. Finally, we show that the developed methodology provides high values of enrichment of known ligands, which outperform that of the individual scoring functions. At the same time, the obtained ensemble can be translated in a sequence of steps that should be followed to maximize the enrichment of dual target dual A_{2A}AR antagonists and MAO-B inhibitors.

.

.

2. Results and Discussion

The receptors, ligands and decoys were prepared for molecular docking calculations as described in the Materials and Methods section. The validation dataset consisting of the combination of the 16 known dual MAO-B inhibitors- A_{2A}AR Antagonists and decoy molecules was docked to both receptor structures following the protocol described in the Materials and Methods.

In all cases analyzed from here on, the best molecular docking protocol was selected as the scoring scheme providing the highest value of BEDROC among those achieving the maximum EF at three different selection sizes (1, 5 and 10 percent of screened data). We separately analyzed the results obtained for the raw and weighted by number of heavy atoms scores. Scoring schemes were produced by fusing the ranks derived from the scoring functions using either arithmetic or geometrical mean as described in the Materials and Methods section.

Different Fusion Schemes (FS) were assayed in this investigation and they can be classified into two groups. The first group consisted in fusing the scoring functions maximizing the enrichment of dual ligands for the A_{2A}AR and MAO-B enzyme separately. The application of the optimal scoring scheme of each target yields one fused ranking of compounds for each one. Then these two fused ranks were aggregated in one final rank. By employing this first fusion scheme we ensure that the final ranking will be based upon information derived from both the A_{2A}AR and the MAO-B enzyme. This fusion scheme will be referred as Fusion Scheme 1 (FS1) from here on.

The second group consisted in evaluating the performance of all possible ensembles resulting from all possible combinations of the individual scoring functions ranks obtained for both targets at the same time. Since the number of scoring functions employed in this study is small, it was possible to evaluate all their possible combinations of size 1 to 2N, being N the number of computed scoring functions per target. For this second approach no constrain is imposed during the modeling process regarding the need of information from both targets in the final ensemble. Therefore, there is the possibility that, in opposition to the expected behavior, the best performing ensemble would contain information from only one of the two molecular targets. This fusion scheme will be referred as Fusion Scheme 2 (FS2) from here on.

As mentioned before, we tested the arithmetic and geometric means as fusion operators. FS1 contains three aggregation steps: the aggregation of $A_{2A}AR$ scoring functions, the aggregation of MAO-B scoring functions and the aggregation of the rankings obtained for both targets. In this case all possible combinations of both fusion operators were tested. That is, scoring functions were first aggregated using the same fusion operator, either arithmetic or geometric mean, for each target separately. Then in the second step the aggregated ranking for each target was fused using both aggregation operators. Considering that the aggregation experiments are conducted with the raw scores and with the scores weighted by number of heavy atoms, the proposed setup provides eight different variants of FS1. These variants are summarized in Table 1.

For FS2, since the scores derived for both targets are considered together, there is only one rankings fusion step. Thus, considering that we studied the raw scores and the scores weighted by number of heavy atoms, four different setups were assayed. The different FS assayed in this scenario are summarized in Table 2.

For each known ligand, 1607 decoys were selected following the procedure described in our previous publication [14]. This amount of decoys provides a ratio of active to decoy compounds of 0.06%, which resembles a real screening scenario [13]. In this case the maximum EF that any of the individual scoring functions can achieve is when the selection size is set to 1% of the ranked list is 6.23. The best performing FS is presented in Table 3.

Our results show that the best scoring schemes are obtained when the docking scores to MAO-B and $A_{2A}AR$ are considered together during the scoring fusion procedure and scores fusion using arithmetic mean provides better results than fusion using geometrical mean. Also, in all the examined cases the best scoring scheme obtained with the raw docking scores outperforms that obtained fusing the scores weighted by the number of heavy atoms. It should also be noted that in almost every case, the best enrichment is derived from more than one scoring function through their fusion.

For the current validation setup the maximum values that the EF can reach are 100, 20 and 10 when 1%, 5% and 8% of screened data are selected respectively. Taking this into consideration it can be seen that if 1% of the screened data is selected for further analyses the resulting virtual screening protocol is able of achieving 31.17% of the theoretical maximum enrichment. Following the same reasoning, when the 5% and 8% of screened data are selected the corresponding virtual screening tools achieve 75% and 87.5% of the theoretical maximum of the EF respectively. Last but not least, the BEDROC values obtained for these virtual screening protocols are away from random (BEDROC=0). The accumulative curves corresponding to the three optimal virtual screening protocols are presented in Figure 1.

The obtained results provide a set of approaches from which we can select the optimal one for the virtual screening of databases of chemical compounds in the search of dual MAO-B inhibitors- $A_{2A}AR$ Antagonists. For example, if we were to select the appropriate virtual screening protocol for screening a database of chemicals and select 1% of data for further analysis, we should follow this procedure:

1. Dock the database to both MAO-B and A_{2A}AR.
2. Select the best pose of each compound in each target according to the grid-based scoring function.
3. Rescore the best poses in A_{2A}AR using the GB/SA Score, Continuous Score; Amber Score (everything rigid) and Amber Score (flexible ligand) scoring functions
4. Rescore the best poses in MAO-B using the. GB/SA Score and SA_Descriptor Score scoring functions.
5. Generate the individual ranking produced by the scoring functions GB/SA Score, Continuous Score; Amber Score (everything rigid) and Amber Score (flexible ligand) for A_{2A}AR and GB/SA Score and SA_Descriptor Score for. MAO-B.
6. Fuse the obtained individual rankings using arithmetic mean.

.

Table 1. Variants of FS1 assayed

Fusion Scheme [a]	Scores Type [b]	Target Scores Fusion [c]	Final fusion [d]
FS1.1	Raw	Arithmetic Mean	Arithmetic Mean
FS1.2	Raw	Arithmetic Mean	Geometric Mean
FS1.3	Raw	Geometric Mean	Arithmetic Mean
FS1.4	Raw	Geometric Mean	Geometric Mean
FS1.5	Weighted	Arithmetic Mean	Arithmetic Mean
FS1.6	Weighted	Arithmetic Mean	Geometric Mean
FS1.7	Weighted	Geometric Mean	Arithmetic Mean
FS1.8	Weighted	Geometric Mean	Geometric Mean

[a] Fusion scheme identifier

[b] Type of score the rankings are derived from, either the raw scores or the scores weighted by the number of heavy atoms

[c] Fusion operator employed to fuse the rankings derived of each scoring function in each target

[d] Fusion operator employed to aggregate the fused rankings obtained for each target

Table 2. Variants of FS2 assayed

Fusion Scheme [a]	Scores Type [b]	Target Scores Fusion [c]	Final fusion [d]
FS1.1	Raw	Arithmetic Mean	Arithmetic Mean
FS1.2	Raw	Arithmetic Mean	Geometric Mean
FS1.3	Raw	Geometric Mean	Arithmetic Mean
FS1.4	Raw	Geometric Mean	Geometric Mean
FS1.5	Weighted	Arithmetic Mean	Arithmetic Mean
FS1.6	Weighted	Arithmetic Mean	Geometric Mean
FS1.7	Weighted	Geometric Mean	Arithmetic Mean
FS1.8	Weighted	Geometric Mean	Geometric Mean

[a] Fusion scheme identifier

[b] Type of score the rankings are derived from, either the raw scores or the scores weighted by the number of heavy atoms

[c] Fusion operator employed to fuse the rankings derived of each scoring function in each target

[d] Fusion operator employed to aggregate the fused rankings obtained for each target

Table 3. Enrichment metrics for the best performing FS

FS Method [a]	EF [b]	BEDROC [c]	AUAC [d]	Fused Scoring Functions [e]
FS2.1	31.17	0.11	0.87	A_{2A}AR: 3, 5, 6 ,7 MAO-B: 3, 4

[a] Employed fusion method. See Tables 2 and 3 for the detailed setup of each method

[b] Enrichment Factor for the best scoring scheme.

[c] BEDROC for the best scoring scheme. Alpha value is set to 160.9.

[d] Area Under the Accumulative Curve for the best scoring scheme.

[e] Scoring functions fused in the best scoring scheme. The following numbering is employed for scoring functions: 1) Grid Score; 2) PB/SA Score; 3) GB/SA Score; 4) SA_Descriptor Score; 5) Continuous Score; 6) Amber Score, everything rigid and 7) Amber Score, flexible ligand.

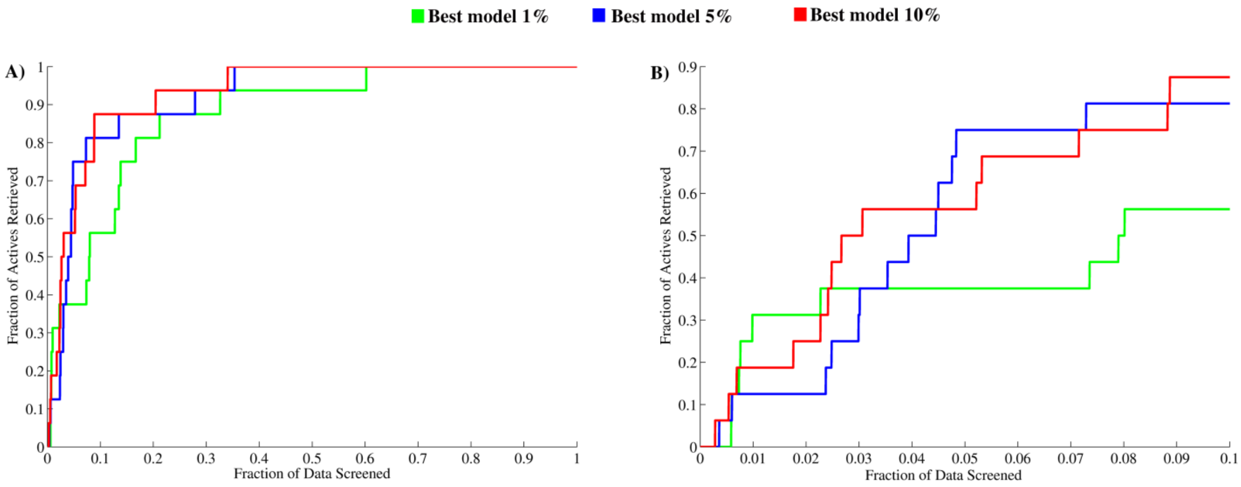


Figure 1. Accumulative curves obtained for the best virtual screening protocol when 1%, 5% and 8% of screened data are selected for further analysis. A) Complete curves. B) Curves for the first 10% of screened data.

3. Materials and Methods

Receptor preparation

The crystallographic structures of the A_{2A}AR in complex with the antagonist ZM241385, PDB code 3PWH and of the MAO-B in complex with a coumarin inhibitor, PDB code 2V61, were obtained from the Protein Data Bank (www.wwpdb.org) database [19]. Receptor preparation was carried out with UCSF Chimera software.[20] During receptor preparation all water molecules and ligands were removed and hydrogen atoms and charges were added. For both receptors the ligand binding pocket was defined as any residue lying at a distance below 5Å from the crystallographic ligand structure.

Ligand preparation

Sixteen known dual MAO-B inhibitors-A_{2A}AR Antagonists were compiled from the literature [21,22]. Three dimensional conformers for the compounds were generated using the OMEGA software [23]. A maximum of 500000 conformations per molecule were generated using an energy window of 100 kcal/mol. All rotatable bonds were considered during the torsion search using the Merck Molecular Force Field (MMFF) and duplicate conformers were discarded based on a RMS value of 0.5 Å. A maximum number of 200 conformers were saved for each compound. Afterwards, AM1-BCC charges were added to each conformer using the MOLCHARGE programs that is part of the QUACPAC package.[24].

Decoy molecules selection

Decoys selection was a based on a desirability-based home-developed algorithm that has been previously employed in the selection of tailored decoy sets for the validation of virtual screening strategies [14]. Decoys were prepared for molecular docking following the same protocol described above for ligands.

Molecular Docking

Molecular docking was performed with the DOCK v6.6 software.[25] A maximum of 2000 orientations per ligand was explored allowing a maximum of two bumps between the ligand and the receptor. Bumps were defined as any pair of atoms closer than the 75% of the sum of their Van der Waals radii. The energy grid-based scoring function was selected for poses quality evaluation. The pose with the lowest score for each ligand conformer was saved, allowing for a maximum of 200 saved poses.

For interaction energies calculation, a grid was pre-computed for the receptor binding pocket region. The grid spacing was set to 0.3 Å and the attractive and repulsive Van der Waals coefficients were set to 6 and 12 respectively. Calculations were performed considering an all-atoms model.

Molecular docking post-processing

The molecular docking protocol described above results in 200 docked conformations of each compound being saved. For every compound the best scored conformation was selected for further rescoring using six scoring functions implemented in DOCK. The scoring functions used for poses rescoring were: PB/SA Score, AMBER Score considering the whole complex as rigid, AMBER Score considering the ligand as flexible, Hawkins GB/SA Score and Solvent Accessible Surface Area (SASA) Score. These rescoring calculations plus the previous grid-based scoring employed for poses evaluation and selection provide seven different ways of evaluating the ligand-receptor interaction energies. In addition to the raw docking scores, the scoring value of each compound was weighted by the number of heavy atoms on it.

The seven computed scoring functions were used for the implementation of a consensus ranking scheme. Instead of combining the raw scoring values coming from different scoring functions, the ranks produced by these scoring functions were combined following the procedure described next. Firstly, the rank derived from each scoring function was produced. Then, for a specific combination of scoring functions, a fused rank was computed as either the arithmetic or geometric mean of the compound’s rank in the individual models.

To evaluate the performance of the developed models in a virtual screening scenario the following metrics were computed: Area Under the Accumulation Curve (AUAC); Area under the Receiver Operating Characteristic Curve (ROC); Enrichment factor (EF) and Boltzmann-enhanced discrimination of ROC (BEDROC).[26] Here the same definitions proposed by Truchon *et al.* are used.[26]

..

4. Conclusions

We investigated different variants of docking scores fusion for maximizing the enrichment of dual target ligands of the Adenosine A2A Receptor and the Monoamine Oxidase B enzyme in virtual screening experiments. Our results show that for achieving high values of dual ligands enrichment, information relative to docking scores to both targets have to be combined. In addition, no single scoring function can be employed for achieving good virtual screening performance. Instead, combining the rankings derived from different scoring functions proved to be a valuable strategy for improving the enrichment relative to single scoring function in virtual screening experiments.

Acknowledgments

MC-M acknowledges the postdoctoral grant [SFRH/BPD/90673/2012] financed by the FCT – Fundação para a Ciência e a Tecnologia, Portugal, co-financed by the European Social Fund.

Conflicts of Interest

The authors declare no conflict of interests.

References and Notes

1. Macalino, S.J.; Gosu, V.; Hong, S.; Choi, S., Role of computer-aided drug design in modern drug discovery. *Archives of pharmacal research* **2015**, *38*, 1686-1701.

2. Zhu, T.; Cao, S.; Su, P.C.; Patel, R.; Shah, D.; Chokshi, H.B.; Szukala, R.; Johnson, M.E.; Hevener, K.E., Hit identification and optimization in virtual screening: Practical recommendations based on a critical literature analysis. *Journal of Medicinal Chemistry* **2013**, *56*, 6560-6572.

3. Castillo-Gonzalez, D.; Mergny, J.L.; De Rache, A.; Perez-Machado, G.; Cabrera-Perez, M.A.; Nicolotti, O.; Introcaso, A.; Mangiatordi, G.F.; Guedin, A.; Bourdoncle, A., *et al.*, Harmonization of qsar best practices and molecular docking provides an efficient virtual screening tool for discovering new g-quadruplex ligands. *Journal of Chemical Information and Modeling* **2015**, *55*, 2094-2110.

4. Miller, Z.; Kim, K.S.; Lee, D.M.; Kasam, V.; Baek, S.E.; Lee, K.H.; Zhang, Y.Y.; Ao, L.; Carmony, K.; Lee, N.R., *et al.*, Proteasome inhibitors with pyrazole scaffolds from structure-based virtual screening. *Journal of Medicinal Chemistry* **2015**, *58*, 2036-2041.

5. Lill, M., Virtual screening in drug design. *Methods in molecular biology (Clifton, N.J.)* **2013**, *993*, 1-12.

6. Ferreira, L.G.; Dos Santos, R.N.; Oliva, G.; Andricopulo, A.D., Molecular docking and structure-based drug design strategies. *Molecules (Basel, Switzerland)* **2015**, *20*, 13384-13421.

7. Yuriev, E.; Holien, J.; Ramsland, P.A., Improvements, trends, and new ideas in molecular docking: 2012-2013 in review. *Journal of molecular recognition : JMR* **2015**, *28*, 581-604.

8. Wang, R.; Lai L Fau - Wang, S.; Wang, S., Further development and validation of empirical scoring functions for structure-based binding affinity prediction.

9. Yang, J.M.; Chen, Y.F.; Shen, T.W.; Kristal, B.S.; Hsu, D.F., Consensus scoring criteria for improving enrichment in virtual screening. *Journal of Chemical Information and Modeling* **2005**, *45*, 1134-1146.

10. Daga, P.R.; Polgar We Fau - Zaveri, N.T.; Zaveri, N.T., Structure-based virtual screening of the nociceptin receptor: Hybrid docking and shape-based approaches for improved hit identification.

11. Parmenopoulou, V.; Kantsadi, A.L.; Tsirkone, V.G.; Chatzileontiadou, D.S.; Manta, S.; Zographos, S.E.; Molfeta, C.; Archontis, G.; Agius, L.; Hayes, J.M.*, et al.*, Structure based inhibitor design targeting glycogen phosphorylase b. Virtual screening, synthesis, biochemical and biological assessment of novel n-acyl-beta-d-glucopyranosylamines. *Bioorganic & Medicinal Chemistry* **2014**, *22*, 4810-4825.

12. Mysinger, M.M.; Carchia, M.; Irwin, J.J.; Shoichet, B.K., Directory of useful decoys, enhanced (dud-e): Better ligands and decoys for better benchmarking. *Journal of Medicinal Chemistry* **2012**, *55*, 6582-6594.

13. Bender, A.; Bojanic, D.; Davies, J.W.; Crisman, T.J.; Mikhailov, D.; Scheiber, J.; Jenkins, J.L.; Deng, Z.; Hill, W.A.; Popov, M.*, et al.*, Which aspects of hts are empirically correlated with downstream success? *Current Opinion in Drug Discovery & Development* **2008**, *11*, 327-337.

14. Perez-Castillo, Y.; Cruz-Monteagudo, M.; Lazar, C.; Taminau, J.; Froeyen, M.; Cabrera-Perez, M.A.; Nowe, A., Toward the computer-aided discovery of fabh inhibitors. Do predictive qsar models ensure high quality virtual screening performance? *Mol Divers* **2014**, *18*, 637-654.

15. Brunton, L.L., *Goodman & gilman's the pharmacological basis of therapeutics*. 11th ed.; The McGraw-Hill: New York, NY, 2007.

16. Youdim, M.B.; Geldenhuys, W.J.; Van der Schyf, C.J., Why should we use multifunctional neuroprotective and neurorestorative drugs for parkinson's disease? *Parkinsonism Rel Disord* **2007**, *13*, S281-S291.

17. Morphy, R.; Kay, C.; Rankovic, Z., From magic bullets to designed multiple ligands. *Drug Discovery Today* **2004**, *9*, 641-651.

18. Petzer, J.P.; Castagnoli, N.; Schwarzschild, M.A.; Chen, J.F.; Van der Schyf, C.J., Dual-target-directed drugs that block monoamine oxidase b and adenosine a2a receptors for parkinson's disease. *Neurotherapeutics* **2009**, *6*, 141-151.

19. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E., The protein data bank. *Nucleic Acids Research* **2000**, *28*, 235-242.

20. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E., Ucsf chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **2004**, *25*, 1605-1612.

21. Stossel, A.; Schlenk, M.; Hinz, S.; Kuppers, P.; Heer, J.; Gutschow, M.; Muller, C.E., Dual targeting of adenosine a(2a) receptors and monoamine oxidase b by 4h-3,1-benzothiazin-4-ones. *J Med Chem* **2013**, *56*, 4580-4596.

22. S., R.; Piersanti, G.; Bartoccini, F.; Diamantini, G.; Pala, D.; Riccioni, T.; Stasi, M.A.; Cabri, W.; Borsini, F.; Mor, M.*, et al.*, Synthesis of (e)‑8-(3-chlorostyryl)caffeine analogues leading to 9-deazaxanthine derivatives as dual a2a antagonists/mao-b inhibitors. *J Med Chem* **2013**, *56*, 1247−1261.

23. OpenEye Scientific Software, I. *Omega*, 2.3.2; Santa Fe, NM, USA, 2008.

24. OpenEye Scientific Software, I. *Quacpac*, 1.3.1; Santa Fe, NM, USA, 2008.

25. Lang, P.T.; Brozell, S.R.; Mukherjee, S.; Pettersen, E.F.; Meng, E.C.; Thomas, V.; Rizzo, R.C.; Case, D.A.; James, T.L.; Kuntz, I.D., Dock 6: Combining techniques to model rna-small molecule complexes. *RNA* **2009**, *15*, 1219-1230.
26. Truchon, J.F.; Bayly, C.I., Evaluating virtual screening methods: Good and bad metrics for the "early recognition" problem. *Journal of Chemical Information and Modeling* **2007**, *47*, 488-508.

SciForum
Mol2Net

Dengue NS5 Global Consensus Sequence Development to Find Conserved Region for Antiviral Drug Development

Shahid Mahmood [1,], Usman Ali ashfaq[2]*

[1] Department of Bioinformatics and Biotechnology, Government College University (GCU), Faisalabad, Pakistan, E-Mail: shahidbnb2013@gmail.com.

[2] Department of Bioinformatics and Biotechnology, Government College University (GCU), Faisalabad, Pakistan,

Address; Department of Bioinformatics and Biotechnology, Government College University (GCU), Faisalabad, Pakistan E-Mails: usmancemb@gmail.com;

* Author to whom correspondence should be addressed; E-Mail: usmancemb@gmail.com; Tel.: +92-331-4728790.

Published: 4 December 2015

Abstract:

Objective: To draw a representing consensus sequence of each DENV serotype, align all four consensus sequences to draw a global consensus sequence and also study the highly conserved residues. Methods: A total of 376 DENV NS3 sequences, belonging to four serotypes, reported from all over the world were aligned to develop global consensus sequence. Results: The active site residues Met343, Thr366, which are involved in nuclear localization and also interact with the NS3 viral, are highly conserved among all the DENV serotypes. Cys450, Gly466 and Ala468, Arg482 are highly conserved in all the serotypes. Structural zinc (Zn1) sited consist of Cys-446, Cys-449, His-441, and the carboxylate group of Glu-437. This pocket is also found near the functionally important residues Ser-710 and Arg-729, which bind to the incoming rNTP. Meth530, Thr543 Asp597, Glu616 and Arg659, Pro671 are structurally conserved in all serotypes. Identification of four out of six conserved sequence motifs accountable for NTP binding and GDD catalytic active site, located in the palm domain. Leu766, Ala776 residues have high conservancy in all serotypes are observed in consensus sequence analysis. The thumb domain also has Zinc binding site (Zn2) and is synchronized by His-712, His-714, Cys-728 of motif E, and Cys-847. Pharmacological blockage of cavity B could potentially lead to suppression of initiation of the viral RNA synthesis and/or inhibition of NS3/NS5 interaction. Thirteen different peptides from the highly conserved regions of DENV NS5 protein were drawn which can be used to develop peptidic inhibitors. Conclusions: In spite of a high mutation rate in DENV, the residues which are present in the Nuclear localization signal (NLS), Di-valet ion binding sites, NTP binding, GDD catalytic active site, Thumb domain, priming loop are highly conserved. These are target sites for the development of antiviral agents or peptide vaccines.

Keywords: Pharmacology, Infectious diseases, Consensus sequence, Antiviral drug target

Mol2Net YouTube channel: *http://bit.do/mol2net-tube*

1. Introduction

Main text paragraph.

Dengue infection has become a major health problem in >100 countries of Africa, Asia, America, Western Pacific and the Eastern Mediterranean [1]. During the recent epoch, dengue infection has caused many endemics in Pakistan and thus, has become a major health issue in Pakistan. The global incidence of dengue infection has increased and estimated 50-100 million cases of dengue infections are reported annually from more than 100 tropical and subtropical countries of the world [2]. Dengue virus has four distinct serotypes and phylogenetically unique dengue viruses (DEN1-DEN4)[3]. DENV-2 serotype was most prevalence circulating serotype in Pakistan. Two types of infections caused by Dengue virus, fluctuating from a dengue fever to a more severe infection that can cause dengue haemorrhagic fever and dengue shock syndrome [1].

The dengue virus belongs to a member of the Flaviviradae that causes a wide range of diseases, including benign febrile illness, dengue fever (DF), plasma leakage syndrome and dengue haemorrhagic fever/dengue shock syndrome (DHF/DSS)[4]. The dengue virus is transmitted to humans by Aedes aegypti and Aedes albopictus mosquitoes [5,6]. The dengue virus is an enveloped, ssRNA positive-strand and comprises 11 kilo base long viral genome[7] having three structural proteins C (Capsid , M (Membrane) and E (envelop) [8] and seven Non-structural proteins NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5 which play an integral part in viral pathology[4,9].

All Non-structural proteins including NS5 are involved in enzymatic reactions which plays an important role in viral replication. NS5 is the most prominent Flavivirus protein due to its high molecular weight is 105 kDa. NS5 proteinsharing the minimum of 67% identity through all serotypes of DENV that’s why NS5 served as highly conserved viral domain. It comprises an N-terminal methyl-transferase domain (MTase) domain covers 1 to 296 amino acid residues.

The MTase activity of NS5 is liable for both guanine N-7 and ribose 2_-O methylations and a C-terminal RNAdependent RNA polymerase (RdRp) domain leads 270 to 900 amino acids and it is creditworthy for synthesizing a transient double-stranded replicative RNA intermediate [10-13]. All Flaviviruses has few highly conserved moieties, which lies between amino acid residues 320 and 368 of NS5 and play an important role in binding with beta-impotin and also with NS3[3]. These conserved residues can be imperative in developing unambiguous antiviral agents and inhibitors against dengue virus. In this study by using a unique approach to produce a substantiation consensus sequence NS5 will be helpful in designing anti-peptide to find a possible cure for dengue infection. The present study is designed to draw a global consensus sequence of the NS5 protein of dengue virus and to study DENV NS5 conserved domain function.

Results and Discussion

NS5 protein comprises of two domains, methyl-transferases (MTase) and RNA dependent RNA polymerases (RdRp). The dengue NS5 MTase domain involves in cap formation which recognized by host cell and RdRp domain plays vital role in viral genome replication. Consensus sequences of NS5 protein of all DENV serotypes (DENV 1-4) was drawn in CLC main workbench. Alignment was done to extract the highly conserved peptides among all serotypes which we were studied. Figure 1 shows the alignment of the consensus sequence of all the four DENV serotypes; the global consensus sequence is presented at the bottom. Conserved residues are shown with their corresponding symbols while the highly variable amino acids are denoted by "x" symbol. The alignment of all the consensus sequences will help us to study the considerably conserved residues in the DENV NS5 protein. Short peptides of 9 to 18 amino acids were designed from the highly conserved regions of the DENV NS5 consensus protein sequences; the sequence and position of these peptides are shown in the Table 1. These are the locations which are highly conserved and are the targets to design peptide vaccines or site specific inhibitors.

Dengue infection has become a global risk to human health to all over the world. Dengue virus has four serotypes and because of four different serotypes, there is no successful vaccine developed. NS5 domains become helpful for novel drug designing due to (1) largest DENV protein (2) its high conservancy (3) sharing minimum of the 67% identity through all serotypes of DENV. NS5 is multifunctional domain and has two domains N-terminal methyl-transferase domain (MTase) domain (1 to 296), C-terminal RNA-dependent RNA polymerase (RdRp) domain (270 to 900) [10, 12, 13] and possess the 37-amino acid inter-domain spacer sequence that contains a functional nuclear localization signal (NLS) between amino acid residues (320 and 405). RdRp contains three subdomains finger subdomain with zinc binding sites

Zn1 (Cys-446, Cys-449, His-441, Glu-437), palm domain with GDD catalytic active sites (Asp-663 and Asp-664), thumb domain with Zinc binding sites is Zn2 (His-712, His-714, Cys-728, Cys-847) and six motifs (A, B, C, D, E, F) at C-terminal region of NS5 that contains five amino acid sequence [14]. The consensus sequence analysis shows that Met343, Thr366 are highly conserved among all serotypes.

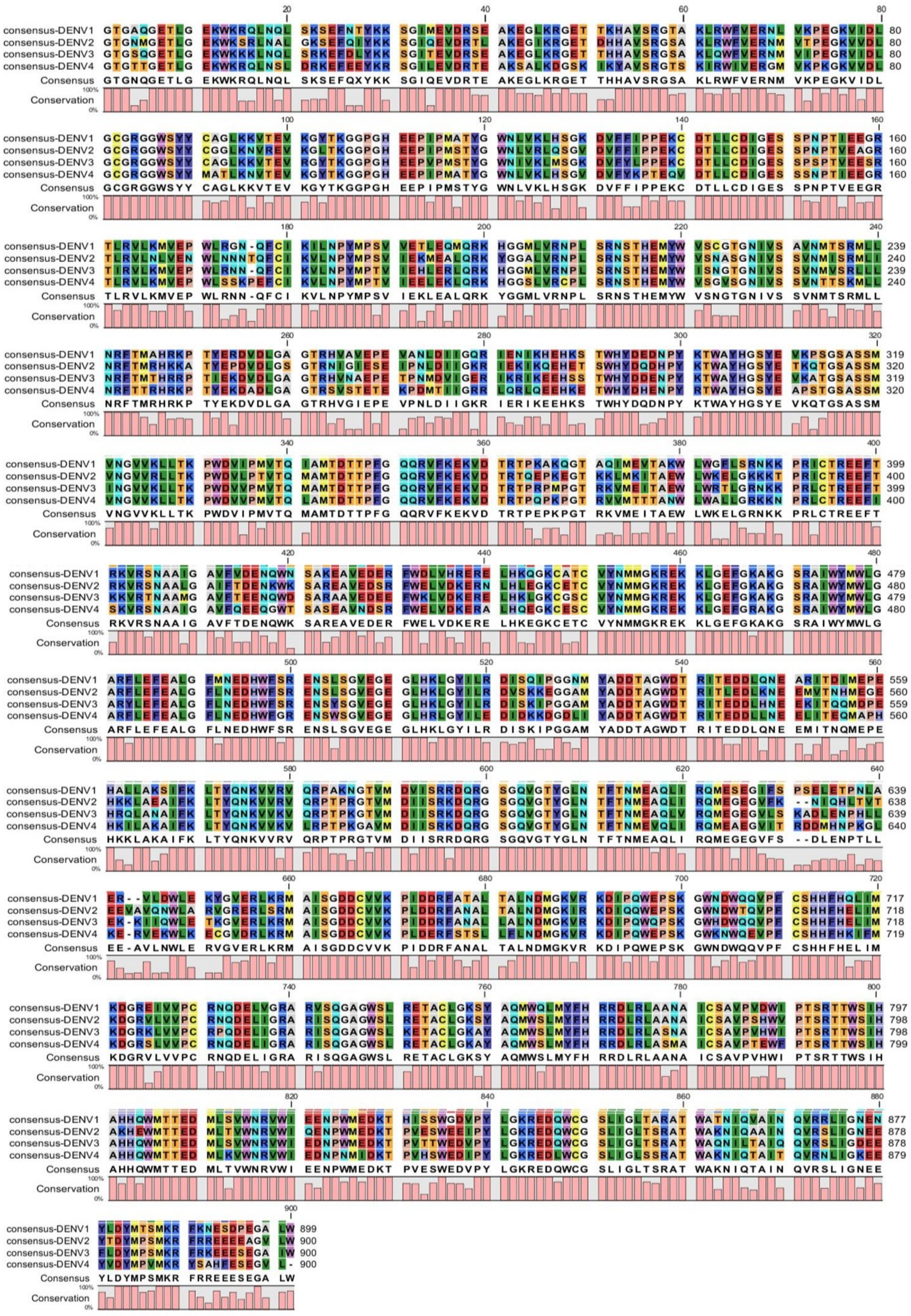
This conserved region involved in nuclear localization sequences (NLS). The NLS has been divided into alpha-NLS (spanning residues 320 to 368) and alpha/beta-NLS (Residues 369 to 405). The alpha-NLS region thought to interact with the NS3 viral helicase [3]. Interestingly, the NLS domain signatures are dispersed between the fingers and thumb subdomains [15]. The region which binds with beta-importin that recognized NLS and carry protein inside nucleus placed between this amino acid [16]. The consensus sequence analysis shows that Cys450, Gly466 and Ala468, Arg482 are highly conserved in all the serotypes. The base of the fingers domain arrangements a concave surface shaped by the solvent-exposed residues of helices (alpha 6, alpha 14, and alpha 15) near the N terminus of the protein. Structural zinc (Zn1) sited in the fingers subdomains is harmonized by Cys- 446, Cys-449, His-441, and the carboxylate group of Glu-437. The zinc ion Zn2 likely donates to the structural stability of the region adjacent motif E of the DENV polymerase. This pocket is also found near the functionally important residues Ser-710 and Arg-729, which bind to the incoming rNTP [11]. The consensus sequence analysis shows that Meth530, Thr543 Asp597, Glu616 and Arg659, Pro671 are conserved in all serotypes and lay under palm domain. It is the collection of a small antiparallel beta-strand platform, beta 4 and beta 5, surrounded by eight helices (alpha 11 to alpha 13 and alpha 16 to alpha 20). The palm domain performs to be the most structurally conserved among all known polymerases, reflecting the salvation of the design of the catalytic site during evolution. Identification of four of six conserved sequence motifs accountable for NTP binding and catalysis, located in the palm domain [17]. The GDD catalytic active site (motif C, comprising Asp-663 and Asp-664) is placed in the turn between strands beta 4 and beta [11]. Mutation in residue Asp-663 and Asp-664 to any negative charge residue can disrupt the function of RdRp. Leu766, Ala776 is thumb domain and has high conservancy in all serotypes. Thumb domain forms the C-terminal end of the RdRp of DENV, is the most structurally mutable among known polymerase structures. It covers two conserved sequence motifs. Motif E forms an antiparallel beta-sheet wedged between the palm domain and several alpha-helices of the thumb domain. A loop spanning amino acids 782 to 809 forms the priming loop that partially occludes the active site [18]. Three mutations (L328A, W859A, and I863A) reduced de novo RNA synthesis by >=85%. These L328A, W859A, and I863A three mutations Table 1.

Position and sequence of the peptides along with potential domain; which can be used as a peptide vaccine. Position of conserved peptides Sequence of peptides Domain encoding 79–90 No putative conserved domain, 104-113 No putative conserved domain, 141-151 No putative conserved domain, 209-220 No putative conserved domain, 342-363 Functional nuclear localization signal (NLS), 450- 466 N-terminus finger subdomain, 468-482 N-terminus finger subdomain, 530-543 Palm domain, 568-578, 597-616 Palm domain, 659-671 Palm domain, 766-776 Thumb domain, 791-801No putative conserved domain reduced viral replication by decreasing the initiation of RNA synthesis. These three mutations decreased RNA synthesis initiation (L328A, W859A, and I863A) are placed at the same line between the delta 1 loop and remaining RdRp DENV thumb subdomain. The K330A mutation abridged the NS3/NS5 interaction. Thumb domain also has Zinc binding site (Zn2) and is synchronized by His-712, His-714, Cys-728 of motif E, and Cys- 847 of helix alpha 26 and mutation in these residue abolish the activity of RdRp.

Table 1. Position and Sequence of the Peptides along with potential domain; which can be used as a Peptide Vaccine

Position of conserved Peptides	**Sequence of peptides**	**Domain encoding**
79 - 90	DLGCGRGGWSYY	No Putative conserved domain
104 - 113	TKGGPGHEEP	No Putative conserved domain
141 - 151	DTLLCDIGESS	No Putative conserved domain
209 - 220	PLSRNSTHEMYW	No Putative conserved domain
342 – 363	MAMTDTTPFGQQRVFKEKVDTRT	**Functional nuclear localization signal (NLS)**
450 - 466	CVYNMMGKREKKLGEFG	**N-terminus finger subdomain**
468 - 482	AKGSRAIWYMWLGAR	**N-terminus finger subdomain**
530 - 543	MYADDTAGWDTRIT	**Palm domain**
568 - 578	IFKLTYQNKVV	
597 - 616	DQRGSGQVGTYGLNTFTNME	**Palm domain**
659 - 671	RMAISGDDCVVKP	**Palm domain**
766 - 776	LMYFHRRDLRLA	**Thumb domain**
791 - 801	PTSRTTWSIHA	No Putative conserved domain

Figure 1. Multiple sequence alignment of consensus sequences of the Dengue NS5 of all serotypes.



3. Materials and Methods

2.1. Drawing consensus sequence of DENV NS5:

A total of 385 sequences of DENV NS5 of all serotypes were retrieved from the NCBI database. All the sequences were imported in CLC main workbench. Consensus sequences of all serotypes were developed by applying multiple sequence alignment feature of CLC main workbench. All four serotypes sequences were retrieved randomly from NCBI protein database. Hundred NS5 sequences of Dengue 1 serotype reported from USA, Indonesia, China, and Australia were used to developed consensus sequence. One hundred sequences of serotype 2 (DENV2) belongs to China, USA, and Taiwan were used to build the serotype two consensus sequence with the support of CLC workbench software. Ninety five sequences of serotype three (DENV3) belonged to the USA, Singapore, and China, were obtained to CLC workbench software to construct consensus sequence. Ninety sequence of serotype four (DENV4) retrieved from NCBI related to Indonesia, USA, China and Thailand were fetched to CLC workbench to produce consensus sequence of DENV4.

2.2. Peptides designing for potential peptide vaccine:

The consensus sequences of all the four serotypes (DENV1-DENV4) were taken up in CLC workbench software. These consensus sequences were aligned in the CLC workbench to develop the global consensus sequence. The consensus sequence was used to study variations in different motifs and domains of the DENV NS5 region. Short peptides from the highly conserved regions of the DENV NS5 protein were selected from the consensus sequence analysis; these peptides are the best targets to be tested as a potential peptide vaccine.

4. Conclusions

Our study suggests that there are certain stretches of amino acids, which take part in binding with divalent cat-ions, RNA Synthesis, ATPase binding, Nuclear Localization, Binding NS5 with NS3 domain and viral replication are highly conserved; and can be used as a potential target for the development of antiviral agents. Pharmacological blockage of cavity B could potentially lead to suppression of initiation of viral RNA synthesis and/or inhibition of NS3/NS5 interaction [19].

Acknowledgments

I would like to thanks all my colleagues at Government College University, Faisalabad.

Author Contributions

Both authors contribute equally.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Idrees, S. and U.A. Ashfaq, A brief review on dengue molecular virology, diagnosis, treatment and prevalence in Pakistan. Genet Vaccines Ther, 2012. 10(6).

2. Siregar, A.R., T. Wibawa, and N. Wijayanti, Early Detection and Serotyping of Dengue Viruses Clinical Isolates Using Reverse Transcription Polymerase Chain Reaction (RT-PCR) 2 Primers. Indonesian Journal of Biotechnology, 2012. 16(2).

3. Halstead, S., Dengue haemorrhagic fever—a public health problem and a field for research. Bulletin of the World Health Organization, 1980. 58(1): p. 1.

4. Guirakhoo, F., et al., Immunogenicity, genetic stability, and protective efficacy of a recombinant, chimeric yellow fever-Japanese encephalitis virus (ChimeriVax-JE) as a live, attenuated vaccine candidate against Japanese encephalitis. Virology, 1999. 257(2): p. 363-372.

5. Gubler, D.J. and G.G. Clark, Dengue/dengue hemorrhagic fever: the emergence of a global health problem. Emerging infectious diseases, 1995. 1(2): p. 55.

6. Johansson, M., et al., A small region of the dengue virus-encoded RNA-dependent RNA polymerase, NS5, confers interaction with both the nuclear transport receptor importin-beta and the viral helicase, NS3. J Gen Virol, 2001. 82(Pt 4): p. 735-45.

7. Rodenhuis-Zybert IA, Wilschut J, and S. JM., Dengue virus life cycle: viral and host factors modulating infectivity. Cell Mol Life Sci, 2010. 67: p. 2773-86.

8. Seema and S.K. Jain, Molecular mechanism of pathogenesis of dengue virus: Entry and fusion with target cell. Indian J Clin Biochem, 2005. 20(2): p. 92-103.

9. Smit, J.M., et al., Flavivirus cell entry and membrane fusion. Viruses, 2011. 3(2): p. 160-171.

10. Iglesias, N.G., C.V. Filomatori, and A.V. Gamarnik, The F1 motif of dengue virus polymerase NS5 is involved in promoter-dependent RNA synthesis. Journal of virology, 2011. 85(12): p. 5745-5756.

11. Yap, T.L., et al., Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. Journal of virology, 2007. 81(9): p. 4753-4765.

12. Egloff, M.P., et al., An RNA cap (nucleoside-2'-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. EMBO J, 2002. 21(11): p. 2757-68.

13. You, S., et al., In vitro RNA synthesis from exogenous dengue viral RNA templates requires long range interactions between 5'- and 3'-terminal regions that influence RNA structure. J Biol Chem, 2001. 276(19): p. 15581-91.

14. Bartholomeusz, A.I. and P.J. Wright, Synthesis of dengue virus RNA in vitro: initiation and the involvement of proteins NS3 and NS5. Arch Virol, 1993. 128(1-2): p. 111-21.

15. Brooks, A.J., et al., The interdomain region of dengue NS5 protein interacts with NS3 and host proteins. 2002.

16. Jans, D.A., C.K. Chan, and S. Huebner, Signals mediating nuclear targeting and their regulation: application in drug delivery. Med Res Rev, 1998. 18(4): p. 189-223.

17. Poch, O., et al., Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. Journal of General Virology, 1990. 71(5): p. 1153-1162.

18. Adachi, T., et al., The essential role of C-terminal residues in regulating the activity of hepatitis C virus RNA-dependent RNA polymerase. Biochimica et Biophysica Acta (BBA)-Proteins & Proteomics, 2002. 1601(1): p. 38-48.

19. Zou, G., et al., Functional analysis of two cavities in flavivirus NS5 polymerase. Journal of Biological Chemistry, 2011. 286(16): p. 14362-14372.

SciForum
Mol2Net

Virtual Screening Tailored Ensembles of QSAR Models for the Discovery of Dual A_{2A} Adenosine Receptor Antagonists / Monoamine Oxidase B Inhibitors

Aliuska Morales Helguera[1], Yunierkis Perez-Castillo[1,2], M. Natália D. S. Cordeiro[3], Eduardo Tejera[4], Cesar Paz-y-Miño[4], Aminael Sánchez-Rodríguez[5], Marta Teijeira Bautista[6,7], Evys Ancede-Gallardo[1], Fernando Cagide[8], Fernanda Borges*[,8], Maykel Cruz-Monteagudo[*,4,8]

[1] Molecular Simulation and Drug Design Group, Centro de Bioactivos Químicos (CBQ), Central University of Las Villas, Santa Clara, 54830, Cuba;

[2] Sección Físico Química y Matemáticas, Departamento de Química, Universidad Técnica Particular de Loja, San Cayetano Alto S/N, EC1101608 Loja, Ecuador;

[3] REQUIMTE, Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal;

[4] Instituto de Investigaciones Biomédicas (IIB), Universidad de Las Américas, 170513 Quito, Ecuador;

[5] Departamento de Ciencias Naturales, Universidad Técnica Particular de Loja, Calle París S/N, EC1101608 Loja, Ecuador;

[6] Departamento de Química Orgánica. Facultade de Química, Universidade de Vigo, 36310 Vigo, Spain;

[7] Instituto de Investigación Biomédica (IVIB), Universidade de Vigo, 36310 Vigo, Spain;

[8] CIQUP/Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Porto 4169-007, Portugal.

* Author to whom correspondence should be addressed; E-Mail: gmailkelcm@yahoo.es (MCM); fborges@fc.up.pt (FB); Tel.: +351 220402502; Fax: +351 220402659.

Published: 4 December 2015

Abstract: Virtual Screening methodologies have emerged as efficient alternatives for the discovery of new drug candidates. At the same time, ensemble methods are nowadays frequently used to overcome the limitations of employing a single model in ligand-based drug design. However, many applications of ensemble methods to this area do not consider important aspects related to both virtual screening and the modeling process. During the application of ensemble methods to virtual screening the proper validation of the models in virtual screening conditions is often neglected. Frequently no analysis is performed of the diversity of the ensemble members or no considerations regarding the applicability domain of the base model are made. In this research we propose a method employing genetic algorithms optimization for the generation of virtual screening tailored ensembles that address problems in the current applications of ensemble methods to virtual screening. The proposed methodology is successfully applied to the generation of ensemble models for the ligand-based virtual screening of dual target A2A adenosine receptor antagonists and MAO-B inhibitors as potential Parkinson's disease therapeutics.

.

Keywords: Dual-target drugs, Virtual screening, MAO-B inhibitors, A_{2A} adenosine receptors antagonist, Ensemble modeling, QSAR

1. Introduction

During the last decades, Virtual Screening (VS) methodologies have emerged as efficient alternatives to the expensive, in terms of time and money, High Throughput Screening (HTS) approaches for the discovery of new drug candidates [1]. In terms of efficiency, the hit rates obtained when VS tools are employed to filter large databases of chemical compounds are considerably higher than those obtained with HTS techniques [2]. Literature reports where VS experiments conducted to the identification of hit molecules in a wide range of application can be found elsewhere [3,4].

VS techniques can be divided into two main categories: Structure-Based VS (SBVS) and Ligand-Based VS (LBVS) [5]. The first one includes all the modeling approaches such as Molecular Docking and Molecular Dynamics that depend on the structure of a molecular receptor. In some cases the structure of the molecular target is not available or the research process focuses in a mechanism where a molecular target cannot be defined. It can also be the case that the amount of data to process is too large to complete a SBVS campaign in a reasonable amount of time. In these cases LBVS tools can aid in the drug discovery process. These types of techniques include Similarity methods, Shape-based methods, Pharmacophore modeling and Quantitative Structure-Activity Relationships (QSAR) studies [5,6]. Specifically, QSAR approaches have been used for VS in the early stages of drug discovery [7].

To exploit the full potential of QSAR modeling some general guidelines have to be followed. These guidelines have been summarized as best practices for QSAR [8]. In addition to these guidelines, in our previous research we pointed out the importance of the proper validation of QSAR models in VS conditions [9].

Ensemble methods (EM) have gained popularity among QSAR practitioners, providing a set of tools for combining the predictions of single models in a more robust and generalizable model [9-11]. Although the success of these methods in Ligand-Based Drug Design (LBDD) and LBVS, the currently published applications of ensemble methods to LBVS suffer from a limitation common to most LBDD workflows: their VS performance is not retrospectively evaluated prior to their use in VS campaigns. In addition, in most reports of applications of ensemble methods to LBVS no considerations are made regarding the diversity of the base classifiers. The optimal size of the developed ensembles is neither considered and all available base classifiers are combined instead. These two last factors can drastically reduce the performance of ensemble methods [12].

On the other side, Parkinson Disease (PD) causes chronic disability and it is the second commonest degenerative condition of the nervous system. The standard treatment for PD is levodopa, which helps to increase the dopamine levels in the brain [13]. However there is a need of finding alternative therapies since levodopa has many side effects and can become ineffective over time. To this end, multicomponent therapies (combination of different drugs) have been used. However the discovery of new multi-target drugs (a single molecule that acts on multiple targets) is attracting more and more attention [14]. Multi-target drugs, compared with the use of combinations of different drugs, have more predictable pharmacokinetic and pharmacodynamic relationships as a consequence of the administration of a single drug [15].

Here, are summarized the results obtained in the development of a GA-based ensemble selection method and its application to the search of ensembles suitable for the VS identification of dual target ligands for PD. It should be mentioned that the application of LBVS tools for the planned design of ligands with a predefined dual-target profile is a recent development and represents an important challenge to medicinal chemists [16].

.

2. Results and Discussion

The collection, curation, class assignment, representation and dataset splitting processes were performed following the procedures described in the Materials and Methods section. Both modeling sets were split into training and test subsets using three sphere exclusion algorithms 1M, 2M and 3M. This partition scheme lead to the selection of 20-23 % of the modeling data for the test set and guaranties the representativeness of each class in the test set.

Fragment descriptors were calculated with ISIDA Fragmentor software [17,18] and they were filtered to discard those with zero or close to zero variance. Afterward, the mRMR algorithm [19] was employed to only keep the 250 more relevant descriptors. ISIDA's header files listing the 250 most relevant fragments for each dataset and their partitions are provided as Supporting Information.

The main goal of this paper is to find models that can be effective in the identification of dual target ligands for PD through VS. To this end, two sets of QSAR models were independently developed for $A_{2A}AR$ binding and MAO-B inhibition following the procedure summarized in the Materials and Methods section. For each endpoint, three feature selection methods and three different classification algorithms were combined to generate nine different classifier types, which were trained and validated using three modeling set partitions obtained with the sphere exclusion algorithms 1M, 2M and 3M. In consequence, 27 different classification experiments per endpoint were conducted. At this point, it is important to remark that the external set was only used to evaluate the real predictive power of the models. In Table 1 are presented the statistics of the optimal models for each partition of the dataset. These results show that the employed QSAR modeling framework provides accurate, robust and generalizable models.

All 16 known dual ligands, MAO-B inhibitors and $A_{2A}AR$ antagonists, were compiled from the literature [20,21]. For each known dual-target compound 1608 decoys were selected based on desirability-based home-developed algorithm that has been previously employed in the selection of tailored decoy sets for the validation of virtual screening strategies [9]. These ligands and decoys were prepared for virtual screening following the same protocol described for the datasets.

To evaluate the performance of the models in virtual screening the following metrics were employed: Area Under the Accumulation Curve (AUAC), Area under the Receiver Operating Characteristic Curve (ROC), Enrichment Factor (EF) and Boltzmann-Enhanced Discrimination of ROC (BEDROC). These metrics were computed as proposed by Truchon *et. al.*[22]

In our previous research [9] the VS-tailored ensembles were obtained using only a very limited set of predictive models that were selected as the best ones derived from each modeling approach. However, during the modeling process a larger number of high quality models are obtained that we did not considered for ensemble modeling in that investigation.

To discard low performing models, the whole pool of classification models was first filtered and a model was considered to be accurate, robust and predictive if it had values of accuracy higher than 0.75 in predicting the train, selection and external sets as well as in the cross validation experiments. The models fulfilling the above criteria are considered as valid models from here on.

For the three modeling set partitions obtained after applying the sphere exclusion algorithms, there were a total of 1526 and 1347 valid models for the A2A adenosine receptor antagonists and MAO-B inhibitors datasets respectively. Since one of the factors negatively influencing the performance of ensemble models is the redundancy of the base classifiers, these valid models were clustered following the protocol above to obtain a representative set of them. This procedure yielded 33 and 48 representative models for the A2A adenosine receptor antagonists and MAO-B inhibitors datasets respectively. The representative model of each cluster was selected as the one having the highest value of BEDROC for α=160.9. The best performance among the representative models corresponded to a value of BEDROC=0.20 for α=160.9.

We then evaluated the proposed GA-guided algorithm for the search of the ensemble maximizing BEDROC at a given fraction of screened data. The algorithm was run five times with different initial populations of 100 individuals each one, yielding 500 solutions. That is, 500 ensembles are obtained. The GA fitness function maximizes BEDROC for α=160.9.

In Table 2 are presented the enrichment metrics for the best individual found after running the GA-guided ensemble generation algorithm presented in this communication, when we search for the ensemble maximizing the initial enrichment of dual ligands. The accumulative curves for this ensemble as well as for its members are shown in Figure 1.

The analysis of these results show that the obtained ensemble is composed by 14 models, five of them related to the prediction of the antagonist activity of the A2A adenosine receptor and nine related to the inhibition of the MAO-B enzyme. This ensemble outperforms the VS metrics of all its members. Specifically, the improvement in the value of BEDROC for α=160.9 for the obtained ensemble relative to the best single model it is composed of is 0.05. The model with the highest value of BEDROC (α=160.9) among all single models within the ensemble is also the one with the best value of this metric among the set of representative valid models. This improvement might seem meaningless, however it can be interpreted as 5%

more probability of retrieving a dual target ligand in the first 1% of the ranked list using the obtained ensemble compared to the best performing individual model [22]. From Table 2 it can also be seen that the values of BEDROC for α=32.2 and α=20 are also higher when the ensemble model is compared with its members. In addition, improvements for the EF metric at the three analyzed fractions of screened data were obtained. The advantages of using the obtained ensemble for VS experiments over the use of single models can be visualized in Figure 1 where the accumulative curves of the ensemble and the models it is composed of are plotted. From this figure it is clear that the ensemble model is able to retrieve more known dual ligands and at lower positions in the ranked list than any of the single models it is composed of.

Another advantage of employing the proposed ensemble for VS tasks is that its applicability domain covers 93% of the whole virtual screening validation set, representing an improvement of 17% of coverage relative to the single model with the highest applicability domain coverage. If the applicability domain coverage of the ensemble is compared to that of the model with the highest BEDROC value for α=160.9, then this improvement increases to 33%. In contrast, the value of ROC of the ensemble is lower than the mean value of ROC across the ensemble members. This last result is a consequence of designing the GA search to find ensembles maximizing the initial enrichment of known dual ligands. In other words, the calculations here performed focus in retrieving the more dual ligands at the very first part of the ranked list and neglect the position of the ligands in the remaining of the list.

.

Table 1. Statistics for the best performing models for each dataset partition.

Target	Method[a]	Size[b]	Train [c]	Test[d]	LOO[e]	Boot[f]	5-Fold[g]	Ext [h]
A2A 1M	GA-AB	10	90 (93/86)	80 (90/72)	87.13	85.42	88.12	81 (81/81)
A2A 2M	GA-LSSVM	11	94 (90/97)	82 (77/88)	88.56	86.32	89.55	84 (84/84)
A2A 3M	GA-LSSVM	10	91 (87/93)	80 (78/83)	88.56	84.67	88.56	86 (84/88)
MAO 1M	GA-LDA	10	85 (90/81)	85 (83/86)	84.26	81.23	83.25	75 (79/67)
MAO 2M	BT-LSSVM	18	95 (93/97)	93 (92/94)	77.50	73.81	77.50	72 (63/88)
MAO 3M	GA-LSSVM	14	93 (94/92)	84 (79/89)	86.07	82.46	86.57	76 (78/73)

[a] Modeling method the classifier is based on, GA stands for Genetic Algorithm, BT for Bagged Trees, AB for Adaboost, LSSVM for Least Squares Support Vector Machines and LDA for Linear Discriminant Analysis.

[b] Number of features in the model.

[c,d,h] Accuracy in predicting the training, test and external sets respectively

[e,f,g] Accuracy of the Leave One Out, Bootstraping and 5-fold cross-validation experiments respectively

Table 2. Enrichment metrics for the ensemble maximizing the initial enrichment of known dual ligands.

Model	**BEDROC Alpha**			**EF %**			**ROC**	**Cov.**
	160.9	**32.2**	**20**	**1**	**5**	**8**		**Domain**
Ensemble	0.25	0.32	0.33	28.47	7.14	4.46	0.45	0.93
Mean Indiv. Models	0.04	0.08	0.10	4.39	2.27	1.42	0.44	0.69

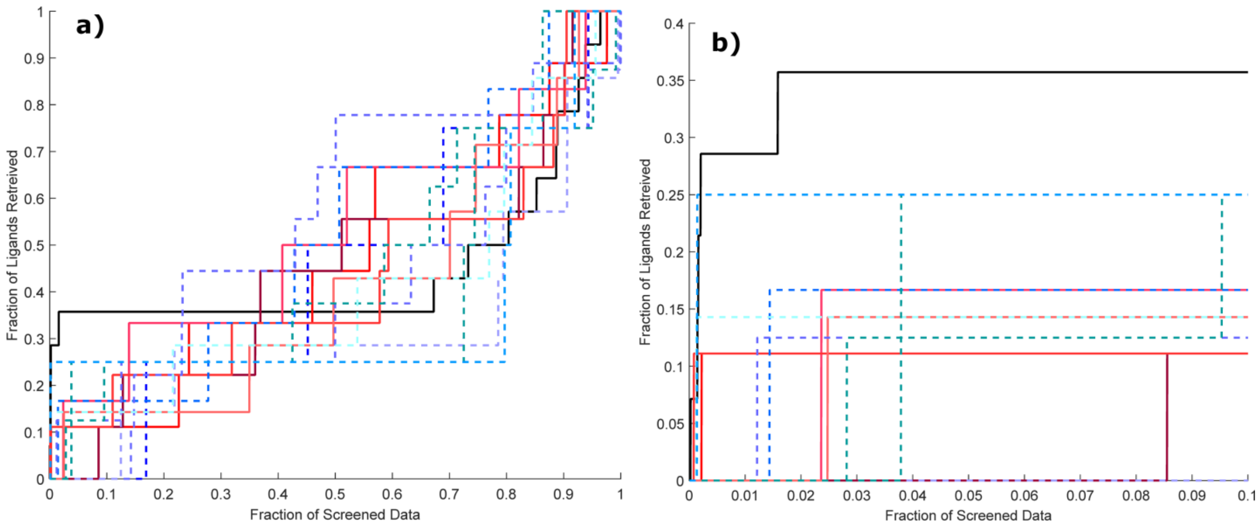


Figure 1. Accumulation curves for the best ensemble as well as for its base classifiers: a) For the whole virtual screening validation dataset and b) For the first 10% of screened data. The black line corresponds to the obtained ensemble, red continuous lines to A2A models and blue discontinuous lines to the MAO-B models

3. Materials and Methods

Data sets

Two data sets were used in this paper; A$_{2A}$AR antagonists and human MAO-B (*h*MAO-B) inhibitors. The A$_{2A}$AR antagonist data set was retrieved from 18 different literature sources. The compounds were divided into two classes according to their Ki values. The first class, designated as potent antagonists, included all chemicals with a Ki ≤ 1000 nM (pKi ≥ 6). The second class, named as weak antagonists, was formed by compounds with Ki >1000 nM (pKi < 6). As result of this categorization a balanced dataset was obtained that included 161 potent antagonists and 166 weak antagonists.

The *h*MAO-B inhibitors data set was compiled from [23] and it contains 474 compounds. the compounds were classified in two groups according with theirs IC$_{50}$ values. The first group, named *h*MAO-B inhibitors, included those chemicals with IC$_{50}$ ≤ 20 µM (*p*IC$_{50}$ ≥ 4.70), while the second one, designated as *h*MAO-B non-inhibitors comprises those compounds with IC$_{50}$ > 20 µM (*p*IC$_{50}$ < 4.70). Thus, the 474 ligands were split in 313 inhibitors and 161 non-inhibitors of *h*MAO-B.

The chemical structures were represented in smiles format and then converted to a SD file (SDF) using the ChemAxon's JChem for Excel (6.3.1.1807) program [24]. Each data set, as well as the decoy compound candidates were curated following the guidelines proposed in the literature [25] using ChemAxon's Standardizer [26]

Datasets splitting

The external sets were randomly selected as 20% of the entire initial datasets. The modeling sets were subsequently partitioned into training and test sets using the three sphere exclusion (SE) algorithms proposed by Golbraikh [27] and implemented in our laboratory that ensure the closeness in chemical spaces of the train and test sets.

The three variants 1M, 2M, and 3M of the sphere exclusion algorithm proposed by the developers and used here to divide the balanced modeling sets were implemented in MATLAB [28]. Unlike the original algorithms, for the SE based partitioning of the data the structure of the compounds was encoded as 1024 bits Chemaxon's Topological Fingerprints from GenerateMD program [26]; and the Tanimoto distance was selected as the distance metric. The radius of the spheres was varied between 0.05 and 1.0 with a step of 0.05

Molecular Descriptors

The ISIDA Fragmentor software (freely available at http://infochim.u-strasbg.fr/spip.php?rubrique49) was used to calculate 2D fragment descriptors [17,18].

Descriptors were calculated for the training dataset. Afterward, fragments with the same value for 99% of the samples were removed. The minimal Redundancy Maximal Relevance (mRMR) algorithm [19] was applied to the reduced data set to keep only the top 500 fragments according to the MIQ score. The same subset of 500 fragments was then computed for the selection and external sets.

Classification-based modeling methodology

QSAR modeling for each dataset was performed using the previously proposed QSAR modeling framework. Here the main steps involved in the modeling process are summarized and the detailed description of the QSAR modeling framework is available in [9]

Methodology of models ensemble for virtual screening

For both targets QSAR models were first filtered, regardless the dataset partition or modeling approach they were obtained with, to ensure that only accurate, robust and predictive models remain eligible as ensemble members. For this step we set the same cutoff value of 0.75 for the accuracy in predicting the training, selection and external datasets as well as in the cross-validation experiments. The models fulfilling these conditions were then clustered using the Hamming distance between the predictions they made on the external dataset to select a representative set of models to ensure diversity in the pool of base models. Clustering was carried out using the K-means algorithm implemented in MATLAB [28]. To obtain the optimal number of clusters we examined the silhouette plot [29] when the number of clusters varied between 3 and 50 and selected the number of clusters corresponding to the maximum of this plot. This procedure leads to the selection of N_{A2A} and N_{MAOB} representative models which form the final pool of diverse models candidates for ensembles.

To build the ensembles for VS we followed the same protocol described in our previous publication [9]. In brief, for a given subset of QSAR models forming the ensemble and a dataset to be evaluated we first search for the samples included within each model applicability domain. Next, the scores produced by each model for the compounds inside its applicability domain are used to obtain a ranking for them and the relative ranking of each sample in each model is computed. Once the relative rank of every sample in each model considering the model's applicability domain is determined, these relative rank values are averaged over the models the compound is inside their applicability domains to obtain the final aggregated score. Finally, the compounds are sorted according to this aggregated score in ascending order to obtain the final ensemble ranking.

Given that to evaluate the virtual screening performance of all ensembles formed by all possible combination of size two to N_{A2A}+N_{MAOB} of the selected models is computationally unfeasible, we implemented a novel GA search strategy which could find combinations of models maximizing the EF at a given fraction of screened data. Each individual for the GA search represents an ensemble and they are encoded as binary vectors of length $N_{A2A}+N_{MAOB}$. In an individual the "on" bits encode the set of models considered for the ensemble while "off" bits represent models excluded from the ensemble. The initial population was set to 100 randomly generated individuals and the population evolved for 100 generations. The crossover and mutation rates were set to 0.7 and 0.3 respectively while the best two individuals survived to the next generation. The selection operator was set to a tournament of size 2. For the crossover operator, the offspring chromosomes were randomly selected position by position from the two selected parents. The mutation operator changed a randomly selected "on" bit to "off" and one randomly selected "off" bit to "on" in the individual. In each case study the GA was run five times using different initial populations. The objective function for the GA was selected as BEDROC for α=160.

..

4. Conclusions

We designed a methodology for the generation of virtual screening tailored ensembles capable of overcoming the previously identified problems. This methodology considered the diversity of the base models for ensemble generation and their applicability domains. The main advantage of the proposed algorithm is that it is able of finding the combination of models providing the best VS performance for a specific problem.

The proposed algorithm was applied to the VS simulation of dual target A2A adenosine receptor antagonists and MAO-B inhibitors. The obtained results showed that the obtained ensemble outperformed the best individual model according to the evaluated enrichment metrics. Thus, confirming the expected improved performance of ensemble models over single ones. In the specific problem being addressed, the results of the ensemble modeling process highlighted the importance of considering information from both targets for the discovery of dual target ligands.

Acknowledgments

MC-M acknowledges the postdoctoral grant [SFRH/BPD/90673/2012] financed by the FCT – Fundação para a Ciência e a Tecnologia, Portugal, co-financed by the European Social Fund.

Conflicts of Interest

The authors declare no conflict of interests.

References and Notes

1. Macalino, S.J.; Gosu, V.; Hong, S.; Choi, S., Role of computer-aided drug design in modern drug discovery. *Archives of pharmacal research* **2015**, *38*, 1686-1701.
2. Zhu, T.; Cao, S.; Su, P.C.; Patel, R.; Shah, D.; Chokshi, H.B.; Szukala, R.; Johnson, M.E.; Hevener, K.E., Hit identification and optimization in virtual screening: Practical recommendations based on a critical literature analysis. *J Med Chem* **2013**, *56*, 6560-6572.
3. Castillo-Gonzalez, D.; Mergny, J.L.; De Rache, A.; Perez-Machado, G.; Cabrera-Perez, M.A.; Nicolotti, O.; Introcaso, A.; Mangiatordi, G.F.; Guedin, A.; Bourdoncle, A., *et al.*, Harmonization of qsar best practices and molecular docking provides an efficient virtual screening tool for discovering new g-quadruplex ligands. *Journal of Chemical Information and Modeling* **2015**, *55*, 2094-2110.

4. Miller, Z.; Kim, K.S.; Lee, D.M.; Kasam, V.; Baek, S.E.; Lee, K.H.; Zhang, Y.Y.; Ao, L.; Carmony, K.; Lee, N.R.*, et al.*, Proteasome inhibitors with pyrazole scaffolds from structure-based virtual screening. *Journal of Medicinal Chemistry* **2015**, *58*, 2036-2041.
5. Lill, M., Virtual screening in drug design. *Methods in molecular biology (Clifton, N.J.)* **2013**, *993*, 1-12.
6. Lavecchia, A.; Di Giovanni, C., Virtual screening strategies in drug discovery: A critical review. *Current medicinal chemistry* **2013**, *20*, 2839-2860.
7. Cherkasov, A.; Muratov, E.N.; Fourches, D.; Varnek, A.; Baskin, I.I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y.C.; Todeschini, R., Qsar modeling: Where have you been? Where are you going to? *Journal of medicinal chemistry* **2014**.
8. Tropsha, A., Best practices for qsar model development, validation, and exploitation. *Mol Inf* **2010**, *29*, 476-488.
9. Perez-Castillo, Y.; Cruz-Monteagudo, M.; Lazar, C.; Taminau, J.; Froeyen, M.; Cabrera-Perez, M.A.; Nowe, A., Toward the computer-aided discovery of fabh inhibitors. Do predictive qsar models ensure high quality virtual screening performance? *Mol Divers* **2014**, *18*, 637-654.
10. Bonet, I.; Franco-Montero, P.; Rivero, V.; Teijeira, M.; Borges, F.; Uriarte, E.; Morales Helguera, A., Classifier ensemble based on feature selection and diversity measures for predicting the affinity of a(2b) adenosine receptor antagonists. *Journal of Chemical Information and Modeling* **2013**, *53*, 3140-3155.
11. Zhang, L.; Fourches, D.; Sedykh, A.; Zhu, H.; Golbraikh, A.; Ekins, S.; Clark, J.; Connelly, M.C.; Sigal, M.; Hodges, D.*, et al.*, Discovery of novel antimalarial compounds enabled by qsar-based virtual screening. *Journal of Chemical Information and Modeling* **2013**, *53*, 475-492.
12. Polikar, R., Ensemble based systems in decision making. *Circuits and Systems Magazine, IEEE* **2006**, *6*, 21-45.
13. Brunton, L.L., *Goodman & gilman's the pharmacological basis of therapeutics*. 11th ed.; The McGraw-Hill: New York, NY, 2007.
14. Youdim, M.B.; Geldenhuys, W.J.; Van der Schyf, C.J., Why should we use multifunctional neuroprotective and neurorestorative drugs for parkinson's disease? *Parkinsonism Rel Disord* **2007**, *13*, S281-S291.
15. Morphy, R.; Kay, C.; Rankovic, Z., From magic bullets to designed multiple ligands. *Drug Discovery Today* **2004**, *9*, 641-651.
16. Wang, Y.; Ge, H.; Li, Y.; Xie, Y.; He, Y.; Xu, M.; Gu, Q.; Xu, J., Predicting dual-targeting anti-influenza agents using multi-models. *Molecular diversity* **2014**, *19*, 123-134.
17. Varnek, A.; Fourches, D.; Hoonakker, F.; Solov'ev, V.P., Substructural fragments: An universal language to encode reactions, molecular and supramolecular structures. *Journal of computer-aided molecular design* **2005**, *19*, 693-703.
18. Varnek, A., Isida-platform for virtual screening based on fragment and pharmacophoric descriptors. *Current Computer Aided-Drug Design* **2008**, *4*.
19. Peng, H.; Long, F.; Ding, C., Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **2005**, *27*, 1226-1238.

20. Stossel, A.; Schlenk, M.; Hinz, S.; Kuppers, P.; Heer, J.; Gutschow, M.; Muller, C.E., Dual targeting of adenosine a(2a) receptors and monoamine oxidase b by 4h-3,1-benzothiazin-4-ones. *J Med Chem* **2013**, *56*, 4580-4596.
21. S., R.; Piersanti, G.; Bartoccini, F.; Diamantini, G.; Pala, D.; Riccioni, T.; Stasi, M.A.; Cabri, W.; Borsini, F.; Mor, M.*, et al.*, Synthesis of (e)‑8-(3-chlorostyryl)caffeine analogues leading to 9-deazaxanthine derivatives as dual a2a antagonists/mao-b inhibitors. *J Med Chem* **2013**, *56*, 1247−1261.
22. Truchon, J.F.; Bayly, C.I., Evaluating virtual screening methods: Good and bad metrics for the "early recognition" problem. *Journal of Chemical Information and Modeling* **2007**, *47*, 488-508.
23. Helguera, A.M.; Pérez-Garrido, A.; Gaspar, A.; Reis, J.; Cagide, F.; Vina, D.; Cordeiro, M.N.D.S.; Borges, F., Combining qsar classification models for predictive modeling of human monoamine oxidase inhibitors. *European Journal of Medicinal Chemistry* **2013**, *59*, 75-90.
24. ChemAxon (http://www.chemaxon.com) *Jchem for excel, version 6.3.1.1807*, Budapest, Hungary, 2013.
25. Fourches, D.; Muratov, E.; Tropsha, A., Trust, but verify: On the importance of chemical structure curation in cheminformatics and qsar modeling research. *Journal of Chemical Information and Modeling* **2010**, *50*, 1189-1204.
26. ChemAxon (http://www.chemaxon.com) *Jchem, version 6.3.1*, Budapest, Hungary, 2013.
27. Golbraikh, A.; Tropsha, A., Predictive qsar modeling based on diversity sampling of experimental datasets for the training and test set selection. *Molecular diversity* **2000**, *5*, 231-243.
28. MATLAB *Version 8.1.0.604 (r2013a)*, The MathWorks Inc.: Natick, Massachusetts, 2009.
29. Rousseeuw, P.J., Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **1987**, *20*, 53-65.

SciForum
Mol2Net

Development of QSAR Models for Identification of CYP3A4 Substrates and Inhibitors

Flavia C. Silva, Ekaterina V. Varlamova, Rodolpho C. Braga, and Carolina H. Andrade*

Labmol – Laboratory for Molecular Modeling and Drug Design, Faculty of Pharmacy, Federal University of Goias, Goiania, Goiás, 74605-170, Brazil.

* Author to whom correspondence should be addressed; E-Mail: carolina@ufg.br. Tel: + 55 62 3209-6451; Fax: + 55 62 3209-6037.

Published: 4 December 2015

Abstract: The pharmacokinetic properties of absorption, distribution, metabolism and excretion (ADME) play a crucial role in drug discovery and development, since many drug candidates fail due to an inappropriate pharmacokinetic profile. Cytochrome P450 enzymes are predominantly involved in Phase 1 metabolism of xenobiotics. Thus, it is important to better understand and prognosticate substrate binding and inhibition of CYP450. The goal of this study was to obtain QSAR (Quantitative Structure-Activity Relationship) models to identify substrates and inhibitors of CYP3A4. The data sets were collected and curated from online available databases and literature. Several QSAR models were obtained and validated according to the recommendations of the Organization for Economic Co-operation Development (OECD). The combination of different descriptors and machine learning methods led to robust and predictive QSAR models with high coverage. The interpretation of developed models was performed using the predicted probability maps (PPMs). These maps help to encode major structural fragments to classify compounds as inhibitors or not inhibitors of CYP3A4. In conclusion, the obtained models can reliably identify substrates and non-substrates, and inhibitors and non-inhibitors of CYP3A4, which is very important in the early stages of the development of new drugs.

Keywords:QSAR; *in silico*; drug metabolism; substrate; inhibitor; CYP3A4.

1. Introduction

Many drug candidates fail during the drug development process in clinical trials due to an inappropriate pharmacokinetic profile. For this reason, the study of the pharmacokinetic properties absorption, distribution, metabolism, excretion, and toxicity (ADME/Tox) of a drug candidate is important to reduce time and increase the chances of success during drug discovery and development[1].

ADME/Tox properties are the major contributors to the failures of new drugs in the development pipeline and often the underlying biological mechanism of toxicity is related to metabolism. Metabolic liability can lead to a number of diverse issues, including drug−drug interactions, in particular enzyme inhibition and induction, which in turn may cause therapeutic failure toxicity, and adverse effects[2].

Cytochrome P450 (CYP) enzymes are predominantly involved in Phase 1 metabolism of xenobiotics. CYP3A4 is the most abundant cytochrome isoenzyme present in liver and is responsible for the metabolism of more than 50% of the marketed drugs[3]. The main goal of this study was to develop robust and predictive models that can be used to classify compound as inhibitor/non-inhibitor or substrate/non-substrate of CYP3A4 for identifying and discarding drug candidates with potential metabolism issues.

2. Results and Discussion

The statistical results of QSAR models generated for substrates of CYP3A4 (dataset I), using the test set compounds, are summarized in Figure 1.

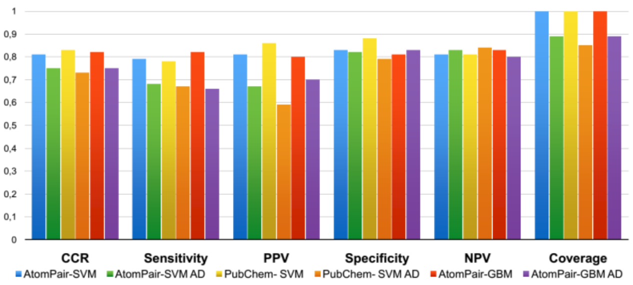


Figure 1. Statistical results of predictions of QSAR models for CYP3A4 substrates evaluated by 5-fold external cross-validation.

The combination of different descriptors and machine learning (ML) methods led to robust and predictive QSAR models for substrates of CYP3A4, with correct classification rate (CCR) values ranging between 0.65-0.83 and coverage of 0.69-0.89. However, among the best three selected models (Atom Pair-SVM; PubChem-SVM; Atom Pair-GBM), the model generated by combining Atom Pair-GBM without considering DA showed a higher sensitivity and lower difference between the values of sensitivity and specificity obtained the best ability to classify correctly both substrates as non-substrates of CYP3A4.

The statistical results of binary and multiclass QSAR models for CYP3A4 inhibitors (data set II) are illustrated on Figure 2.

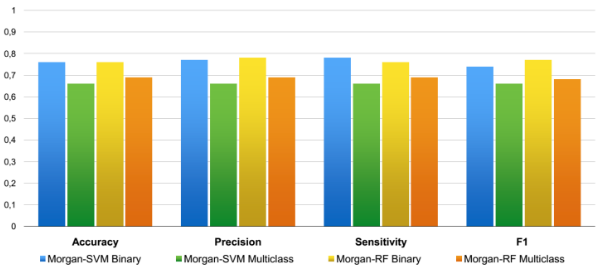


Figure 2. Statistical results of predictions for the best binary and multiclass QSAR models for CYP3A4 inhibitors evaluated by 5-fold external cross-validation.

The two best binary and multiclass models were generated using a combination of Morgan-SVM and Morgan-RF. These binary models showed equal values of accuracy 0.76, which corresponds to the percentage of molecules that are correctly classified by model. Furthermore, they showed sensitivity values of 0.74 and 0.77, respectively. The accuracy of these models was 0.77 and 0.78, respectively, whereas F1 was 0.76 and for both models. The multiclass models were also generated using the combination of Morgan-SVM and Morgan-RF. The Morgan-RF model presented precision value 0.69, while the Morgan-SVM was 0.66. The Morgan-RF model was also slightly higher in relation to F1 value, with value of 0.69, compared to the value of 0.66 for the Morgan-SVM. However, multiclass and binary QSAR models showed similar statistical results. Therefore, both models were considered the best models to evaluate the inhibition of CYP3A4. In addition, predicted probability maps (PPMs) were generated by Morgan-RF models. The maps for drugs ketoconazole, tioconazole and miconazole are presented in Figure 3.

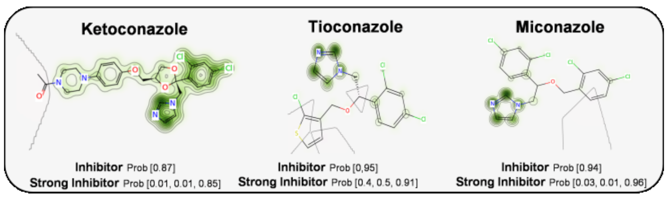

Ketoconazole
Tioconazole
Miconazole
Inhibitor Prob [0.87]
Strong Inhibitor Prob [0.01, 0.01, 0.85]
Inhibitor Prob [0,95]
Strong Inhibitor Prob [0.4, 0.5, 0.91]
Inhibitor Prob [0.94]
Strong Inhibitor Prob [0.03, 0.01, 0.96]

Figure 3. PPMs for selected antifungal drugs generated using Morgan-RF models. Green atoms/fragments have favorable contribution in the property (CYP3A4 inhibition); Gray: no contribution; Pink atoms/fragments have unfavorable contribution in the property (CYP3A4 non-inhibition). The bit vector size of Morgan was 1024 bits.

Miconazole, ketoconazole and tioconazole are antifungal drugs and CYP3A4 inhibitors. These three drugs were classified by the binary model as CYP3A4 inhibitors, and multiclass model considered the three drugs as strong inhibitors with high probability. The imidazole fragment in their structures outlined in green indicate that this fragment has favorable characteristics for the investigated property. These fragments have atoms which are capable of coordinating with heme group iron. The phenyl and thiophene rings are outlined in gray color, which features neutral contribution to the property. Gray isolines demarcate the separation of regions that have favorable and unfavorable contribution.

3. Materials and Methods

In this study, two large datasets were collected for profiling the CYP3A4 activity. The dataset I contained 8,214 compounds, in which 475 are substrates of CYP3A4 and 7,739 are non-substrates (inactive). The annotated dataset was gathered from the literature[4] and PubChem bioassay (Assay ID: 1851). The dataset II contained 9,186 compounds, in which 4,962 are inhibitors de CYP3A4 and 4,224 are non-inhibitors. The annotated dataset was gathered from ChEMBL340 assay. All the molecular modeling studies were performed using a workflow in KNIME platform developed in our laboratory. The dataset curation (removal of duplicates, structural conversion, normalization of specific chemotypes etc.) was performed using Indigo Open Source Standardizer following the workflow described by Fourches et al.[5] including the duplicate analysis. Binary and multiclass QSAR models were developed and validated according to the OECD principles. For generation of QSAR models we used the qsaR package fully integrated workflow KNIME 2.9[6]. The cross-validation procedure 5-fold was used to estimate the robustness of the model using the training set, while the test set was used to validate and estimate the predictive power of the generated models.

Because dataset I was highly unbalanced, it was not recommended to build binary QSAR models for the entire dataset. Therefore, a linear under-sampling strategy was used to investigate the more adequate dataset balancing. We generated five under-sampled datasets with substrates-to-non-substrates ratios of 1:1, 1:2, 1:3, 1:4, 1:8, and the unbalanced dataset. From the six different datasets splits generated, the balancing with proportion of 1:1 and the total unbalanced dataset were selected because of the best statistical results and covering the largest chemical space. Thus, various QSAR models were generated using different types of descriptors and algorithms, in order to use more information from QSAR models. Four different types of molecular fingerprints were utilized in this study (Atom Pair[7], PubChem[8], MACCS[9] and FeatMorgan[10]), as well as four ML algorithms (SVM[11], GBM[12], PLSDA[13] and *k*NN[14]) were used to model generation, totaling in 16 different QSAR models.

For dataset II, the models for CYP3A4 inhibitors were generated using a 5-fold technique, *i.e.*, spliting the data set in modeling set and external validation set. We used only one type of molecular descriptor (Morgan) and two ML methods (SVM and RF[15]). For construction of multiclass models, the threshold activity was defined as follows: strong inhibitor ≤ 1 μM; weak-moderate inhibitor, property between 1 μM and 10 μM; non-inhibitor ≥ 10 μM[16].

PPMs[17] were generated for visualization of favorable (positive) and unfavorable (negative) structural fragments for compound to be inhibitor or non-inhibitor of CYP3A4.

4. Conclusions

The largest publicly available data sets for substrates and inhibitors of CYP3A4 were collected, prepared and balanced. Robust and predictive QSAR models were generated for the identification of substrates (binary models) and inhibitors (binary and multiclass models). Obtained models can be used for identifying substrates and inhibitors of CYP3A4 in early stages of drug development. PPMs showed important contribution of some fragments probably responsible for interaction with the heme group of CYP3A4.

Acknowledgments

We are grateful to Coordination for the Improvement of Higher Education Personnel (CAPES), the National Counsel of Technological and Scientific Development (CNPq), and the State of Goiás Research Foundation (FAPEG) for their financial support and fellowships. We are also grateful to ChemAxon (Budapest, Hungary) for providing us with the academic license for their software.

Author Contributions

Conceived and designed the experiments: FCS, EV, RCB, CHA. Performed the experiments: FCS, EV, RCB. Analyzed the data: FCS, EV, RCB, CHA. Contributed analysis tools: FCS, EV, RCB, CHA. Wrote the paper: FCS, EV, RCB, CHA.

Conflicts of Interest

The authors declare no conflict of interest. The funders had no role in the study design, data collection, analysis, decision to publish, or preparation of this manuscript.

References and Notes

1. STEPAN, A. F.; MASCITTI, V.; BEAUMONT, K.; KALGUTKAR, A. S. Metabolism-guided drug design. **MedChemComm**, 2013, v, 4, p. 631-652.

2. LI, H.; SUN, J.; FAN, X.; SUI, X.; ZHANG, L.; WANG, Y.; HE, Z. Considerations and recent advances in QSAR models for cytochrome P450-mediated drug metabolism prediction. **Journal of computer-aided molecular design**, 2008, 11, p. 843–855.
3. KIRCHMAIR, J.; WILLIAMSON, M. J.; TYZACK, J. D.; TAN, L.; BOND, P. J.; BENDER, A.; GLEN, R. C. Computational prediction of metabolism: sites, products, SAR, P450 enzyme dynamics, and mechanisms. **Journal of chemical information and modeling**, 2012, 3, p. 617–648.
4. ZARETZKI, J.; RYDBERG, P.; BERGERON, C.; BENNETT, K. P.; OLSEN, L.; BRENEMAN, C. M. RS-Predictor models augmented with SMARTCy reactivities: robust metabolic regioselectivity predictions for nine CYP isozymes. **Journal of chemical information and modeling**, 2012, 6, p. 1637–1659.
5. FOURCHES, D.; MURATOV, E.; TROPSHA, A. Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. **Journal of chemical information and modeling**, 2010, 7, p. 1189–1204.
6. BRAGA, R. C.; ALVES, V. M.; SILVA, A. C.; LIAO, L. M.; ANDRADE, C. H. Virtual Screening Strategies in Medicinal Chemistry: The state of the art and current challenges. **Current topics in medicinal chemistry**, 2014, 16, p. 1899–1912.
7. CARHART, R. E.; SMITH, D. H.; VENKATARAGHAVAN, R. Atom Pairs as Molecular Features in Structure-Activity Studies : Definition and Applications. **Journal of chemistry information and computer sciences**, 1985, 2, p. 64-73.
8. STEINBECK, C.; HAN, Y.; KUHN, S.; HORLACHER, O.; LUTTMANN, E.; WILLIGHAGEN, E. The Chemistry Development Kit (CDK): an open-source Java library for Chemo- and Bioinformatics. **Journal of chemical information and computer sciences**, 2003, 2, p. 493–500.
9. TODESCHINI, R., CONSONNI, V. **Molecular Descriptors for Chemoinformatics**. 2 rd ed. MANNHOLD, R; KUBINYI, H; FOLKERS, G. Wiley-VCH: Weinheim, Germany, 2009, p-1-1257.
10. MORGAN, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. **Journal of Chemical Documentation**, 1965, 2, p. 107–113.
11. CORTES, C.; VAPNIK, V. Support-vector networks. **Machine Learning**, 1995, 20 (3), p. 273-297.
12. FRIEDMAN, J. H. Greedy Function Approximation: A Gradient Boosting Machine. **Annals of Statistics,** 2001, 5, p. 1189–1232.
13. Barker, M.; Rayens, W. Partial Least Squares for Discrimination. **Journal Chemometrics**.**2003**, *3*, p.166–173.
14. ALTMAN, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. **The American Statistician,** 1992, 46 (3), p. 175–185.
15. BREIMAN, L. Random Forests. **Machine Learning**, 2001, 45 (1), p. 5–32
16. YAN, Z.; CALDWELL, G. W. Metabolism profiling, and cytochrome P450 inhibition & induction in drug discovery. **Current topics in medicinal chemistry**, 2001, 5, p. 403–425.

17. RINIKER, S.; LANDRUM, G. A. Similarity maps - a visualization strategy for molecular fingerprints and machine-learning methods. **Journal of cheminformatics**, 2013, 1, p. 43.

SciForum
Mol2Net

Uptake of Different Organic Pollutants by Carrot

Ekhiñe Bizkarguenaga [1],*, Luis Ángel Fernández [1], Olatz Zuloaga [1] and Ailette Prieto [1]

[1] Department of Analytical Chemistry, Faculty of Science and Technology, University of the Basque Country (UPV/EHU), P.O. Box 644, 48080 Bilbao (Spain), Address; E-Mail: luis-angel.fernandez@ehu.eus (Luis Angel Fenández), olatz.zuloaga@ehu.eus (Olatz Zuloaga), ailette.prieto@ehu.eus (Ailette Prieto)

* Author to whom correspondence should be addressed; ebizkarguenaga@gmail.com.

Published: 4 December 2015

Abstract: In this study the uptake of different organic pollutants, including musk fragrances (tonalide and galaxolide), polybrominated diphenyl ethers (PBDEs), perfluorocarboxylic acids (PFCAs), perfluorosulfonic acids (PFSAs) and perfluorosulfonamide (FOSA) by carrot samples was compared. The bioconcentration factors (BCFs), defined as ratio of the concentration in the dry plant tissue and the concentration in the compost amended soils, were compared and correlation with the water solubility of the target compounds was studied. A good correlation was obtained between the water solubility and the BCFs in the different plant tissues (carrot root peel, root core and leaves). Besides, while the target analytes with the lowest solubility (musk fragrances and PBDEs) tended to accumulate in the peel of the carrot, the most water soluble target analytes (the perfluorinated compounds) tended to translocate to the carrot leaves.

.

.

Keywords: plant uptake; carrot; musk fragrances; polybrominated diphenyl ethers, perfluorocarboxylic acids; perfluorosulfonic acids; perfluorosulfonamide;

1. Introduction

Under the Urban Wastewater Treatment Directive (UWWTD), towns and cities within the 28 European Union (EU-28) members are required to collect and treat their urban wastewater. The reuse of the sludge is also encouraged, and final disposal to surface waters has been banned (1). However, wastewater treatment plants (WWTPs), also called “biological treatments”, are demonstrated not to be effective enough in contaminant removal (2). Not all the chemicals entering the WWPTs are completely degraded and are either removed by sorption and deposition to the final sludge, by volatilization or by discharge onto a surface water body, if they remain in the wastewater effluent stream (2).

In this sense, contaminants of emerging concern (CECs) have been detected in effluents discharges from municipal and/or industrial wastewater treatment plants (WWTPs), including polybrominated diphenyl ethers (PBDEs), musk fragrances or perfluorinated compounds (2, 3).

Land applications of sewage sludge and/or compost derived of them have been adopted worldwide as an option for sludge management. Crops grown in soils amended or irrigated with wastewater containing CECs are exposed to contaminant uptake (4), which then become and entrance of pollutants in the food chain. Within this context, and taking into account that plants form an essential basis of animal and human diet, an evaluation of the uptake and accumulation of potential harmful organic contaminants in plants is of importance for risk assessment.

In this sense, the aim of the present work is to study the uptake in terms of BCFs of different organic pollutants, including musk fragrances (tonalide and galaxolide), PBDEs, perfluorocarboxylic acids (PFCAs), perfluorosulfonic acids (PFSAs) and perfluorooctanosulfonamide (FOSA) by different carrot compartments (peel, core and leaves) and to evaluate the correlation between BCFs and the target analytes water solubility.

.

2. Results and Discussion

The BCFs for tonalide (water solubility 1200 µg/L), galaxolide (water solubility 1800 µg/L), BDE-138 (water solubility 19 µg/L) and BDE-209 (water solubility 0.14 µg/L), perfluorooctanoic acid (PFOA, water solubility $11\cdot10^{6}$ µg/L), perfluorooctasulfonate (PFOS, water solubility $75\cdot10^{5}$ µg/L), perfluorootanosulfonamide (FOSA, water solubility 0.029 µg/L), perfluoro-n-nonanoic acid (PFNA, water solubility 2.5 10^{6} µg/L), perfluoro-n-heptanoic acid (PFHpA, water solubility 5.1 10^{7} µg/L), perfluorohexyl phosphonic acid (PFHxPA, water solubility 2.3 10^{8} µg/L), perfluoro-n-pentanoic acid (PFPeA, water solubility 9.5 10^{8} µg/L) and perfluoro-n-butanoic acid (PFBA, water solubility 10 10^{8} µg/L) obtained in our laboratories for the same carrot (*Daucus carota ssp sativus*) specie (Chantenay) were correlated with the logarithm of their water solubilities. According to the results in Figure 1 (a), an exponential correlation between water solubility and $BCF_{Total,}$ as well as BCF_{Peel} (determination coefficients of r^2=0.32), BCF_{Core} (r^2=0.74) and BCF_{Leaves} (r^2=0.60) was observed for all the analytes. The accumulation was higher with the water solubility increment observing a dramatically bioconcentration increased for the analytes (PFHxPA, PFPeA and PFBA) with a water solubility higher than 8. In order to confirm these results, the previously mentioned analytes were discarded and only the rest included in the graphic (see Figure 1 (b)) observing the similar behaviour for all the analytes.

It should also be highlighted, while the target analytes with the lowest solubility (musk fragrances, AHTN and HHCB, and PBDEs, BDE-138 and BDE-209) tended, in general, to accumulate in the peel of the carrot, the most water soluble target analytes (PFOS and PFOA as examples) tended to translocate to the carrot leaves (see BCFs included in Table 1). However, as can be clearly observed from Table 1, while BDE-138 accumulated exclusively in the peel, BDE-209 accumulated mainly in the leaves (BCF_{leaves}) when present at a low concentration.

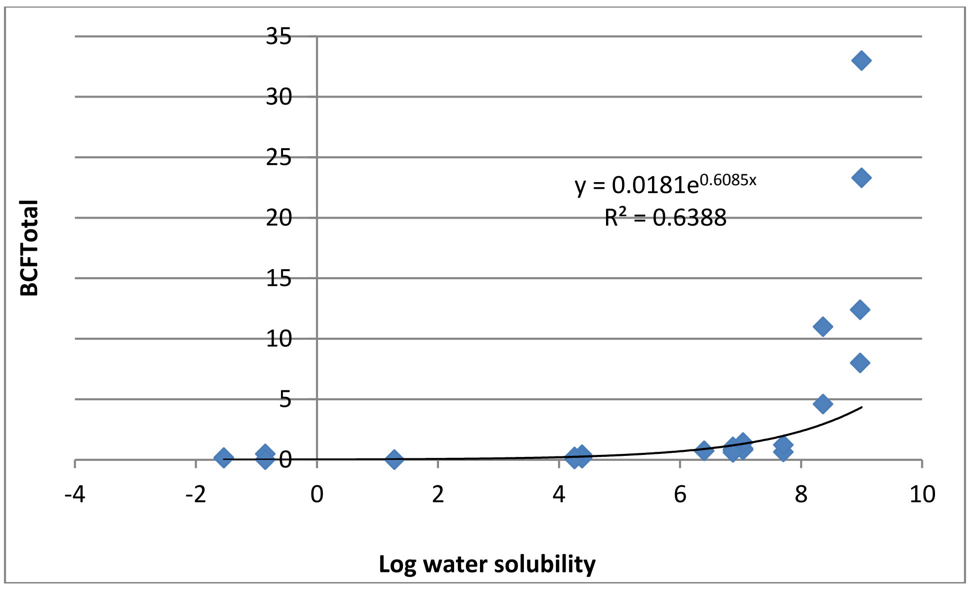
Accumulation in the leaves could be due to translocation after root uptake or by foliar uptake from the air. According to the values obtained from the blanks, no appreciable contribution from foliar uptake was observed for any of the target analytes and, thus, foliar uptake was discarded.

Main text paragraph.

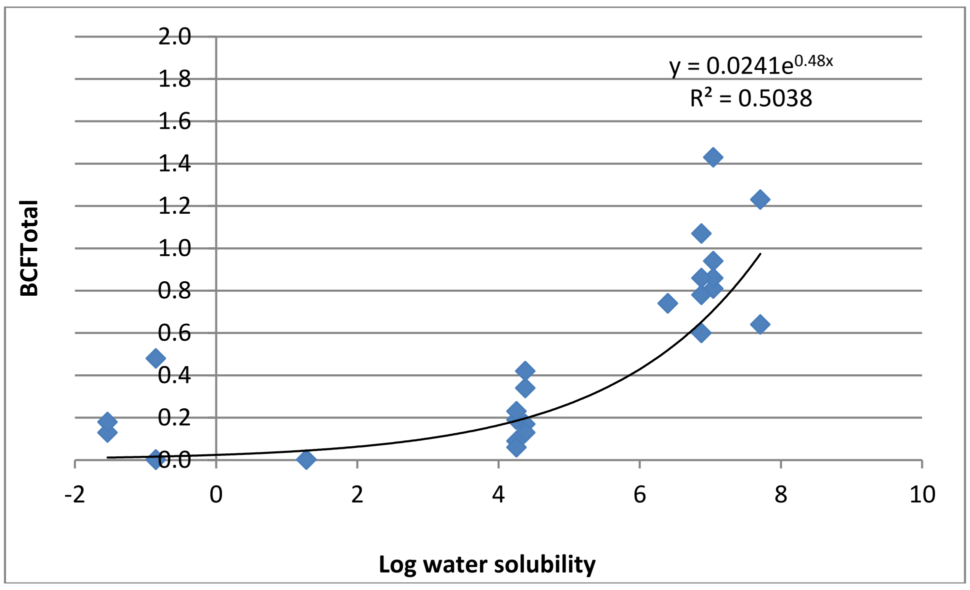
Table 1. Bioconcentration factors (BCFs) of the analytes in the different carrot (*Chantenay* specie) compartments and the total BCFs (BCF$_{Total}$).

An average nominal concentration of 5000 ng/g (musks fragrances) and 7500 ng/g (BDE-209) (high concentration level) was used as fortification level. As the compost already contained the musks (16-38 ng/g) and BDE-209 (7-20 ng/g) at a low concentration level, fortification was unnecessary. BDE-138 was added in order to adjust to a nominal concentration of 120 ng/g (low level). A nominal concentration of 500 ng/g (low level) was used in the case of PFOA and PFOS. MDLs: method detection limits.

Experiment	Compartment	BCF$_{HHCB}$	BCF$_{AHTN}$	BCF$_{BDE-138}$	BCF$_{BDE-209}$	BCF$_{PFOA}$	BCF$_{PFOS}$
Low level	Peel	1.25-1.54	0.59-0.67	0.01-0.02	< MDLs	0.12-0.61	0.39-0.43
	Leaves	0.44-0.52	0.30-0.32	< MDLs	1.44	0.81-3.34	1.67-1.93
	Core	<MDLs	<MDLs	< MDLs	< MDLs	0.05-0.36	0.52-0.64
	Total	0.34-0.42	0.19-0.23	0.002-0.003	0.48	0.81-0.94	0.86-1.07
High level	Peel	0.84-0.87	0.42-0.46		0.001-0.009		
	Leaves	0.02-0.02	0.01-0.01		0.003-0.004-		
	Core	0.03-0.04	0.01-0.02		< MDLs		
	Total	0.13-0.17	0.06-0.09		0.001-0.003		



(a)



(b)

Figure 1. BCF_{Total} versus the logarithm of water solubility of (a) BDE-138, BDE-209, PFOA, PFOS, PFOSA, PFNA, PFHpA, PFHxPA, PFPeA and PFBA and (b) all the analytes except PFHxPA, PFPeA and PFBA.

3. Materials and Methods

Plant Cultivation

Pots (n=2) with 2 kg of the (95:5) soil:compost mixtures were sown with previously germinated (~14 days) carrot seeds. For germination, petri dishes were covered with moistened filter paper and the seeds were evenly distributed in the petri dish. Afterwards, seeds were covered with another piece of moistened filter paper. The number of plants per pot was 3−4.

Control (n=1) plants of carrots grown in the non-fortified compost amended soil 2.4 mixture were placed in between the fortified amended soil pots. The cultivation of the carrot was performed under controlled greenhouse conditions.

Temperature was set to 25 °C during the day and at 18 °C during the night with a 14-h day length and a relative humidity of 50 % and 60 % during the day and overnight, respectively, and they were regularly watered with distilled water and Hoagland nutritive solution (5). Carrots were harvested during a period of three months reflecting the minimum time to produce relatively mature crops and all plants per pot were collected and pooled to one sample. Each plant was dissected into roots (peel and core) and leaves. Fresh weight of all plant fractions was recorded, followed by rinsing with tap water. Carrots were peeled with a vegetable peeler (depth of ~2 mm).

Sample treatment and analysis

Carrot samples (peel, core and leaves) were freeze-dried using a Cryodos-50 laboratory freeze-dryer (Telstar Instrumat, Sant Cugat del Valles, Barcelona, Spain). In the case of the compost amended soil, this was air-dried for approx. 48 hours. Both, the dried plant and compost amended soil samples were stored at -20

°C until analysis. Analyses were performed in triplicate under the conditions described in previous works for musk fragrances (6), PBDEs (7) and perfluorinated compounds (8).

4. Conclusions

It has been demonstrated that uptake of several organic pollutants is dependent on their water solubility and that the BCFs in the different carrot compartments is exponentially related to the logarithm of the water solubility of the target analytes. A dramatically bioconcentration increase for the analytes (PFHxPA, PFPeA and PFBA) with a water solubility higher than 8 was observed. In general, while the target analytes with the lowest solubility tended to accumulate in the carrot peel, the most water soluble target analytes tended to translocate to the leaves.

Acknowledgments

The Spanish Ministry of Science and Technology financially supported this work through the CTM2011-24094 project. E. Bizkarguenaga is for her pre-doctoral fellowships Technical and human support by PhD. Azucena González, Phytotron Service, SGIker (UPV/EHU) is gratefully acknowledged.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Kelessidis, A.; Stasinakis, A.S. Comparative study of the methods used for treatment and final disposal of sewage sludge in European countries. *Waste Manage.* **2012**, 32, 1186-1195.
2. Ratola, N.; Cincinelli, A.; Alves, A.; Katsoyiannis, A. Occurrence of organic microcontaminants in the wastewater treatment process: A mini review. *J. Hazard. Mater.* **2012,** 239, 1-18.
3. Loos, R.; Carvalho, R.; António, D. C.; Comero, S.; Locoro, G.; Tavazzi, S.; Paracchini, B.; Ghiani, M.; Lettieri, T.; Blaha, L.; Jarosova, B.; Voorspoels, S.; Servaes, K.; Haglund, P.; Fick, J.; Lindberg, R. H.; Schwesig, D.; Gawlik, B. M., EU-wide monitoring survey on emerging polar organic contaminants in wastewater treatment plant effluents. *Water Res.* **2013,** 47, 6475-6487.
4. Olofsson, U.; Bignert, A.; Haglund, P. Time-trends of metals and organic contaminants in sewage sludge. *Water Res*. **2012**, 46, 4841-4851.
5. Epstein, E.; Bloom, A.J. Mineral nutrition of plants: principles and perspectives. *second ed. Sinauer Associates, Sunderland, MA* **2005**.
6. Aguirre, J.; Bizkarguenaga, E.; Iparraguirre, A.; Fernández, L.Á.; Zuloaga, O.; Prieto, A. Development of stir-bar sorptive extraction thermal desorption gas chromatography-mass spectrometry for the analysis of musks in vegetables and amended soils. *Anal. Chim. Acta* **2014**, 812, 74-82.
7. Iparragirre, A.; Rodil, R.; Quintana, J.B.; Bizkarguenaga, E.; Prieto, A.; Zuloaga, O.; Cela, R.; Fernández, L.A., Matrix solid-phase dispersion for the extraction of polybrominated diphenyl ethers and their hydroxylated and methoxylated analogues in vegetables and soils. *J. Chromatogr. A* **2014**, 1360, 57-65.

8. Zabaleta I.; Bizkarguenaga E.; Iparraguirre A.; Navarro P.; Prieto A.; Fernández L.A.; Zuloaga O.; Focused ultrasound solid-liquid extraction for the determination of perfluorinated compounds in fish, vegetables and amended soil., *J. Chromatogr. A.* **2014** 1331, 27-37.

SciForum
Mol2Net

Histones Bind, Aggregate and Fuse Phosphoinositides Containing Bilayers

Marta G. Lete[1], Hasna Ahyayauch[1,2], Jesús Sot[1], Félix M. Goni[1] and Alicia Alonso[1*]

[1] Unidad de Biofísica (CSIC, UPV/EHU) and Departamento de Bioquímica, Universidad del País Vasco, Leioa, Spain

[2] Institut de Formation aux Carrieres de Sante de Rabat (IFCSR), Rabat, Morocco

* Author to whom correspondence should be addressed.

Published: 4 December 2015

1. Introduction

Phosphoinositides (PIPns) are negatively charged phospholipids mainly found at the cytosolic surface of membranes. They are considered as minor components of cell membranes because they represent less than a 15 % of the total phospholipids found in eukaryotic cells. However, phosphoinositides are recognised as direct signalling molecules, which can act as second messengers by interacting with effector proteins either electrostatically or via specific phosphoinositides binding domains. This family of lipids is formed from seven members, phosphorylated in different positions, which are constantly being turned over by an array of kinases and phosphatases. Each of them has a unique subcellular distribution (1).

Moreover, the existence of a nuclear pool of phosphoinositides (2) has been described, whose physical state and location has been a matter of controversy, but they are part of the nucleoplasm and perhaps located on invaginations of the nuclear envelope (NE) that penetrate the nucleus. These invaginations are known as the nucleoplasmic reticula.

The roles of phosphoinositides inside the nucleus are unclear but in DNA and RNA polymerase activity upon the addition of phospholipids changes have been observed (3). Lately, it has also been demonstrated that they play important roles in membrane dynamics (4). Recent studies have shown that the NE formation is also dependent on these phospholipids (5). In order to investigate the detailed mechanism of the NE assembly, cell-free systems, which mimic the steps of the NE assembly have been used. These systems have shown that there is a NE precursor membrane vesicle population (MV1), that does not derive from the endoplasmic reticulum (ER), but is essential to the NE assembly and is highly enriched in PIPns (up to a 60 mol %) (5). The NE formation is a vital process that occurs in every mitotic cycle and during fertilisation, and defects induce diseases, such as specific cancers or premature aging diseases (6).

The nuclear envelope encapsulates chromatin, which is highly condensed by histones, the most abundant basic proteins present within the nucleus. Histones contain 25 to 35 % basic amino acid residues. Given their positive net charge, interactions with PIPn is expected. Histones are commonly seen as static molecules that pack DNA but they are highly mobile and dynamic proteins (7).

We used an *in vitro* biophysical approach to obtain a more detailed understanding of the mechanism of interaction between histones and the PIPn.

As a first approach, given the presence and location of phosphoinositides in the eukaryotic cell nucleoplasm, we studied the interaction of these lipids with histones, mainly linker histone H1. Using model membranes, turbidity measurements were performed, revealing that a variety of histones caused a dose-dependent aggregation of phosphatidylcholine (PC) vesicle containing negatively-charged phospholipids. 5 mol % PtdIns(4)P was enough to cause extensive aggregation, while with PtdIns at least 20 mol % (Figure 1) was necessary to obtain a similar effect. With confocal microscopy we were able to visualise H1 binding to vesicles and vesicle aggregation (Figures 2). In order to compare

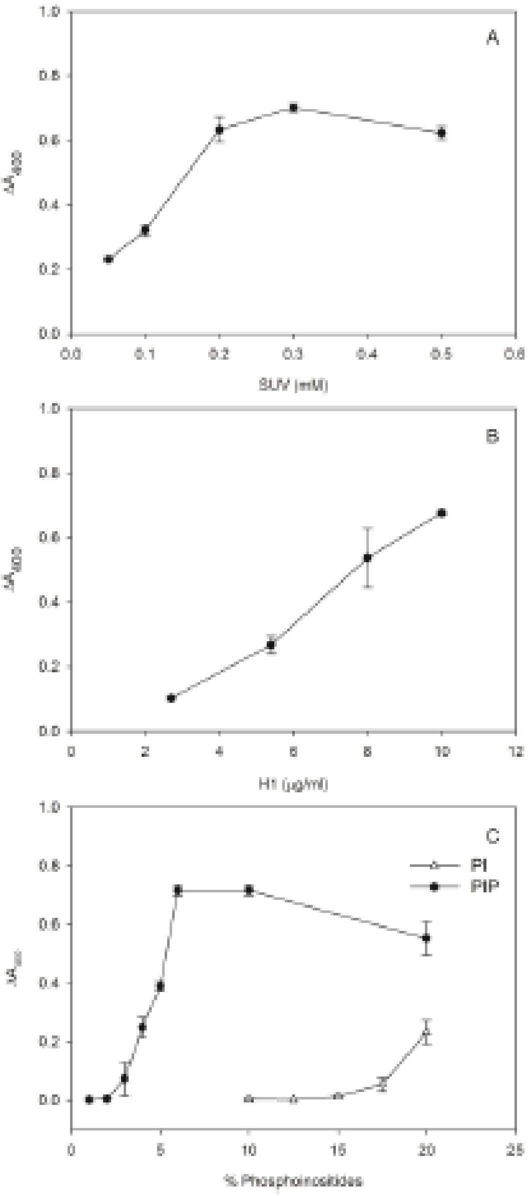


Figure 1. H1-induced aggreation of PIPns contining vesicles. Extents of vesicle aggregation measured as changes in turbidity (Δ A400). **(A)** Increasing concentrations of Small Unilamellar Vesicles (SUV) containing 10 mol % of PtdIns(4)P were treated with 10 µg/ml H1. **(B)** Increasing concentrations of H1 added to 0.3 mM of vesicles containing 10 mol % of PtdIns(4)P. (C) SUV with increasing concentrations of PIPns at 0.3 mM total lipid concentration treated with 10 µg/ml H1. Each point corresponds to the mean (n = 3) ± S. E.

the binding affinities of H1 for vesicles containing PtdIns or PtdIns(4)P, ITC studies were performed, and revealed that the PtdIns(4)P-H1 association constant was one order of magnitude higher than that of PtdIns-H1, and the corresponding lipid/histone stoichiometries were ~ 0.5 and ~ 1.0, respectively (Table 1). This indicated that the two negative charges of PtdIns(4)P are involved in histone binding.

Although these *in vitro* studies indicate that these molecules interact by electrostatic interactions, the fact that both the PIPns and histones are present in the nucleoplasm may suggest also a specific biological function.

Table 1. ITC parameters for the interaction of histone H1 with vesicles of two different lipid compositions. Results in cal/mol injectant. Average values ± S.D. (n = 3)

	PC:PI (9:1)	PC:PIP (9:1)
n	0.99 ± 0.17	0.53 ± 0.13
K_a	$1.08 \pm 0.3\cdot10^{4}$	$9.31 \pm 0.44\cdot10^{4}$
ΔH°	$-4.25 \pm 1.3\cdot10^{6}$ cal/mol	$-1.41 \pm 0.27\cdot10^{6}$ cal/mol
ΔS°	$-1.42 \pm 0.45\cdot10^{4}$ cal/K·mol	$-4.71 \pm 0.91\cdot10^{3}$ cal/K·mol
ΔG°	$-2.99 \pm 0.32\cdot10^{3}$ cal/mol	$-7.9 \pm 0.35\cdot10^{3}$ cal/mol

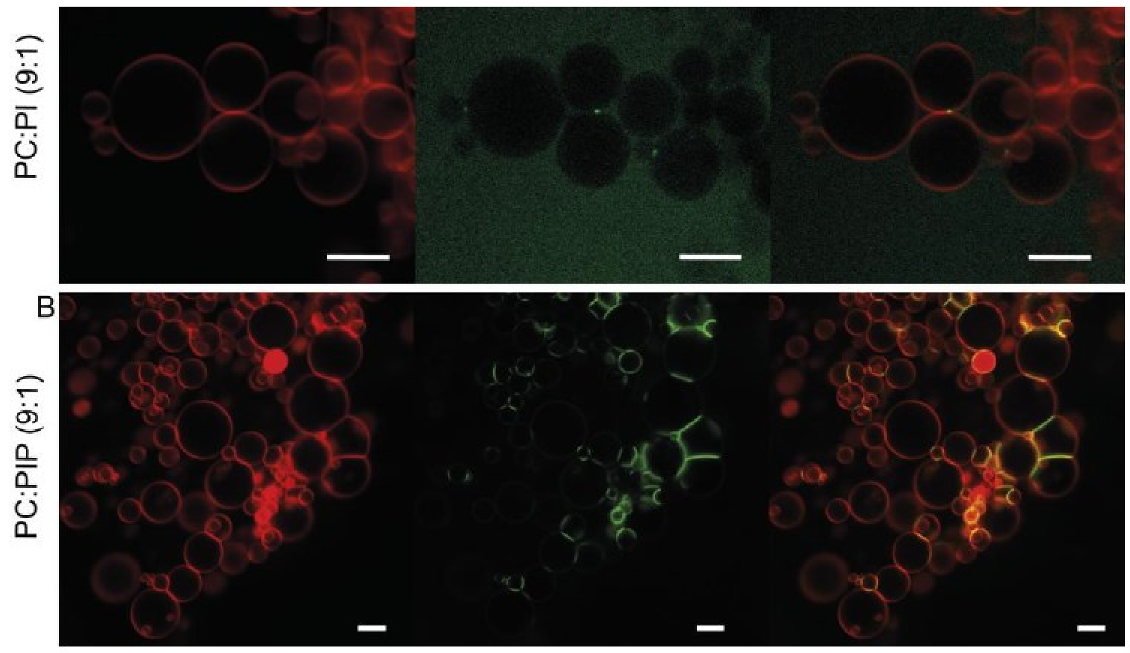


Figure 2. Histone binding to PIPns containing membranes. Representative Giant Unilamellar Vesicles (lipid composition at the left-hand side) imaged by confocal microscopy. (i) Rho-PE for membrane labelling, (ii) H1-Alexa488, and (iii) colocalization of both fluorescent probes. Scale bars 10 μm.

Due to the fact that over every mitotic cycle the NE disassembles and reassembles, fission and fusion processes are required. Therefore we explored whether histones could participate in these events by inducing fusion of phosphoinositides enriched vesicles. This part of our investigation was based on our observation of vesicle-vesicle aggregation, a requisite for *in vitro* vesicle fusion events. To test whether, in addition to aggregation, fusogenic events were taking place, we used fluorescent probes and monitored the change in their spectroscopic properties. We have shown that, in the presence of PtdIns(4)P, histones induce not only intervesicular lipid mixing, but also inner monolayer lipid mixing (Figure 3), which is a diagnostic for *in vitro* membrane fusion.

Even if these results cannot directly demonstrate *in situ* that histones are promoting fusion events at the nuclear membrane, they point towards the likelihood of being involved during nuclear membrane fusion. Our *in vitro* studies corroborate with work performed in the studies by Garnier-Lhomme *et al.* (8) and Byrne *et al.* (5) where the importance of elevated levels of PIPns in NE assembly was demonstrated.

In conclusion, the findings demonstrate that phosphoinositides can interact *in vitro* with the nuclear proteins, histones (specifically H1), not only producing Both, vesicle aggregation and fusion.

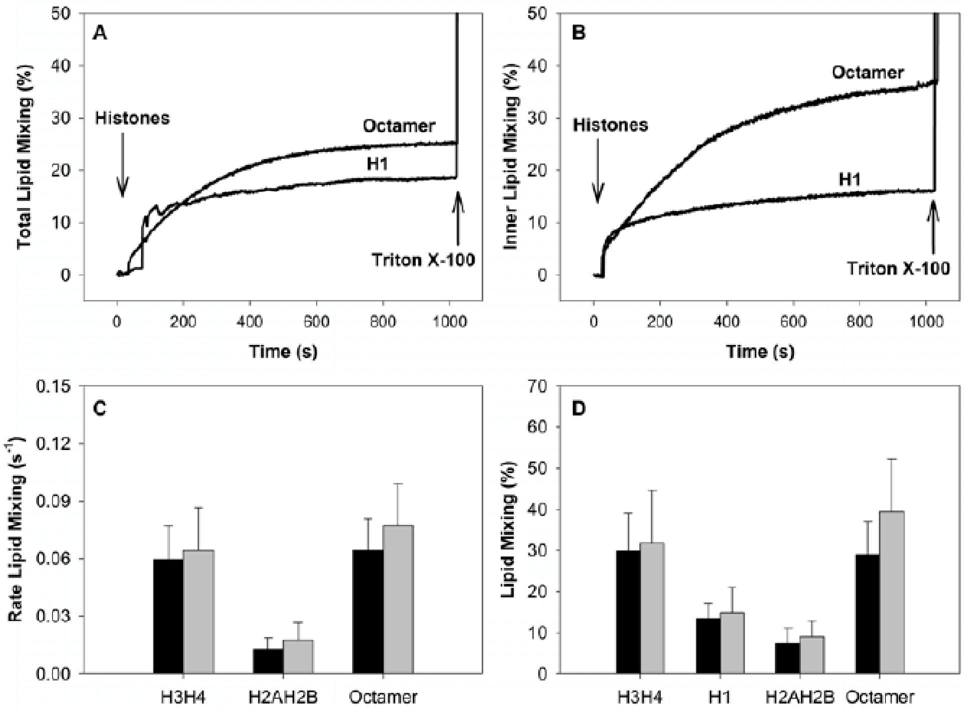


Figure 3. Membrane fusion measured as lipid mixing. Histones induce intervesicular lipid mixing of the inner and outer monolayer. **(A)** Representative total lipid mixing time course. **(B)** Representative inner lipid mixing time course. Arrows indicate the protein or detergent addition. **(C)** Initial rates of histone-induced lipid mixing. **(D)** Extent of histone-induced lipid mixing at equilibrium (20 min). Black corresponds to total lipid mixing and gray correspond to inner lipid mixing. Average values + S.E.M (n =3).

References

1. Di Paolo, G., and P. De Camilli. 2006. Phosphoinositides in cell regulation and membrane dynamics. Nature 443:651-657.

2. Martelli, A. M., L. Manzoli, and L. Cocco. 2004. Nuclear inositides: facts and perspectives. Pharmacology & therapeutics 101:47-64.
3. Cocco, L., A. M. Martelli, and R. S. Gilmour. 1994. Inositol lipid cycle in the nucleus. Cellular signalling 6:481-485.
4. Larijani, B., F. Hamati, A. Kundu, G. C. Chung, M. C. Domart, L. Collinson, and D. L. Poccia. 2014. Principle of duality in phospholipids: regulators of membrane morphology and dynamics. Biochemical Society transactions 42:1335-1342.
5. Byrne, R. D., M. Garnier-Lhomme, K. Han, M. Dowicki, N. Michael, N. Totty, V. Zhendre, A. Cho, T. R. Pettitt, M. J. Wakelam, D. L. Poccia, and B. Larijani. 2007. PLCgamma is enriched on poly-phosphoinositide-rich vesicles to control nuclear envelope assembly. Cellular signalling 19:913-922.
6. Malhas, A. N., and D. J. Vaux. 2014. Nuclear envelope invaginations and cancer. Advances in experimental medicine and biology 773:523-535.
7. Bustin, M., F. Catez, and J. H. Lim. 2005. The dynamics of histone H1 function in chromatin. Molecular cell 17:617-620.
8. Garnier-Lhomme, M., R. D. Byrne, T. M. Hobday, S. Gschmeissner, R. Woscholski, D. L. Poccia, E. J. Dufourc, and B. Larijani. 2009. Nuclear envelope remnants: fluid membranes enriched in sterols and polyphosphoinositides. PloS one 4:e4255.

SciForum
Mol2Net

New Theoretical Model for the Study of New β-Secretase Inhibitors

Jan-carlo Miguel Díaz-González[1], Francisco J. Aguirre-Crespo[1], Xerardo García-Mera[2] and Francisco J. Prado-Prado[1]*

[1] Biomedical Sciences Department, Health Sciences Division, University of Quintana Roo, UQROO, 77039, Mexico.

[2] Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, 15782, Spain.

* Author to whom correspondence should be addressed; Prado-Prado, Francisco: fenol1@hotmail.com.

Published: 4 December 2015

Abstract: Alzheimer's disease (AD) is the most prevalent form of dementia, and current indications show that twenty-nine million people live with AD worldwide, a figure expected rise exponentially over the coming decades. AD is characterize with several pathologies this disease, amyloid plaques, composed of the β-amyloid peptide and γ-amyloid peptide are hallmark neuropathological lesions in Alzheimer's disease brain. Indeed, a wealth of evidence suggests that β-amyloid is central to the pathophysiology of AD and is likely to play an early role in this intractable neurodegenerative disorder. For this reason, we developed a new QSAR (QSAR) model to discover new drugs. A public database ChEMBL contain Big Data sets of inhibitors of *β*-secretase. We revised QSAR studies using method of Artificial Neural Network (ANN) in order to understand the essential structural requirement for binding with receptor for *β*-secretase inhibitors.

.

.

Keywords: QSAR; *β*-secretase inhibitors; Alzheimer's disease (AD).

1. Introduction

Alzheimer´s disease (AD) (1) is a serious and degenerative disorder that causes a the gradual loss of neurons, and in spite of the efforts realized by the big pharmaceutical companies of the world, the origen of this pathology is still not very clear. We can see in this paper that the development of theoretical and QSAR models to study γ-secretase inhibitors are usually not many achieved so far, and most of these works present docking studies. Watching this situation we need to develop QSAR models with γ-secretase inhibitors. In this sense, quantitative structure-activity relationships (QSAR) could play an important role in studying these γ-secretase inhibitors; QSARs can be used as predictive tools for the development of molecules (2, 3). Computer-aided drug design techniques based on Quantitative Structure-Activity Relationships (QSAR) could play an important role in drug discovery programs. The QSAR approach involves the development of models that relate the structure of drugs with their biological activity against different targets (4, 5). In principle, there are currently more than 1600 molecular descriptors that may be generalized and used to solve the problem outlined above (6). Numerous different molecular descriptors have been reported to encode chemical structures in QSAR studies. Furthermore, there are multiple chemometric approaches that can, in principle, be selected for this step. Multiple linear regression (MLR), linear discriminant analysis (LDA) (7), partial least squares (PLS) and different kinds of artificial neural networks (ANN) can be used to relate molecular structure (represented by molecular descriptors) with biological properties. The ANNs are particularly useful in QSAR studies in which the linear models fit poorly due to high data complexity [17;] [18;], an example was the work of Prado-Prado *et. al*. In which four types of artificial neural networks (ANN) were developing for γ-secretase inhibitors, ANNs was constructing from more than 15 000 cases with more than 1 500 different molecules inhibitors of γ-secretase obtained from ChEMBL database http://www.ebi.ac.uk/ChEMBLdb/index.php. We used spectral moments molecular descriptors calculated with Modeslab software (8).

2. METHOD

2.1. Data set

The data set used in this article was obtained from ChEMBL database. It has more than 30 000 cases and more than 1 500 different compounds inhibitors of γ-secretase. In total we used more than 10 000 different molecules to develop the QSAR models obtained in ChEMBL. This is a database of bioactive drug-like small molecules, it contains 2D structures, calculated properties (e.g. logP, Molecular Weight, Lipinski Parameters, etc.) and abstracted bioactivities (e.g. binding constants, pharmacology and ADMET data). ChEMBL normalises the bioactivities into a uniform set of end-points and units where possible, and also tags the links between a molecular target and a published assay with a set of varying confidence levels. The data is abstracted and curated from the primary scientific literature, and covers a significant fraction of the SAR and discovery of modern drugs..

1.2.ANN models

The ANN models are non-linear models useful to predict the biological activity of a large datasets of molecules. This technique is an alternative to linear methods such as LDA. **Figure 1** depicts the networks maps for some of the ANN models. In general, at least one ANN of every types tested was statically significant. However, one must note that the profiles of each network indicate that these are highly nonlinear and complicated models.

There are several different kinds of ANN and these include multilayer perceptron (MLP), radial basis functions (RBF) and PNNs; the latter ANN is a variant of RBF systems. In particular, PNN is a type of neural network that uses a kernel-based approximation to form an estimate of the probability density functions of classes in a classification problem (9).

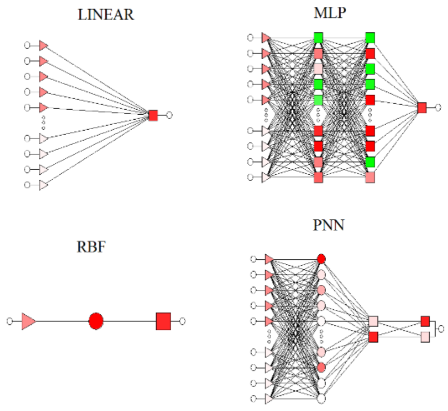


Figure 1. Topology of some ANN models trained in this work.

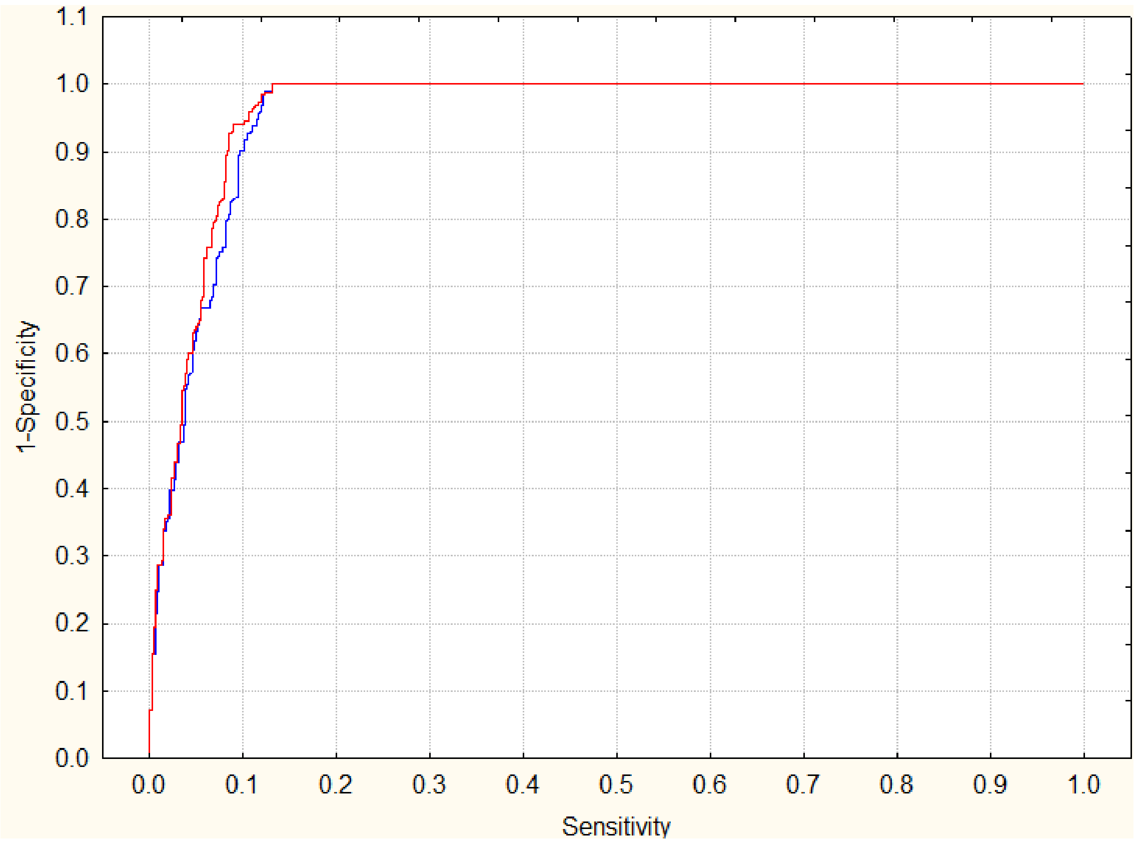


Figure 2. ROC Curve for classifier.

Model	Train			Stat.	Validation		
profile	active	Non-active	%	Par.	active	Non-active	%
LNN	983	129	88.40	Sn	475	81	85.43
14:14-1:1	2943	22415	88.39	Sp	1545	11187	87.87
			88.39	Ac			87.76

PNN	0	1112	0	Sn	0	556	0
14:14-26470-2-2:1	0	25358	100	Sp	0	12732	100
			95.80	Ac			95.82
MLP	781	331	70.23	Sn	396	160	71.22
15:15-9-1:1	7610	17748	69.99	Sp	3879	8853	69.53
			70.00	Ac			69.60
RBF	482	74	86.69	Sn	964	148	86.69
15:15-691-1:1	0	12697	100	Sp	0	25393	100
			99.44	Ac			99.44

Table 1. Comparison of different ANNs classification models.

3. RESULTS AND DISCUSSION

The network found was RBF and it showed training performance higher than 99%. We compare different types of networks to obtain a better model; **Table 1** shows the classification matrix of the different networks. 15:15-691-1:1 was taken as the main network because it presented a wider range of variables, 15 inputs in the first layer and 15 neurons in second layer, and two sets of cases (Training and Validation). Another tested networks found were MLP 15:15-9-1:1, LNN 14:14-1:1 presented low accuracy and PNN 14:14-26470-2-2:1 had a very low percentage of non-active leading to possible errors in the model although its accuracy very good, see **Table 1.** In **Figure 2**, we depict the ROC-curve (10) for LNN tested (ROC=0.96). Notably, almost model presented and an area under curve higher than 0.5 (the value for a random classifier). The vitality of this type of procedures developing ANN-QSAR models has been demonstrated before(11); see, for instance, the work of Fernandez and Caballero (12). The same is true about the ANNs tested, where is illustrated ROC-curve of ANN LNN with an area higher than 0.93. To show how important is this result, we compared the present model with other model used to address the same problem. We processed our data with Artificial Neural Networks (ANNs) looking for a better model. In general, the ANN RBF tested was statically significant.

4. CONCLUSIONS

Theoretical studies such as QSAR models have become a very useful tool in this context substantially reduce time and resources consuming experiments. The functions of γ-secretase and its implication in Alzheimer's disease have triggered an active search for potent and selective γ-secretase inhibitors. In this sense, QSAR could play an important role in studying these γ-secretase inhibitors. QSARs can be use as predictive tools for the development of molecules. In this work, we developed a new ANN RBF model using the ModesLab descriptors, based on a large database using about 10,000 different drugs obtained from the ChEMBL server. A very good model obtained, and it predict near to 99% γ-secretase inhibitors. This model could be a goal to discover new drugs to treat AD.

.

ACKNOWLEDGEMENTS

The authors thank the sponsorship DCS-UQROO PROFOCIE (P/PROFOCIE-2014-23MSU0140Z-09DCS) and sponsorships for research project PRODEP-Fortalecimiento Cuerpos Académicos 2012..

REFERENCES

1. Salmon SA, Watts JL. Minimum inhibitory concentration determinations for various antimicrobial agents against 1570 bacterial isolates from turkey poults. Avian Dis. 2000 Jan-Mar;44(1):85-98.

2. Chou KC. Review: Structural bioinformatics and its impact to biomedical science. Current Medicinal Chemistry. 2004;11:2105-34.

3. Chou KC, D. Q. Wei, Q. S. Du, S. Sirois, and W. Z. Zhong. . Review: Progress in computational approach to drug development against SARS. Current Medicinal Chemistry 2006;13:3263-70.

4. Prado-Prado FJ, Gonzalez-Diaz H, de la Vega OM, Ubeira FM, Chou KC. Unified QSAR approach to antimicrobials. Part 3: first multi-tasking QSAR model for input-coded prediction, structural back-projection, and complex networks clustering of antiprotozoal compounds. Bioorg Med Chem. 2008 Jun 1;16(11):5871-80.

5. Prado-Prado FJ, de la Vega OM, Uriarte E, Ubeira FM, Chou KC, Gonzalez-Diaz H. Unified QSAR approach to antimicrobials. 4. Multi-target QSAR modeling and comparative multi-distance study of the giant components of antiviral drug-drug complex networks. Bioorg Med Chem. 2009;17:569–75.

6. Kubinyi H. Quantitative structure-activity relationships (QSAR) and molecular modelling in cancer research. J Cancer Res Clin Oncol. 1990;116(6):529-37.

7. Prado-Prado FJ, Borges F, Perez-Montoto LG, Gonzalez-Diaz H. Multi-target spectral moment: QSAR for antifungal drugs vs. different fungi species. Eur J Med Chem. 2009 May 5.

8. Estrada E. On the topological sub-structural molecular design (TOSS-MODE) in QSPR/QSAR and drug design research. SAR QSAR Environ Res. 2000;11(1):55-73.

9. Mosier PD, Jurs PC. QSAR/QSPR studies using probabilistic neural networks and generalized regression neural networks. J Chem Inf Comput Sci. 2002 Nov-Dec;42(6):1460-70.

10. Lombardi G, Gramegna G, Cavanna C, Michelone G. Fluconazole vs amphotericin B: "in vitro" comparative evaluation of the minimal inhibitory concentration (MIC) against yeasts isolated from AIDS patients. Microbiologica. 1990 Jul;13(3):201-6.

11. Prado-Prado FJ, Garcia-Mera X, Gonzalez-Diaz H. Multi-target spectral moment QSAR versus ANN for antiparasitic drugs against different parasite species. Bioorg Med Chem. 2010 Mar 15;18(6):2225-31.

12. Fernandez M, Caballero J, Tundidor-Camba A. Linear and nonlinear QSAR study of N-hydroxy-2-[(phenylsulfonyl)amino]acetamide derivatives as matrix metalloproteinase inhibitors. Bioorg Med Chem. 2006 Jun 15;14(12):4137-50.

SciForum
Mol2Net

Two QSAR Paradigms- Congenericity Principle *versus* Diversity Begets Diversity Principle- Analyzed Using Computed Mathematical Chemodescriptors of Homogeneous and Diverse Sets of Chemical Mutagens

Subhash C. Basak [1]* Subhabrata Majumdar[2]

[1] University of Minnesota Duluth-Natural Resources Research Institute (UMD-NRRI) and Department of Chemistry and Biochemistry, University of Minnesota Duluth, 5013 Miller Trunk Highway, Duluth, MN 55811, USA; sbasak@nrri.umn.edu

[2] School of Statistics, University of Minnesota Twin Cities, Minneapolis, MN 55414, USA

* Author to whom correspondence should be addressed; E-Mail: sbasak@nrri.umn.edu
Tel.: +1-218-727-1335

Published: 4 December 2015

Abstract: The age old paradigm of quantitative structure-activity relationship (QSAR) is the congenericity principle which states that similar structures usually have similar properties. But these days a lot of large and structurally diverse data sets of chemicals with the same experimental data (dependent variable) are available. Starting with the same classes of descriptors we extracted the two subsets of the most significant predictors for the formulation of QSARs for two sets of chemicals: A homogeneous set of 95 amine mutagens and a diverse set of 508 structurally diverse mutagens. The predictors included calculated topostructural (TS), topochemical (TC), geometrical, and quantum chemical (QC) indices. Whereas for the homogeneous amines, a small group of descriptors were sufficient for QSAR development, for the 508 diverse set we needed a large and diverse set of indices for effective QSAR formulation. This empirical study thus vindicates the DIVERSITY BEGETS DIVERSITY paradigm of QSAR.

Keywords: quantitative structure-activity relationship (QSAR); Congenericity principle; Diversity begets diversity paradigm; Aromatic amine mutagens; Structurally diverse mutagens; Topostructural (TS) indices; Topochemical (TC) indices; Geometrical indices; Quantum chemical indices.

1. Introduction

Quantitative structure-activity relationships (QSARs) pertaining to the prediction of physicochemical, pharmacological, and toxicological properties of chemicals are mathematical models developed for the prediction of properties of chemicals from their physical properties or structural descriptors [1-10]. The basic idea underlying QSAR development can be conveniently expressed by the following equation:

P = f (S) ………….. Eq. 1

where P is any physical, biological, medicinal or toxicological property of interest and S represents the relevant aspect of the structure that determines the property. A look at the recently published literature would show that various classes of calculated properties, viz., topological, geometrical, quantum chemical, substructural, are used routinely in QSAR formulation [2-15].

A perusal of QSAR literature would indicate that in many cases QSARs are developed based on properties of a set of structurally related molecules. This is based on the "congenericity principle" or the "structure-property similarity principle" which states that similar structures usually have similar properties [11]. But in many practical situations one has to develop models for the prediction of property/ bioactivity of sets of chemicals which are structurally diverse instead of being homogeneous [12]. In the course of carrying out principal components analysis (PCA) of homogeneous and diverse data sets, Basak et al [13, 14] noted that a larger number of PCs are required to explain an analogous percentage of variance for diverse data sets as compared to homogeneous collection of molecules. From such QSAR studies Basak [15] formulated the

"diversity begets diversity principle," which states that we need a diverse collection of descriptors independent variables) if we want to develop a QSAR for structurally diverse sets of chemicals. We have tested this hypothesis using two data sets: a) Mutagenicity of a homogeneous set of 95 aromatic and heteroaromatic amines, and b) Mutagenic activity of a large diverse set of 508 chemicals.

2. Results and Discussion

The results of QSAR based on the 95 aromatic amine mutagens [16] and 508 diverse mutagens [17] are shown in Table 1 based on calculated descriptors described in Table 2. After QSAR development of the two sets of mutagens, one congeneric and the other structurally diverse, we sorted the descriptors based on their significance as measured by [t] values which are given in Table 3. As evident from data in Table 3, for the set of congeneric mutagens, only seven molecular descriptors of limited class diversity were sufficient to give a reasonably good QSAR. Starting from the same set of calculated descriptors (Table 2), the significant descriptors for the 508 diverse mutagens needed 42 descriptors for good QSAR development. Whereas the indices needed for 95 amines fall into some narrow classes, those needed for 508 chemical set need not only higher number of descriptors, but also heterogeneous types of descriptors. For example, the diverse set of mutagens needed triplet indices (ASV1), information theoretic indices of neighborhood complexity (IC, SIC, CIC indices of different orders), and the quantum chemical descriptors (HOMO, LUMO) which were not selected for the congeneric amine data set. This supports the dichotomy in the QSAR paradigms: Congenericity principle for congeneric data set and diversity begets diversity principle for structurally diverse situations.

Table 1: Results of QSAR based on the 95 aromatic amine mutagens and 508 diverse mutagens

508 compound	**No of predictors**	**% Correct classification**	**Sensitivity**	**Specificity**
TS+ TC Model (Ridge Regression)	**298**	**76.97**	**83.98**	**69.84**
TS + TC Model (using ITC+ Ridge Regression)	**298**	**73.23**	**77.34**	**69.05**
95 Aromatic amine mutagens				
TS+ TC Model (Ridge Regression)	**266**	**84.21**	**77.36**	**92.86**
TS + TC Model (using ITC+ Ridge Regression)	**266**	**89.47**	**92.45**	**85.71**

Table 2: Molecular Descriptors used for QSAR development

	Topostructural (TS)
I_D^W	Information index for the magnitudes of distances between all possible pairs of vertices of a graph
$\overline{I_D^W}$	Mean information index for the magnitude of distance
W	Wiener index = half-sum of the off-diagonal elements of the distance matrix of a graph
I^D	Degree complexity
H^V	Graph vertex complexity
H^D	Graph distance complexity
$\overline{IC}$	Information content of the distance matrix partitioned by frequency of occurrences of distance h
M_1	A Zagreb group parameter = sum of square of degree over all vertices
M_2	A Zagreb group parameter = sum of cross-product of degrees over all neighboring (connected) vertices
${}^h\chi$	Path connectivity index of order $h = 0\text{-}10$
${}^h\chi_C$	Cluster connectivity index of order $h = 3\text{-}6$
${}^h\chi_{PC}$	Path-cluster connectivity index of order $h = 4\text{-}6$
${}^h\chi_{Ch}$	Chain connectivity index of order $h = 3\text{-}10$
P_h	Number of paths of length $h = 0\text{-}10$
J	Balaban's J index based on topological distance
nrings	Number of rings in a graph
ncirc	Number of circuits in a graph

DN^2S_y	Triplet index from distance matrix, square of graph order, and distance sum; operation y = 1-5
DN^21_y	Triplet index from distance matrix, square of graph order, and number 1; operation y = 1-5
$AS1_y$	Triplet index from adjacency matrix, distance sum, and number 1; operation y = 1-5
$DS1_y$	Triplet index from distance matrix, distance sum, and number 1; operation y = 1-5
ASN_y	Triplet index from adjacency matrix, distance sum, and graph order; operation y = 1-5
DSN_y	Triplet index from distance matrix, distance sum, and graph order; operation y = 1-5
DN^2N_y	Triplet index from distance matrix, square of graph order, and graph order; operation y = 1-5
ANS_y	Triplet index from adjacency matrix, graph order, and distance sum; operation y = 1-5
$AN1_y$	Triplet index from adjacency matrix, graph order, and number 1; operation y = 1-5
ANN_y	Triplet index from adjacency matrix, graph order, and graph order again; operation y = 1-5
ASV_y	Triplet index from adjacency matrix, distance sum, and vertex degree; operation y = 1-5
DSV_y	Triplet index from distance matrix, distance sum, and vertex degree; operation y = 1-5
ANV_y	Triplet index from adjacency matrix, graph order, and vertex degree; operation y = 1-5
kp_0	Kappa zero
kp_1-kp_3	Kappa simple indices
Topochemical (TC)	
O	Order of neighborhood when IC_r reaches its maximum value for the hydrogen-filled graph
O_{orb}	Order of neighborhood when IC_r reaches its maximum value for the hydrogen-suppressed graph
I_{ORB}	Information content or complexity of the hydrogen-suppressed graph at its maximum neighborhood of vertices
IC_r	Mean information content or complexity of a graph based on the r^{th} (r = 0-6) order neighborhood of vertices in a hydrogen-filled graph
SIC_r	Structural information content for r^{th} (r = 0-6) order neighborhood of vertices in a hydrogen-filled graph
CIC_r	Complementary information content for r^{th} (r = 0-6) order neighborhood of vertices in a hydrogen-filled graph
$^h\chi^b$	Bond path connectivity index of order h = 0-6
$^h\chi^b_C$	Bond cluster connectivity index of order h = 3-6
$^h\chi^b_{Ch}$	Bond chain connectivity index of order h = 3- 6
$^h\chi^b_{PC}$	Bond path-cluster connectivity index of order h = 4-6
$^h\chi^v$	Valence path connectivity index of order h = 0-10
$^h\chi^v_C$	Valence cluster connectivity index of order h = 3-6
$^h\chi^v_{Ch}$	Valence chain connectivity index of order h = 3-10
$^h\chi^v_{PC}$	Valence path-cluster connectivity index of order h = 4-6
J^B	Balaban's J index based on bond types
J^X	Balaban's J index based on relative electronegativities
J^Y	Balaban's J index based on relative covalent radii
AZV_y	Triplet index from adjacency matrix, atomic number, and vertex degree; operation y = 1-5
AZS_y	Triplet index from adjacency matrix, atomic number, and distance sum; operation y = 1-5
ASZ_y	Triplet index from adjacency matrix, distance sum, and atomic number; operation y = 1-5

AZN$_y$	Triplet index from adjacency matrix, atomic number, and graph order; operation y = 1-5
ANZ$_y$	Triplet index from adjacency matrix, graph order, and atomic number; operation y = 1-5
DSZ$_y$	Triplet index from distance matrix, distance sum, and atomic number; operation y = 1-5
DN^{2}Z$_y$	Triplet index from distance matrix, square of graph order, and atomic number; operation y = 1-5
nvx	Number of non-hydrogen atoms in a molecule
nelem	Number of elements in a molecule
fw	Molecular weight
si	Shannon information index
totop	Total Topological Index *t*
sumI	Sum of the intrinsic state values *I*
sumdelI	Sum of delta-*I* values
tets2	Total topological state index based on electrotopological state indices
phia	Flexibility index (kp_1* kp_2/nvx)
Idcbar	Bonchev-Trinajstić information index
IdC	Bonchev-Trinajstić information index
Wp	Wiener *p*
Pf	Platt *f*
Wt	Total Wiener number
knotp	Difference of chi-cluster-3 and path/cluster-4
knotpv	Valence difference of chi-cluster-3 and path/cluster-4
nclass	Number of classes of topologically (symmetry) equivalent graph vertices
NumHBd	Number of hydrogen bond donors
NumHBa	Number of hydrogen bond acceptors
SHCsats	E-State of C *sp*^{3} bonded to other saturated C atoms
SHCsatu	E-State of C *sp*^{3} bonded to unsaturated C atoms
SHvin	E-State of C atoms in the vinyl group, *=CH-*
SHtvin	E-State of C atoms in the terminal vinyl group, $=CH_2$
SHavin	E-State of C atoms in the vinyl group, *=CH-*, bonded to an aromatic C
SHarom	E-State of C *sp*^{2} which are part of an aromatic system
SHHBd	Hydrogen bond donor index, sum of Hydrogen E-State values for *–OH*, *=NH*, -NH_2, *-NH-,-SH*, and *#CH*
SHwHBd	Weak hydrogen bond donor index, sum of *C-H* Hydrogen E-State values for hydrogen atoms on a C to which a F and/or Cl are also bonded
SHHBa	Hydrogen bond acceptor index, sum of the *E*-State values for *–OH*, *=NH*, -NH_2, *-NH-*, *>N*, *-O-*, *-S-*, along with –F and –Cl
Qv	General Polarity descriptor
NHBint$_y$	Count of potential internal hydrogen bonders (y = 2-10)
SHBinty	E-State descriptors of potential internal hydrogen bond strength (*y =2-10*)
ka_1-ka_3	Kappa alpha indices

Electrotopological State index values for atom types:

	SHsOH, SHdNH, SHsSH, SHsNH2, SHssNH, SHtCH, SHother, SHCHnX, Hmax, Gmax, Hmin, Gmin, Hmaxpos, Hminneg, SsLi, SssBe, Sssss, Bem, SssBH ,SsssB, SssssBm, SsCH3, SdCH2, SssCH2, StCH, SdsCH, SaaCH, SsssCH, SddC, StsC, SdssC, SaasC, SaaaC, SssssC, SsNH3p, SsNH2, SssNH2p, SdNH, SssNH, SaaNH, StN, SsssNHp, SdsN, SaaN, SsssN, SddsN, SaasN, SssssNp, SsOH, SdO, SssO, SaaO, SsF, SsSiH3, SssSiH2, SsssSiH, SssssSi, SsPH2, SssPH, SsssP, SdsssP, SssssssP, SsSH, SdS, SssS, SaaS, SdssS, SddssS, SssssssS, SsCl, SsGeH3, SssGeH2, SsssGeH, SssssGe, SsAsH2, SssAsH, SsssAs, SdsssAs, SssssAs, SsSeH, SdSe, SssSe, SaaSe, SdssSe, SddssSe, SsBr, SsSnH3, SssSnH2, SsssSnH, SssssSn, SsI, SsPbH3, SssPbH2, SsssPbH, SssssPb
	Geometrical (3-D)
^{3D}W	3D Wiener number based on the hydrogen-suppressed geometric distance matrix
$^{3D}W_H$	3D Wiener number based on the hydrogen-filled geometric distance matrix
V_W	Van der Waal's volume
	Quantum Chemical (QC)
E_{HOMO}	Energy of the highest occupied molecular orbital
E_{HOMO-1}	Energy of the second highest occupied molecular
E_{LUMO}	Energy of the lowest unoccupied molecular orbital
E_{LUMO+1}	Energy of the second lowest unoccupied molecular orbital
ΔHf	Heat of formation
μ	Dipole moment

Table 3: Most significant descriptors for the two sets based on *t*-ratio.

508 compound dataset (42 descriptors have significant t-ratios)

t-ratio	*Descriptor name*	*t-ratio*	*Descriptor name*
-30.78	ASV_1	4.75	kp_2
-21.26	$SHBint_6$	-4.09	*StN*
-19.59	*S1*	-3.76	*SddC*
-17.50	*HF*	-3.67	$\overline{IC}$
11.49	*S2*	-3.47	SIC_4
-10.80	SIC_3	3.39	kp_3
-10.51	E_{HOMO}	-3.32	*totop*
9.35	E_{LUMO+1}	-3.14	*SHdNH*
-8.75	E_{LUMO+1}	3.06	*SsF*
8.53	IC_6	-3.05	S_6
7.93	*SHssNH*	2.96	*nelem*
7.83	*SHHBa*	2.96	CIC_4
-7.72	ka_1	-2.61	CIC_6
-7.68	*SHarom*	2.53	*SssssC*
7.24	*SaaaC*	-2.35	*SaasC*
6.64	DN^21_2	-2.33	$DS1_1$

-6.35	*DN^21_3*	*-2.33*	*$NHBint_3$*
6.00	*μ*	*-2.22*	*SsNH2*
-5.50	*SIC_0*	*2.19*	*DSZ_2*
-5.39	*kp_1*	*2.04*	*knotpv*
4.85	*SssH*	*2.03*	*SaaCH*

95 compound dataset
(7 descriptors have significant t-ratios)

t-ratio	**Descriptor name**
-8.89	**SsNH2**
-5.72	**NHBint3**
4.65	**NHBint9**
-3.63	**NumHBd**
-3.20	**NumHBd**
-3.17	**NumHBd**
2.54	**SssNH**

3. Materials and Methods

A machine learning method called Interrelated Two-way clustering (ITC), originally developed for application in gene microarray data [72], is used for variable selection, and resulting predictors are fed into a ridge regression model to get final predictions. The ITC algorithm involves the following steps:

i. Predictors are clustered into separate functional groups, say $G_1, G_2, \dots, G_k$, which are substituted by several types of descriptors in QSAR;
ii. After that samples are clustered into two classes using each functional group, Say $S_{i,a}$ and $S_{i,b}; i = 1,2, \dots, n$;
iii. All possible intersections of the 2^k clusters are taken. For example, for $k = 2$ the intersections are:
$$C_1 = S_{1,a} \cap S_{2,a};\ C_2 = S_{1,b} \cap S_{2,a};$$
$$C_3 = S_{1,a} \cap S_{2,b};\ C_4 = S_{1,b} \cap S_{2,b}$$
iv. These are divided into heterogeneous groups: pairs of intersections with no common elements, e.g. $H_{14} = (C_1, C_4)$ and $H_{23} = (C_2, C_3)$ above;
v. For each $H_{st} = (C_s, C_t)$, cosine distances of subvectors with predictors from this heterogeneous group are calculated with the two model vectors: one with C_s zeros and C_t ones, and another with C_s ones and C_t zeros. Each distance vector is sorted in decreasing order, top one-third of predictors are taken from each of these vectors and are merged.

The algorithm is then repeated with selected predictors, and terminated when 90% of total number of samples is covered by the largest heterogeneous group, or maximum number of iterations reached. This is done because through the algorithm the functional groups become more and more similar, so sample classifications using them become more and more similar, thus heterogeneous groups cover an increasing proportion of total number of samples.

To tackle high collinearity among different predictors, after variable selection through ITC we use ridge regression to build the predictive models. Given n samples and p variables, the $n \times p$ data matrix of predictors $\mathbf{X}$ and $n \times 1$ vector of 0/1 responses $\mathbf{Y}$, the vector of coefficients obtained by ridge regression is defined as:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{Y} \qquad (1)$$

Where $k > 0$ is the ridge constant, chosen by cross-validation.

3. Conclusions

The aim of research in this paper was to test the hypothesis that congeneric data sets need a structurally narrow set of descriptors for QSAR formulation as compared to diverse sets which require diverse collection of independent variables for model building. The results derived from two data sets, viz. a congeneric set of 95 amines and a diverse set of 508 structurally diverse mutagens appear to support the “diversity begets diversity” hypothesis. Further QSAR studies on other congeneric and diverse data sets are necessary to test the validity of this hypothesis.

Acknowledgments.

Author Contributions

Subhash C. Basak developed the hypothesis and Subhabrata Majumdar carried out the data analysis for the paper.

Conflicts of Interest

“The authors declare no conflict of interest”.

References and Notes

[1] Hansch, C.; Leo, A., Exploring QSARs: Fundamentals and Applications in Chemistry and Biology, American Chemical Society, Washington, DC 1995.-

[2] Kier, L.B.; Hall, L. Molecular Structure Description: The Electrotopological State; Academic Press: San Diego, CA, 1999, pp. 245.

[3] Devillers, J.; Balaban, A.T., Eds. Topological Indices and Related Descriptors in QSAR and QSPR; Gordon and Breach: Amsterdam, 1999, pp. 811.

[4] Diudea, M.V., Ed. QSPR / QSAR Studies by Molecular Descriptors; Nova: Huntington, N.Y., 2001, pp. 438.

[5] Karelson, M. Molecular Descriptors in QSAR/QSPR; Wiley-Interscience: New York, 2000, pp. 448.

[6] Balaban, A.T., Ed. From Chemical Topology To Three-Dimensional Geometry; Plenum Press: 1997,

pp. 420.

[7] Todeschini, R.; Consonni, V. Molecular Descriptors for Chemoinformatics; Wiley-VCH: Weinheim, 2009,Vol. I, pp. 967; Vol. II, pp. 257.

[8[Hawkins, D. M.; Basak, S. C.; Kraker, J. J.; Geiss, K. T.; Witzmann, F. A., Combining chemodescriptors and biodescriptors in quantitative structure-activity relationship modeling, J. Chem. Inf. Model., 2006, 46, 9–16.

[9] Basak, S. C., Role of Mathematical Chemodescriptors and Proteomics-Based Biodescriptors in Drug Discovery, Drug Develop Res, 2010, 72, 1-9.

[10] Basak, S. C.; Mills, D.; Hawkins, D. M., Predicting allergic contact dermatitis: A hierarchical structure-activity relationship (SAR) approach to chemical classification using topological and quantum chemical descriptors. J. Comput. Aided Mol. Des., 2008, 22, 339-343.

[11] Johnson, M, Basak, S. C.;, Maggiora, G., A characterization of molecular similarity methods for property prediction. Mathl. Comput. Modelling 1988, 11, 630–634.

[12] Basak, S. C.; Grunwald, G. D.; Host, G.; Niemi, G. J.; Bradbury, S. P. A comparative study of molecular similarity, statistical and neural network methods for predicting toxic modes of action of chemicals, Environ. Toxicol. Chem., 1998, 17, 1056–1064.

[13] Basak, S. C.; Mills, D.; Gute, B. D.; Balaban, A. T.; Basak, K.; Grunwald, G. D. Use of Mathematical Structural Invariants in Analyzing, Combinatorial Libraries: A Case Study with psoralen Derivatives, Current Computer-Aided Drug Design, 2010, 6, 240-251.

[14] Basak, S. C.; Grunwald, G. D., A COMPARATIVE STUDY OF GRAPH INVARIANTS, TOTAL SURFACE AREA AND VOLUME IN PREDICTING BOILING POINTS OF ALKANES, Math Modelling & Sci. Computing, 1993, 2, 735-740.

[15] Basak, S. C., Mathematical Descriptors for the Prediction of Property, Bioactivity, and Toxicity of Chemicals from their Structure: A Chemical-Cum-Biochemical Approach, Current Computer-Aided Drug Design, 2013, 9, 449-462.

[16] Debnath, A.K.; Debnath, G.; Shusterman, A.J.; Hansch, C. A QSAR Investigation of the Role of Hydrophobicity in Regulating Mutagenicity in the Ames Test: 1. Mutagenicity of Aromatic and Heteroaromatic Amines in Salmonella typhimurium TA98 and TA100. Environ. Mol. Mutagen. 1992, 19, 37-52.

[17] Soderman, J.V. CRC Handbook of Identified Carcinogens and Noncarcinogens: Carcinogenicity-Mutagenicity Database, CRC Press: Boca Raton, FL, 1982.

SciForum
Mol2Net

Intrinsic Dimensionality of Chemical Space: Characterization and Applications

Subhash C. Basak [1]* Gregory D. Grunwald[2] and Subhabrata Majumdar[3]

[1] University of Minnesota Duluth-Natural Resources Research Institute (UMD-NRRI) and Department of Chemistry and Biochemistry, University of Minnesota Duluth, 5013 Miller Trunk Highway, Duluth, MN 55811, USA; sbasak@nrri.umn.edu

[2] University of Minnesota Duluth-Natural Resources Research Institute (UMD-NRRI), 5013 Miller Trunk Highway, Duluth, MN 55811, USA

[3] School of Statistics, University of Minnesota, Twin Cities, Minneapolis, MN 55414

* Author to whom correspondence should be addressed; E-Mail: sbasak@nrri.umn.edu
Tel.: +1-218-727-1335

Published: 4 December 2015

Abstract:

One popular method for the representation and characterization of chemical structure is through the use of their computed mathematical descriptors. Such descriptors, often called molecular descriptors, quantify different aspects of molecular structure, viz., size, shape, branching, cyclicity, bonding patterns, etc. Applications of discrete mathematics in the development of molecular descriptors began in the middle of the twentieth century and the trend is going on in an unabated manner even today. While in the 1970s only a few descriptors could be calculated, currently available software can calculate a large number of descriptors for molecules or biomolecules like DNA/ RNA, proteins. When p molecular descriptors are calculated for n molecules, the data set can be viewed as n vectors in p dimensions, each chemical being represented as a point in R^p. Because many of the descriptors are strongly correlated, the n points in R^p will lie on a subspace of dimension lower than p. Methods like principal components analysis (PCA) can be used to characterize the intrinsic dimensionality of chemical spaces. Since the early 1980s, Basak et al have carried out PCA of various congeneric and diverse data sets relevant to new drug discovery and predictive toxicology. Principal components (PCs) derived from mathematical chemodescriptors have been used in the formulation of quantitative structure-activity relationships (QSARs), clustering of large combinatorial libraries as well as quantitative molecular similarity analysis (QMSA). This presentation will review the results of PCA carried out by Basak and coworkers since the early 1980s to the present time in the characterization and visualization of chemical spaces with special reference to five data sets, both congeneric and structurally diverse: 1) A large and structurally diverse set of 3,692 chemicals which was a subset of the Toxic Substances Control Act (TSCA) Inventory maintained by the United States Environmental Protection Agency (USEPA), 2) A set of 74 alkanes, 3) A virtual library of 248,832 psoralen derivatives, 4) A congeneric set of 95 aromatic and heteroaromatic amine mutagens, and 5) A structurally diverse collection of 508 chemicals mutagens.

Keywords: Computed mathematical descriptors; Principal components analysis (PCA); Congeneric set; Diverse set; Psoralen derivatives; alkanes, Subset of the Toxic Substances Control Act (TSCA) Inventory; Structurally diverse mutagens; Congeneric set of aromatic and heteroaromatic amine mutagens; Intrinsic dimensionality of chemical space; Quantitative structure-activity relationship (QSAR); Clustering of combinatorial libraries; Quantitative molecular similarity analysis (QMSA).

Mol2Net YouTube channel*: http://bit.do/mol2net-tube*

1. Introduction

Mathematical chemistry, or more correctly discrete mathematical chemistry, had its beginning at the middle of the twentieth century probably with the publication of the seminal paper by Harry Wiener [1] on the calculation of .structural indices for the prediction of molecular properties. Although representation of chemical species by graphs was explored by Sylvester [2] as early as 1878, the characterization of molecular structure by graph invariants has made great strides during the past half century or so [3-16] following the seminal work of Wiener [1]. Invariants of graphs associated with molecules and biomolecules quantify certain aspects of their structure and have been used in the characterization and comparison of such structures as well as prediction of their properties [4, 17, 18, 19]. Specifically, such invariants and orthogonal factors like principal components PCs) derived from them have found applications in quantitative structure-activity relationship (QSAR) studies [3-15, 20], quantitative molecular similarity analysis (QMSA) research [21-24], clustering of large libraries of structures into smaller subsets [23, 24], and in the discrimination of pathological structures like isospectral graphs [17]. One of the authors of this paper (Basak) has been involved since the early 1970s in the development of novel numerical graph invariants or topological indices (TIs) [6, 7, 11, 25-27] as well as biodescriptors derived from DNA/ RNA sequences [28] and proteomics maps [29]. Basak's research [20] carried out with colleagues at the University of Calcutta, India, in the 1970s involved mainly formulation of QSARs of congeneric sets of chemicals using their own information theore5tic indices and topological indices developed by Bonchev & Trinajstić [4, 5], Randic [9-12] & Kier and Hall [3] as well as physical properties like van der Waals' volume, calculated or experimental hydrophobicity (log P, octanol water) [20]. In the early 1980s, after Basak joined the University of Minnesota Duluth, the software POLLY [30] was developed and large scale calculation of TIs for QSAR and QMSA analyses was initiated. In one of the earliest studies of its kind, Basak et al [31] used POLLY for the calculation of ninety TIs for a collection of 3,692 structurally diverse chemicals which was a subset of the Toxic Substances Control Act (TSCA) Inventory of the United States Environmental Protection Agency (USEPA). The authors carried out principal components analysis (PCA) on this data set and asked the question: *What is the intrinsic dimensionality of chemical structure measured by the large number of TIs*? This line of research, i.e., PCA and use of principal components (PCs) derived from different collection of TIs calculated by POLLY [30], MolConnZ [32], Triplet [33, 34], and APProbe [35] in QSAR and QMSA, has continued to this day. This paper summarizes the results and the lessons learned from a few of these studies using both congeneric and structurally diverse sets of chemicals, viz., 1) A large and structurally diverse set of 3,692 chemicals mentioned above, and 2) A data set of 74 alkanes, 3) A virtual library of 248,832 psoralen derivatives, 4) A congeneric set of 95 aromatic and heteroaromatic amine mutagens, and 5) A structurally diverse collection of 508 chemicals mutagens.

.

2 Results and Discussion

2.1 A large and structurally diverse set of 3,692 chemicals.

For this data set, 90 TIs were calculated by the POLLY [30] software and PCA was performed. For details of the list of the particular TIs calculated for this study see Basak et al [21, 31]. Results showed that first ten PCs with eigenvalues greater than or equal to 1 explained 92.6% of the variance in the data and PC$_1$-PC$_4$ explained 78.3% of the variation in the original variables. Regarding the correlation profiles of the original variables or TIs with the first four important PCs, Table 1 below gives the data:

Table 1: Correlation of the first four PCs with the original variables including topological indices.

PC1	PC2	PC3	PC4
K$_1$ (.96)	SIC$_3$ (0.97)	$^4\chi^b{}_C$ (.69)	$^4\chi_{CH}$ (.85)
$^2\chi$ (.95)	CIC$_4$ (-.96)	$^4\chi^b{}_C$ (.69)	$^4\chi^b{}_{CH}$ (.84)
$^3\chi$ (.95)	CIC$_3$ (-.95)	$^5\chi^b{}_C$ (.68)	$^4\chi^v{}_{CH}$ (.80)
K$_2$ (.95)	SIC4 (.95)	4χ_C (.68)	$^3\chi_{CH}$ (.75)
K$_0$ (.95)	SIC$_2$ (.94)	$^3\chi^v{}_C$ (.67)	$^3\chi^b{}_{CH}$ (.75)
$^1\chi$ (.94)	CIC$_5$ (-.94)	$^5\chi_C$ (.64)	$^4\chi^b{}_{CH}$ (.74)
$^3\chi^b$ (.94)	CIC$_6$ (-.92)	$^6\chi_C$ (.64)	$^3\chi^v{}_{CH}$ (.72)
$^4\chi$ (.94)	SIC5 (.92)	$^3\chi_C$ (.61)	$^5\chi_{CH}$ (.71)
$^4\chi^b$ (.93)	SIC6 (.89)	$_6\chi^b{}_C$ (.60)	$^5\chi^v{}_{CH}$ (.67)
$^0\chi$ (.93)	CIC$_2$ (-.87)	$^5\chi^v{}_C$ (.60)	$^6\chi^b{}_{CH}$ (.47)

It is clear from the data in Table 1 that PC1 is strongly correlated with those indices which are related to size of chemicals. It is noteworthy that for the set of 3,692 chemicals PC_1 was also highly correlated (r = 0.81) with molecular weight. PC_2 may be interpreted as an axis of molecular complexity as encoded by the higher order information theoretic indices [27]. PC_3 is most highly related to the cluster/ path-cluster type molecular connectivity indices which quantify information regarding molecular branching. The data in Table 1 clearly show that PC_4 is strongly correlated with the cyclicity terms of the connectivity type.

2.2 A data set of 74 alkanes

For boiling point estimation lf alkanes, twenty six TIs, total surface area (TSA), and volume (V) were calculated for a set of 74 alkanes [36]. Table 2 below gives the three different two-parameter regression models for the prediction of boiling point.

Table 2: Results of three 2-parameter models in predicting the boiling points of 74 alkanes

Parameter	R^2	s. e.	F
$^1\chi$, CIC_2	99.4	3.72	5620
PC_1, PC_3	98.1	6.48	1828
V, TSA	97.0	8.17	1136

It appears from the data in Table 2 that the individual TIs, PCs derived from them as well as the calculated physical properties like volume and total surface area give good QSARs for this congeneric set of molecules. The TIs and PCs derived from them give a little bit superior models as compared to the properties.

2.3 A virtual library of 248,832 psoralen derivatives

A virtual library of 248,832 psoralen derivatives [23] was created and analyzed using PCs derived from TIs. For this study, a set of 92 topological indices was calculated by POLLY [30]. The set of TIs consisted of 37 topostructural and 55 topochemical indices. We define topostructural indices as those invariants which are derived from simple (unweighted) molecular graphs. Such graphs do not distinguish among different types of bonds or atoms. The Wiener index; cluster, path-cluster, and simple connectivity indices; and path length indices are examples of topostructural parameters. Topochemical indices, on the contrary, are indices defined on weighted molecular graphs such that the various types of atoms and bonds are weighted to reflect their nature and contribution to chemical bonding. The SIC, CIC, and IC indices as well as both bonding and valence connectivity indices are all examples of topochemical indices. For this data set, the top 3 PCs explained 89.2% of the variance in the data; first 6 PCs explained 95.5% of the variance of the original calculated indices. The PCs were used to cluster the large set of chemicals into a smaller subset as an exercise of managing combinatorial explosion that can happen in the drug design scenarios when one wants to create a large pool of derivatives of a lead compound. For details of the outcome of clustering of the 248,832 psoralen derivatives, please see [23].

For the large but congeneric set of 248,832 psoralen derivatives, first 6 PCs explained 95.5% of the variance of the original calculated indices.

2.4 A homogeneous set of aromatic amines

<Data description>

Table 3: Top 10 variables behind each of first 4 PCs (loadings in brackets) for 95 compound dataset

PC1	PC2	PC3	PC4
ΔHf (-0.20)	*ΔHf* (0.75)	E_{HOMO} (-0.52)	*fw* (-0.76)
DN^21_4 (-0.14)	$^9\chi$ (0.23)	*ΔHf* (-0.50)	E_{LUMO} (0.38)
I^W_D (-0.13)	$^{10}\chi$ (0.23)	E_{LUMO} (-0.36)	*SaaN* (-0.32)
Wt (-0.13)	E_{LUMO} (0.21)	$^{10}\chi$ (0.19)	E_{HOMO-1} (-0.27)
EDH (-0.13)	$^8\chi$ (0.20)	$^9\chi$ (0.16)	*SdO* (0.24)
$DS1_4$ (-0.13)	E_{HOMO} (0.18)	*fw* (0.15)	*nvx* (0.23)
$AS1_4$ (-0.12)	*SaaaC* (0.12)	$^8\chi$ (0.14)	*phia* (-0.19)
DN^2Z_4 (-0.12)	$^7\chi$ (0.11)	*SaaN* (-0.13)	μ (0.14)
AZS_2 (-0.12)	$SHBint_3$ (-0.09)	$^7\chi$ (0.12)	*ΔHf* (0.14)
ED (-0.12)	*XP8* (0.07)	μ (-0.12)	$^9\chi$ (-0.14)

For the 95 aromatic amine set, PC_1 is correlated with different original variables including some triplet indices and some indices related to molecular size. PC_2 is most strongly correlated with the heat of formation, ΔH*f* (r = 0.75); the energy of the highest occupied molecular orbital, E_{HOMO} (r = -0.52) and ΔH*f* (r = -0.50) are most highly corr4elated with PC_3 whereas PC_4 is strongly correlated with *fw (r* = -0.76) which is the molecular weight of the chemical species.

2.5 A diverse set of 508 chemicals

<Data description>

Table 4: Top 10 variables behind each of first 4 PCs (loadings in brackets) for 508 compound diverse dataset

PC1	PC2	PC3	PC4
DN^21_4 (-9.89×10^{-2})	*ΔHf* (-0.77)	*ΔHf* (-0.62)	E_{HOMO} (0.96)
I^W_D (-9.51×10^{-2})	$^{10}\chi$ (- 0.17)	$^8\chi$ (0.18)	$E_{HOMO\text{-}1}$ (-0.21)
Wt (-9.47×10^{-2})	$^9\chi$ (-0.17)	$^9\chi$ (0.18)	E_{LUMO} (-0.09)
EDH (-9.38×10^{-2})	$^8\chi$ (-0.16)	$^{10}\chi$ (0.17)	E_{LUMO+1} (-0.07)
$AS1_4$ (-8.94×10^{-2})	$^7\chi$ (-0.15)	$^7\chi$ (0.17)	*ΔHf* (-0.04)
IdC (-8.59×10^{-2})	$^6\chi$ (-0.12)	$^6\chi$ (0.14)	μ (0.03)
ED (-8.52×10^{-2})	E_{LUMO+1} (0.12)	E_{LUMO} (-0.13)	$^5\chi$ (-0.03)
fw (-8.52×10^{-2})	E_{LUMO} (0.12)	$^5\chi$ (0.13)	DSZ_2 (0.03)
W (-8.43×10^{-2})	$^5\chi$ (-0.11)	$^6\chi_{PC}$ (0.12)	$^6\chi$ (-0.03)
$AN1_4$ (-8.37×10^{-2})	$^6\chi_{PC}$ (-0.09)	E_{LUMO+1} (-0.12)	$^4\chi$ (-0.03)

For the diverse 508 chemical mutagen set, the energy of the highest occupied molecular orbital, E_{HOMO} (r = .96) is most strongly correlated with PC_4; PC_3 is highly correlated with ΔHf (r = -.62) which is also negatively correlated with PC_2 (r = -.77); PC_1 is loaded with some triplet indices and those invariants which reflect molecular size.

3. Materials and Methods

For materials and methods of data collection and
statistical analyses, see [19-21, 23, 27, 30-35].

4. Conclusions

In this paper, we reviewed our over three decades of research on the use of topological indices and principal components analysis in the characterization of five data sets: 1) A large and structurally diverse set of 3,692 industrial chemicals which was a subset of the Toxic Substances Control Act (TSCA) Inventory of the United States Environmental Protection Agency (USEPA), 2) A data set of 74 alkanes, 3) A virtual library of 248,832 psoralen derivatives, 4) A congeneric set of 95 aromatic and heteroaromatic amine mutagens, and 5) A structurally diverse collection of 508 chemicals mutagens.

The results show that the PCs derived from the TIs can be used for the development of QSARs as exemplified in Table 2 with 74 alkanes. PCs derived from TI have been used in the clustering of large set of psoralens [23]. Basak et al also used both PCs and individual TIs for analog selection [24] and characterization of isospectral graphs [17]. The data presented here show the usefulness of TIs and PCs derived from them in the clustering/ characterization of chemical libraries as well as QSAR. Details of QMSA analyses using PCs derived from TIs are not given here for brevity.

Basak [37] recently noted: “Mathematical chemistry or more accurately discrete mathematical chemistry had a tremendous growth spurt in the second half of the twentieth century and the same trend is continuing now. This growth was fueled primarily by two major factors: 1) Novel applications of discrete mathematical concepts to chemical and biological systems, and 2) Availability of high speed computers and associated software whereby *hypothesis driven* as well as *discovery oriented* research on large data sets could be carried out in a timely manner. This led to the development of not only a plethora of new concepts, but also to various useful applications to such important areas as drug discovery, protection of human as well as ecological health, and chemoinformatics. Following the completion of the Human Genome Project in 2003, discrete mathematical methods were applied to the “omics” data to develop descriptors relevant to bioinformatics, toxicoinformatics, and computational biology.”

Initially, TIs were used for the discrimination of structure and QSAR studies of congeneric and small sets of structures. For example, Randic’s [9] first order connectivity index (1χ), the information theoretic indices developed by Bonchev and Trinajstić [38] and those developed by Raychaudhury et al [7] were used to discriminate the set of alkanes and they worked well in those cases. In the case of 18 octanes, the molecules do not vary from each other with respect to size, but primarily in terms of branching patterns. Therefore, the indices [7, 9, 38] were interpreted based on the data as reflecting molecular branching. But when PCA was carried out with a diverse set of 3,692 chemical structures, the results entered an uncharted territory and were counterintuitive, to say the least. As shown from the correlation of the original variables with PC_1, 1χ and related indices were now strongly correlated with molecular size in the large and diverse set, not to molecular branching. PC_3 emerged as the axis representing branching and was strongly correlated to the cluster type molecular connectivity indices. Further studies with both congeneric and diverse data sets are need to understand the utility of TIs and PCs derived from them in structure-activity analyses.

Author Contributions

Since the early 1970sn Subhash C. Basak has been involved in the development of novel topological indices and their applications in QSARs. Gregory D. Grunwald has been a collaborator of Basak since the early 1980s in the use of TIs in QSAR and QMSA analyses. Subhabrata Majumdar has been involved in the applications of calculated mathematical chemodescriptors in QSAR in collaboration with Basak and Grunwald for the past five years and the standardization of novel methods like Interrelated Two-way Clustering (ITC) for use in QSAR.

Conflicts of Interest

"The authors declare no conflict of interest".

References and Notes

1. Wiener, H. 1947. Structural determination of paraffin boiling points. Journal of the American Chemical Society, 1947, 69, 17-20.

2. Sylvester, J.J. On an application of the new atomic theory to the graphical representation of the invariants and covariants of binary quantics, with three appendices. American Journal of Mathematics, 1878, 1, 64-125.

3. Kier, L.B.; Hall, L. Molecular Structure Description: The Electrotopological State; Academic Press: San Diego, CA, 1999, pp. 245.

4. Trinajstić, N. Chemical Graph Theory, 2nd ed.; CRC Press: Boca Raton, FL, 1992, pp. 352

5. Bonchev, D. Information Theoretic Indices for Characterization of Chemical Structures; Research studies Press: Chichester, U.K.; 1983.

6. Basak, S.C. Use of molecular complexity indices in predictive pharmacology and toxicology: A QSAR approach. Med. Sci. Res., 1987, 15, 605-609.

7. Raychaudhury, C.; Ray, S.K.; Ghosh, J.J.; Roy, A.B.; Basak, S.C. Discrimination of isomeric structures using information-theoretic topological indices. J. Comput. Chem., 1984, 5, 581-588

8. Hosoya, H. Topological Index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. Bull. Chem. Soc. Jpn., 1971, 44, 2332-2339.

9. Randic, M., Characterization of molecular branching. J. Am. Chem. Soc., 1975, 97, 6609-6615.

10. Balaban, A. T. "Distance Connectivity Index." Chem. Phys. Lett., 1982, 89, 399-404.

11. Basak, S. C., Restrepo, G. and Villaveces, J. L., Eds, Advances in Mathematical Chemistry and Applications, volume 1 & 2 , Bentham eBooks, Bentham Science Publishers and Elsevier, , 2015.

12. Devillers, J.; Balaban, A.T., Eds. Topological Indices and Related Descriptors in QSAR and QSPR; Gordon and Breach: Amsterdam, 1999, pp. 811.

13. Karelson, M., Molecular Descriptors in QSAR/QSPR, Wiley-Interscience, New York, 2000.

14. Todeschini, R.; Consonni, V., Molecular Descriptors for Chemoinformatics, Wiley-VCH, New York, 2009.

15. Gonzalez-Diaz, H.; Munteanu, C. R. (Editors), Topological Indices for Medicinal Chemistry, Biology, Parasitology, Neurological and Social Newworks, Transworld Research Neywork, 2011.

16. Janežič, D.; Miličević, A.; Nikolić, S. & Trinajstić, N.: 2007, Graph theoretic matrices in chemistry, Kragujevac: University of Kragujevac.

17. Balasubramanian, K.; Basak, S.C., Characterization of isospectral graphs using graph invariants and derived orthogonal parameters', Journal of Chemical Information and Computer Sciences, 1998, 38, 367-73.

18. Nandy, A.; Harle, M. & Basak, S.C., Mathematical descriptors of DNA sequences: Development and application, Arkivoc, 2006, 9, 211-38.

19. Basak, S. C. Philosophy of Mathematical Chemistry: A Personal Perspective, HYLE--International Journal for Philosophy of Chemistry, 2013, 19, 3-17.

20. Basak, S. C., Mathematical Descriptors for the Prediction of Property, Bioactivity, and Toxicity of Chemicals from their Structure: A Chemical-Cum-Biochemical Approach, Current Computer-Aided Drug Design, 2013, 9, 449-462.

21. Basak, S. C.; Magnuson, V. R.; Niemi, G. J.; Regal, R. R., Determining structural similarity of chemicals using graph-theoretic indices. Discrete Appl. Math. 1988, 19, 17–44.

22. Lajiness M (1990) Molecular similarity-based methods for selecting compounds for screening. In *Computational Chemical Graph Theory*, Ed. Rouvray DH (Nova, New York), pp 299-316.

23. Basak, S. C.; Mills, D.; Gute, B. D.; Balaban, A. T.; Basak, K.; Grunwald, G. D. Use of Mathematical Structural Invariants in Analyzing, Combinatorial Libraries: A Case Study with psoralen Derivatives, Current Computer-Aided Drug Design, 6, 240-251, 2010.

24. Basak, S. C. Molecular Similarity and Hazard Assessment of Chemicals: A Comparative Study of Arbitrary and Tailored Similarity Spaces, J. Eng. Sci. Manage. Educ. 7, 178-184, 2014

25. Balaban, A. T.; Mills, D.; Ivanciuc, O.; Basak, S. C. Reverse Wiener indices, Croat. Chim. Acta, 73, 923–941 (2000).

26. Nikolic, S.; Trinajstic, N.; Amic, D.; Beslo, D.; Basak S. C. Modeling the solubility of aliphatic alcohols in water. Graph connectivity indices versus line graph connectivity indices, In QSAR/QSPR Studies by Molecular Descriptors, M.V. Diudea, Ed., Nova Science Publishers, Huntington, New York, USA, pp. 63–81 (2001).

27. Basak, S. C. Information theoretic indices of neighborhood complexity and their applications, In Topological Indices and Related Descriptors in QSAR and QSPR, J. Devillers and A.T. Balaban, Eds., Gordon and Breach Science Publishers, The Netherlands, pp. 563–593 (1999).

28. Randic, M.; Vracko, M.; Nandy, A.; Basak, S. C. On 3-D graphical representation of DNA primary sequences and their numerical characterization, J. Chem. Inf. Comput. Sci., 40, 1235–1244 (2000).

29. Basak, S. C.; Gute, B. D. Mathematical descriptors of proteomics maps: Background and applications, Curr. Opin. Drug Discov. Devel., 11, 320-326 (2008).

30. Basak, S. C.; Harriss, D. K.; Magnuson, V. R. 1988. POLLY v. 2.3: 1988; Copyright of the University of Minnesota.

31. Basak, S. C.; Magnuson, V. R.; Niemi, G. J.; Regal, R. R.; Veith, G. D. Topological indices: their nature, mutual relatedness, and applications, Mathematical Modelling, 1987 8, 300–305..

32. MolConnZ, Version 4.05, 2003; Hall Ass. Consult.; Quincy, MA.

33. Basak, S.C.; Grunwald, G.D.; Balaban, A.T.TRIPLET: Copyright of the Regents of the University of Minnesota, 1993

34. Filip, P. A.; Balaban,T. S.; Balaban, A. T. A new approach for devising local graph invariants: derived topological indices with low degeneracy and good correlation ability. J. Math. Chem. 1987, 1, 61-83.

35. Basak, S. C.; Grunwald, G. D., APProbe. 1993; Copyright of the University of Minnesota

36. Basak, S. C.; Grunwald, G. D., A COMPARATIVE STUDY OF GRAPH INVARIANTS, TOTAL SURFACE AREA AND VOLUME IN PREDICTING BOILING POINTS OF ALKANES, Math Modelling & Sci. Computing, 1993, 2, 735-740.
37. Basak, S. C. Mathematical Structural Descriptors of Molecules and Biomolecules: Background and Applications, In, Advances in Mathematical Chemistry and Applications, volume 1, Basak, S. C., Restrepo, G. and Villaveces, J. L., Eds., pp. 3-23, Bentham eBooks, Bentham Science Publishers and Elsevier, , 2015.
38. Bonchev, D. & Trinajstić, N.: 1977, 'Information theory, distance matrix and molecular branching', Journal of Chemical Physics, 67, 4517-33.

SciForum
Mol2Net

Hierarchical Quantitative Structure-Activity Relationships (HiQSARs) for the Prediction of Physicochemical and Toxicological Properties of Chemicals Using Computed Molecular Descriptors

Subhash C. Basak[1*] and Subhabrata Majumdar[2]

[1] University of Minnesota Duluth-Natural Resources Research Institute (UMD-NRRI) and Department of Chemistry and Biochemistry, University of Minnesota Duluth, 5013 Miller Trunk Highway, Duluth, MN 55811, USA; sbasak@nrri.umn.edu

[2] School of Statistics University of Minnesota Twin Cities Minneapolis, MN 55414

* Author to whom correspondence should be addressed; E-Mail: sbasak@nrri.umn.edu; Tel.: +1-218-727-1335

Published: 4 December 2015

Abstract: Attempts have been made to formulate quantitative structure=activity relationships (QSARs) for the prediction of property/ bioactivity of chemicals from their experimental test data as well as properties that can be computed directly from molecular structure without the input of any other experimental property. Because both in drug design and hazard assessment of chemical scenarios relevant experimental data for property/ bioactivity estimation are not available for the majority of candidate chemicals, QSARs based on computed molecular descriptors are emerging as methods of choice for property/ bioactivity estimation in many cases. Numerical graph invariants or topological indices, viz., topostructural (TS) indices, topochemical (TC) indices, as well as three-dimensional (3-D) descriptors, and quantum chemical (QC) indices have been used for QSAR formulation based on computed descriptors. In the 1990s, Basak et al formulated the concept of hierarchical quantitative structure=activity relationships (HiQSAR) in which TS, TC, 3-D, and QC descriptors were used in a graduated manner, the more computationally demanding descriptors being used only if the simpler ones did not give acceptable QSAR models. Our experience with a substantial number of HiQSARs for physical, pharmacological, and toxicological properties of different congeneric as well diverse sets chemicals indicate that the combinations of TS + TC descriptors are capable of giving good quality QSARs in most situations. The addition of 3-D or QC descriptors make marginal or no improvement in model quality after the use of TS+ TC descriptors. At this age of "big data screening and analysis" this is a good news because QSARs derived from the less expensive and practically useful TS+ TC combination can be effective tools in the screening of large chemical libraries.

Keywords: keyword; keyword; keyword (3–10 keywords separated by semi colons)
Quantitative structure-activity relationship (QSAR); Hierarchical quantitative structure-activity relationship (HiQSAR); Topostructural (TS) indices; Topochemical (TC) indices; Three-dimensional (3-D) descriptors; Quantum chemical (QC) indices; Vapor pressure; Mutagenicity; Ridge Regression; Interrelated two-way clustering (ITC).

1. Introduction

A contemporary trend in quantitative structure-activity/ property relationship (QSAR/ QSPR) studies is the use of properties which can be computed from structure without the input of any other data [1-7]. The underlying reason for this is that for the majority of candidate chemicals that need to be screened for both new drug discovery and hazard assessment of environmental pollutants, experimental properties needed for QSAR formulation are not available [4-7]. Table 1 gives a partial list of physical and bi9ochemical/ toxicological properties needed for the prediction of bioactivity/ toxicity of chemicals. In the realm of hazard assessment of industrial chemicals currently listed in the Toxic Substances Control Act (TSCA) Inventory of the United States Environmental Protection Agency (USEPA), Auer et al [8] reported that for most of the chemicals under investigation the majority of the properties needed for hazard estimation were not available. Over the years, after the publication of this summary by Auer et al [8] in 1990, the availability of good quality experimental data needed for the risk assessment of chemicals has probably became worse with time. Therefore, quantitative structure-activity/ property relationships (QSAR/ QSPR) remain one important source of property data itemized in Table 1. Because property-property relationships (PPRs) are not practical in many situations arising out of the paucity of the predictor property, QSARs derived from computed molecular descriptors have emerged as useful tools in the screening of chemicals.

Over the years Basak and coworkers have used different combinations of topostructural (TS) indices, topochemical (TC) indices, 3-D descriptors as well as and quantum chemical (QC) indices for QSAR formulation in a hierarchical manner (Figure 1). In the hierarchical QSAR (HiQSAR) approach [4-7], TS, TC, geometrical, and quantum chemical descriptors are used for model building in a graduated manner, the latter and more complex levels being used when the earlier ones fail to give reasonable QSAR. Basak et al [4-7] divided the topological indices (TIs) into two major groups: Topostructural (TS) indices and topochemical (TC) indices. TS descriptors are indices which are calculated from skeletal graph models of molecules that do not distinguish among different types of atoms in a molecule or the various types of chemical bonds, e.g.; single bond, double bond, triplet bond, etc. Thus, TS descriptors quantify information regarding the connectivity, adjacency, and distances between vertices of molecular graphs, ignoring their distinct chemical nature. TC indices, on the other hand, are sensitive to both the pattern of connectedness of the vertices (atoms), as well as their chemical/bonding characteristics. Therefore, the TC indices are more complex than the TS descriptors. Figure 1 represents the full hierarchical scheme of QSAR formulation involving different levels of chemodescriptors and biodescriptors, the latter being derived from the omics data. Basak et al used various combination of TS, TC, 3-D, and QC indices for the development of QSAR over the years. For a review please see recent references [1, 4-7]. The indices used by Basak and coworkers have been calculated by the software POLLY [9], MolConnZ [10], APProbe [11], Triplet [12, 13], MOPAC [14], and Gaussian [15].

In this paper we discuss our HiQSAR approach for two sets of chemicals: Vapor pressure of a set 476 diverse molecules and Ames' mutagenicity of a heterogeneous group of 508 chemicals.

.

2. Results and Discussion

.

2.1 QSAR for vapor pressure) for a diverse set of 476 chemicals.

The HiQSAR results provided in Table 2 show that of all classes of molecular descriptors the TC class of indices gave the most effective models. The TS+TC combination makes some improvement in model quality over the TC only QSAR. The model developed using all indices which consisted of (TS+ TC+ 3-D) combination plus dipole moment calculated by Sybyl [18] as well as a hydrogen bonding descriptor HB_1 [19, 20] could not outperform the model derived from the TS+TC combination. For details for this analysis, see [17].

2.2 HiQSAR modeling of a diverse set of 508 chemical mutagens

TS, TC, 3D, and QC descriptors for 508 chemical were calculat4ed and QSARs were formulated hierarchically using the four types of descriptors. For details of calculations and model building, see ref. [7]. The method Interrelated two way clustering, ITC [21], was used for variable selection. Table 3 gives results of ridge regression (RR) alone as well as those where RR was used on descriptors selected by ITC. For both RR only and ITC+ RR analysis the TS + TC combination gave the best models for predicting mutagenicity of the 508 diverse chemicals. The addition of 3-D and QC descriptors to the set of independent variables made minimum or no improvement in model quality.

.

.

Table 1. Important properties needed for evaluation of chemicals

Physicochemical	Pharmacological / Toxicological
Molar volume	Macromolecule level
Boiling point	: Receptor binding
Melting point	: Michaels constant (Km)
Vapor pressure	: Inhibition constant (Ki)
Water solubility	: DNA alkylation
Dissociation constant (pK$_a$)	: Unscheduled DNA synthesis
Partition coefficient	Cell level
: Octanol-water (log P)	: Salmonella mutagenicity
: Air-water	: Mammalian cell transformation
: Sediment-water	Organism level (acute)
Reactivity (electrophile)	: Algae
	: Invertebrates
	: Fish
	: Birds
	: Mammals
	Organism level (chronic)
	: Bioconcentraton factor
	: Biodegradation
	: Carcinogenicity
	: Reproductive toxicity
	: Delayed neurotoxicity

Table 2: Summary of the Regression Results for the Training Set and the Prediction Results for the Test Set for the Hierarchical Analysis of log VP

	Training Set (342)			Test Set (134)		
Parameter Class	F	R^2	S	F	R^2	S
Topostructural (TS)	104.6	48.1	0.56		57.9	0.46
Topochemical (TC)	126.3	79.2	0.36		85.8	0.27
Geometrical	168.9	51.8	0.53		62.2	0.44
TS+ TC	112.5	80.4	0.35		84.7	0.28
All Indices	117.4	79.6	0.35		84.2	0.28

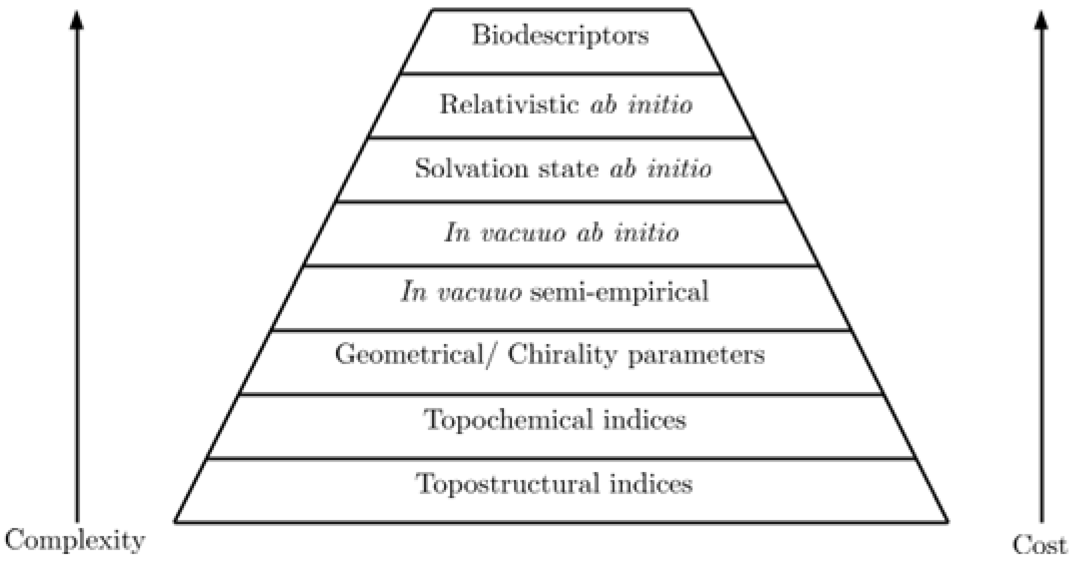
Table 3. HiQSAR model (RR and ITC+RR) for a diverse set of 508 chemical mutagens

Model type	Predictor Type	Predictor Number	% Correct classification	Sensitivity	Specificity
RR	TS	103	53.14	52.34	53.97
	TS+TC	298	76.97	83.98	69.84
	TS+TC+3D+QC	307	77.17	84.38	69.84
ITC+ RR	TS	103	66.34	73.83	58.73
	TS+TC	298	73.23	77.34	69.05
	TS+TC+3D	301	74.80	77.34	72.22
	TS+TC+3D+QC	307	72.05	76.17	67.86

Table 4. Major chemical classes (not mutually exclusive) within the 508 mutagen/non-mutagen database.

Chemical class	Number of compounds
Aliphatic alkanes, alkenes, alkynes	124
Monocyclic compounds	260
Monocyclic carbocycles	186
Monocyclic heterocycles	74
Polycyclic compounds	192
Polycyclic carbocycles	119
Polycyclic heterocycles	73
Nitro compounds	47
Nitroso compounds	30
Alkyl halides	55
Alcohols, thiols	93
Ethers, sulfides	38
Ketones, ketenes, imines, quinones	39
Carboxylic acids, peroxy acids	34
Esters, lactones	34
Amides, imides, lactams	36
Carbamates, ureas, thioureas, guanidines	41
Amines, hydroxylamines	143
Hydrazines, hydrazides, hydrazones, traizines	55
Oxygenated sulfur and phosphorus	53
Epoxides, peroxides, aziridines	25

Figure 1. Hierarchical QSAR development scheme involving different levels of chemodescriptors and biodescriptors, the latter being derived from the omics sciences



3. Materials and Methods

3.1 QSAR for vapor pressure) for a diverse set of 476 chemicals.

Measured vapor pressure (VP) values for 476 subset of the Toxic Substances Control Act (TSCA) Inventory were obtained from the ASTER (Assessment Tools for the Evaluation of Risk) database [16]. Due to the size of the dataset being used in this study, the VP data for these chemicals will not be listed in this paper. The set of 92 TIs was partitioned into 38 topostructural indices and 54 topochemical indices. For details of this study see [17]. Because the number of data points was reasonably large, the data was split into a training set (342 compounds) and a test set (134 compounds), an approximately 75/25 split. Models were developed using the training set of chemicals and then used to predict the VP values of the test chemicals, the results being shown in Table 2.

3.2 HiQSAR modeling of a diverse set of 508 chemical mutagens

The data were taken from the CRC Handbook of Identified Carcinogens and Non-carcinogens [22]. The response variable is Ames mutagenicity, the sample available being 508 compounds classified as not mutagenic (scored 0) or mutagenic (scored 1). The set of 508 is comprised of 256 mutagens and 252 non-mutagens. Table 4 gives an idea regarding the diversity of the chemicals in this database in terms of chemical types and functional groups. Ridge regression was used for model building because it is a sound method, in the rank deficient case in particular. For the ITC+RR modeling, ITC was first used for variable selection and then RR was employed for model building.

.

.

4. Conclusions

The objective of HiQSAR research reported in this paper was to study the relative effectiveness of topological (TS, TC), geometrical, and quantum chemical descriptors in the development of useful QSAR models. Results derived for two large data sets, viz. vapor pressure of a group of 476 diverse chemicals and a structurally diverse set of 508 mutagens, show that the computationally less expensive TS and TC descriptors give QSARs of reasonable quality. The addition of 3-D or QC descriptors after the use of TS+TC combination does not make any improvement in model quality. We previously observed this trend in different properties of other data sets [23-30]. At this age of "big data screening and analysis" [31], this is a good news because QSARs derived from the less expensive TS+ TC combination can be effective tools in the fast and effective screening of large chemical libraries. Further QSAR research is in progress to validate the broad applicability of the HiQSAR paradigm based on mathematical structural descriptors [32].

Acknowledgments

One of the authors (Basak) would like to thank Gregory D. Grunwald, Brian D. Gute, and Denise Mills for sustained technical support and collaboration in his research.

.

Author Contributions

Subhash C. Basak formulated the HiQSAR concept in the 1990s and applied it to QSAR of various congeneric and diverse sets of chemicals. Subhabrata Majumdar has been carrying out collaborative research with Basak in QSAR during the past five years.

Conflicts of Interest

"The authors declare no conflict of interest".

References and Notes

[1] Basak, S. C., Role of Mathematical Chemodescriptors and Proteomics-Based Biodescriptors in Drug Discovery, Drug Develop. Res., 2010, 72, 1-9.

[2] Kier, L.B.; Hall, L. Molecular Structure Description: The Electrotopological State; Academic Press: San Diego, CA, 1999.

[3] Devillers, J.; Balaban, A.T., Eds. Topological Indices and Related Descriptors in QSAR and QSPR; Gordon and Breach: Amsterdam, 1999.

[4] Basak, S. C., MATHEMATICAL STRUCTURAL DESCRIPTORS OF MOLECULES AND BIOMOLECULES: BACKGROUND AND APPLICATIONS, in Advances in Mathematical Chemistry and Applications, volume 1, pp. 3-23, Basak, S. C., Restrepo, G. and Villaveces, J. L., Editors, Bentham eBooks, Bentham Science Publishers, 2015.

[5] Basak, S. C., Gute, B. D., Grunwald, G. D., Relative effectiveness of topological, geometrical, and quantum chemical parameters in estimating mutagenicity of chemicals, , in Quantitative Structure-activity Relationships in Environmental Sciences VII, F. Chen and G. Schuurmann, Eds., SETAC Press, Pensacola, FL., pp. 245–261 (1998).

[6] Basak, S. C., Mathematical Descriptors for the Prediction of Property, Bioactivity, and Toxicity of Chemicals from their Structure: A Chemical-Cum-Biochemical Approach, Current Computer-Aided Drug Design, 2013, 9, 449-462.

[7] Basak, S. C.; Majumdar, S. Current landscape of hierarchical QSAR modeling and its applications: Some comments on the importance of mathematical descriptors as well as rigorous statistical methods of model building and validation, in Advances in Mathematical Chemistry and Applications, volume 1, pp. 251-281, Basak, S. C., Restrepo, G. and Villaveces, J. L., Editors, Bentham eBooks, Bentham Science Publishers, 2015.

[8] Auer, C. M.; Nabholz, J. V.; Baetcke, K. P. Mode of action and the assessment of chemical hazards in the presence of limited data: use of structure-activity relationships (SAR) under TSCA, Section 5. Environ. Health Perspect. 1990, 87,183–197.

[9] Basak, S. C.; Harriss, D. K.; Magnuson, V. R. 1988. POLLY v. 2.3: 1988; Copyright of the University of Minnesota.

[10] MolconnZ, Version 4.05, 2003; Hall Ass. Consult.; Quincy, MA..

[11] Basak, S. C.; Grunwald, G. D., APProbe. 1993; Copyright of the University of Minnesota.

[12] Filip, P. A.; Balaban, T. S..; Balaban, A. T. A new approach for devising local graph invariants: derived topological indices with low degeneracy and good correlation ability.1987, J. Math. Chem.1, 61-83.

[13] Basak, S.C.; Grunwald, G.D.; Balaban, A.T.TRIPLET: Copyright of the Regents of the University of Minnesota, 1993.

[14] Stewart, J.J.P. MOPAC Version 6.00, QCPE #455, Frank J Seiler Research Laboratory, US Air Force Academy, CO, 1990.

[15] Frisch, M. J. et al. Gaussian 98 (Revision A.11.2). 1998. Pittsburgh, PA, Gaussian, Inc.

[16] Russom, C. L.; Anderson, E. B.; Greenwood, B. E.; Pilli, A. ASTER: An Integration of the AQUIRE Data Base and the QSAR System for Use in Ecological Risk Assessments. Sci. Total Environ. 1991, 109/110, 667-670.

[17] Basak, S. C.; Gute, B. D.; Grunwald, G. D. Use of Topostructural, Topochemical, and Geometric Parameters in the Prediction of Vapor Pressure: A Hierarchical QSAR Approach, 1997, J. Chem. Inf. Comput. Sci., 37, 651-655.

[18] SYBYL Version 6.2; Tripos Associates, Inc.: St. Louis, MO, 1994.

[19] Basak, S. C. H-Bond; Copyright of the University of Minnesota, 1988.

[20] Gu, Y-C.; Cuyang, Y.; Lien, E.J. (1986). Examination of quantitative relationship of partition coefficient (log P) and molecular weight, dipole moment and hydrogen bond capability of miscellaneous compounds. 1986, J. Mol. Sci., 4, 89- 95.

[21] Tang, C.; Zhang, L.; Zhang, A.; Ramanathan, M. In: Interrelated Two-way Clustering: An Unsupervised Approach for Gene Expression Data Analysis, Proceedings of BIBE 2001: 2nd IEEE International Symposium on Bioinformatics and Bioengineering, Bethesda, Maryland, November 4-5, 2001; Bilof, R.; Palagi, L., Eds.; IEEE Computer Society: Los Alamitos, CA, 2001; pp. 41-48.

[22] Soderman, J.V. CRC Handbook of Identified Carcinogens and Noncarcinogens: Carcinogenicity-Mutagenicity Database, Boca Raton, Florida, 1982.

[23] Gute, B. D.; Basak, S. C. Predicting acute toxicity of benzene derivatives using theoretical molecular descriptors: a hierarchical QSAR approach, SAR QSAR Environ. Res., 1997, 7, 117–131.

[24] Gute, G. D.; Grunwald, G. D.; Basak, S. C. Prediction of the dermal penetration of polycyclic aromatic hydrocarbons (PAHs): A hierarchical QSAR approach, B.D. SAR QSAR Environ. Res., 1999, 10, 1–15.

[25] Basak, S. C.; Mills, D. R.; Balaban, A. T.; Gute, B. D. Prediction of mutagenicity of aromatic and heteroaromatic amines from structure: A hierarchical QSAR approach, , J. Chem. Inf. Comput. Sci., 2001, 41, 671–678.

[26] Hawkins, D. M.; Basak, S. C.; Mills, D. QSARs for chemical mutagens from structure: ridge regression fitting and diagnostics, Environ. Toxicol. Pharmacol., 2004, 16, 37–44.

[27] Gute, B. D.; Basak, S. C.; Balasubramanian, K.; Geiss, K.; Hawkins, D. M. Prediction of halocarbon toxicity from structure: A hierarchical QSAR approach, Environ. Toxicol. Pharmacol., 2004, 16, 121–129.

[28] Basak, S. C.; Natarajan, R.; Mills, D. Structure-activity relationships for mosquito repellent aminoamides using the hierarchical QSAR method based on calculated molecular descriptors, Conference proceedings, WSEAS Transactions on Information Science and Applications, 2005, 7, 958–963.

[29] Basak, S. C.; Mills, D.; Hawkins, D. M.; El-Masri, H. A. Prediction of tissue: air partition coefficients: A comparison of structure-based and property-based methods, SAR QSAR Environ. Res., 2002, 13, 649–665.

[30] Basak, S. C.; Mills, D.; Mumtaz, M. M.; Balasubramanian, K. Use of topological indices in predicting aryl hydrocarbon (Ah) receptor binding potency of dibenzofurans: A hierarchical QSAR approach. Indian. J. Chem., 2003, 42A, 1385–1391.

[31] Basak, S. C.; Bhattacharjee, A. K.; Vracko, M. Big Data and New Drug Discovery: Tackling "Big Data" for Virtual Screening of Large Compound Databases. Current Computer-Aided Drug Design, 2015, 11, 197-201.

[32] Basak, S. C. Philosophy of Mathematical Chemistry: A Personal Perspective, HYLE--International Journal for Philosophy of Chemistry, 2013, 19, 3-17.

SciForum
Mol2Net

Synthesis and Characterization of Shape Memory Polyurethanes

Míriam Sáenz-Pérez[1, 2]*, José Manuel Laza[1], Jorge García-Barrasa[2], Luis Manuel León[1]and José Luis Vilas[1]

[1] Macromolecular Chemistry Research Group. Dept. of Physical Chemistry. Faculty of Science and Technology. University of the Basque Country (UPV/EHU), Leioa 48940, Spain. E-Mail: josemanuel.laza@ehu.es (J.M.L.), luismanuel.leon@ehu.eus (L.M.L.) and joseluis.vilas@ehu.eus (J.L.V.)

[2] The Footwear Technology Center of La Rioja, Calle Raposal 65, Arnedo 26580, Spain; E-Mail: jgarcia@ctcr.es (J.G.)

* M. Sáenz-Pérez; E-Mail: msaenz@ctcr.es; Tel.: +34 94 601 5534.

Published: 4 December 2015

Abstract: Shape memory polymers (SMPs) have attracted extensive attention from basic and fundamental research to industrial and practical applications because they have emerged as a cheap and efficient alternative to well-known metallic shape-memory alloys. Among them, shape memory polyurethanes (SMPUs) own different applications such as the textile finishing, adhesives, coatings, automotive, furniture, construction, and thermal insulation and footwear industries, due to it can be synthesized with different types of molecular architectures by manipulating their composition and choosing properly the chemical structure of their components. In this work, the synthesis and characterization of shape memory polyurethanes, based on two-step polymerization, is reported. The hard segment of SMPU was composed of diisocyanate and a chain extender. On the other hand, the soft segment was prepared by polyols with different molecular weights. Depending on the structure of the synthetized polyurethanes, the materials presented different properties. Thermal characterization was performed by means of Differential Scanning Calorimetry (DSC) and Thermogravimetric Analysis (TGA). Furthermore, mechanical properties and shape memory effect were also determined by Dynamic Mechanical Analysis (DMA) and Thermo-Mechanical Analysis (TMA).

Keywords: shape memory behavior; polyurethane; mechanical properties; thermal properties.

1. Introduction

Shape memory polymers (SMPs) present the ability of modifying their shape in a predefined manner in response to externally imposed stimuli. A deformation-induced temporary shape is transformed to an initial equilibrium configuration defined by the chemical or physical crosslinkings within the polymer. SMPs are cheap, easy processing, light, could be deformed to high strains and high recovery ratios are achieved, in contrast to shape memory alloys. Up to date, light, thermal, electrical, magnetic stimuli have been mainly used for trigger the shape-memory effect in polymeric materials[1,2].

Usually, shape memory process consists of two different phases known as "programming" and "recovery". During the programming, the material is deformed above the softening temperature, where the conformational entropy of the material is decreased. Subsequently the material is cooled to temperatures below the segmental transition under constrained conditions to reach the "temporary shape" owing to the reduced molecular mobility at temperatures below transition temperature (T_{trans})[3,4]. Finally, the material recovers its initial permanent "fixed shape" upon the application of the external stimulus which triggers the shape memory effect (Figure 1). Depending on their microstructure and chemical nature several classes of SMPs could be found. Among all the available types of materials showing a shape-memory effect, shape memory polyurethanes (SMPUs) have shown suitable physico-mechanical properties to be used in applications as varied as stents, microactuators and wrinkle free fabrics among others[5,6].

Though extensive work has been devoted in developing SMPUs, little attention has been paid in understanding how the chain microstructure affects the shape-memory behavior and mechanical properties of the resulting material. In this framework, this work deals with the synthesis and physic-mechanical characterization of toluene 2,4-diisocyanate (TDI) based SMPUs. Transition temperatures have been determined by Differential Scanning Calorimetry (DSC) and Dynamic Mechanical Thermal Analysis (DMTA). Thermomechanical programming experiments were carried out to examine the shape-memory effect of developed materials. In overall, results reveal a marked influence of the soft-hard segments over the transition temperature and shape-memory effect of TDI-based polyurethanes.

2. Results and Discussion

2.1. Thermogravimetric Analysis (TGA)

The representative TGA curves obtained from SMPUs with different molar ratio are shown in Figure 2 and the initial decomposition temperatures, T_i, are listed in Table 1.

It is generally accepted that the thermal degradation process in polyurethanes is a two-stage or three-stage decomposition, which mainly depends on the chemical structure, and the composition of polyols, diisocyanates, and chain extenders[7,8]. The three synthetized polyurethanes display typical two-stage degradation. The first one may be attributed to the PU hard segments, whereas in the second stage the degradation is caused by the soft segments. The obtained results show good thermal stability for the three samples, indicating that the choice both the diisocyanate and the butanodiol influence on the thermal decomposition of polyurethanes.

On the other hand, the residue of the SMPUs increased as the molar relation between diisocyanate and butanodiol increase. This may be due in part to favorable interactions between the hard domain interface and the liquid crystalline phase[9,10].

Samples	T_i (ºC)	$T_{g,\ DSC}$ (ºC)	$T_{g,\ DMTA}$ (ºC)
SMPU-3.5	279	-1.2	44.6
SMPU-4.5	275	19.8	58.1
SMPU-5.5	285	29.5	62.9

Table 1. Thermal properties of synthetized polyurethanes.

Figure 1. Schematic representation of the mechanism of the shape memory effect.

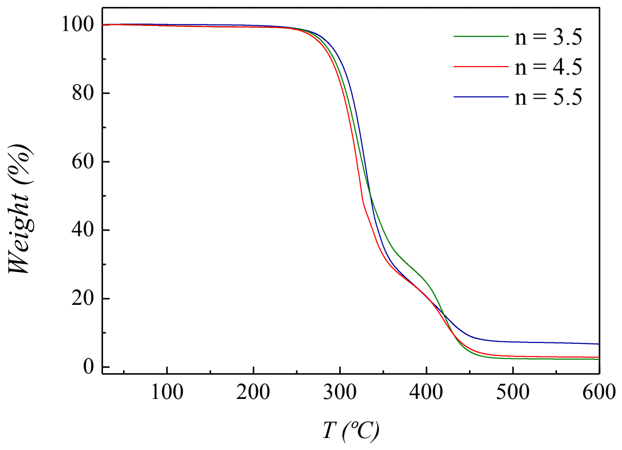


Figure 2. TGA curves for synthetized polyurethanes.

2.2. Differential Scanning Calorimetry (DSC)

The DSC scans of the SMPUs are shown in Figure 3 and the measured glass transition temperatures, $T_{g,\ DSC,}$ are summarized in Table 1. DSC results show that as increase the hard segment content (higher n), higher is the glass transition temperature[11,12]. This suggests that hard segment interaction among polymeric chains were strengthened at low hard segment content, due to these hard segments can achieve a more well-oriented position within the polymeric structure.

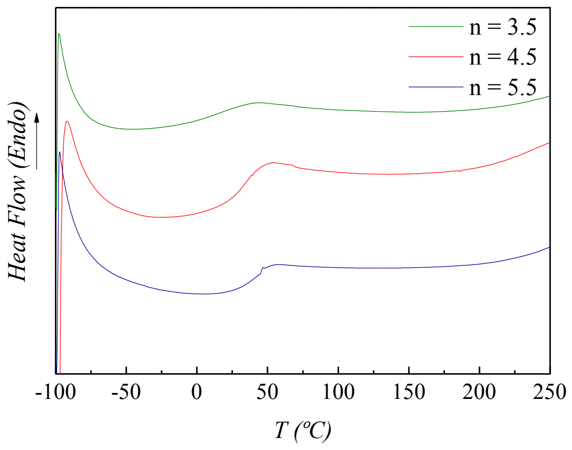


Figure 3. DSC curves for synthetized polyurethanes.

2.3. Dynamic Mechanical Thermal Analysis (DMTA)

DMTA analysis is a sensitive method to study the thermomechanical behavior of polymers. The major peak in the loss factor, *tanδ*, is usually used to designate T_g. This peak corresponds with a sharp drop in the storage modulus (*E'*), so this is also an important parameter in order to evaluate the structure-property relationship of these materials.

Figure 4 shows the variation of the storage modulus (*E'*) and the loss factor, *tanδ*, as a function of the temperature for the synthetized polyurethanes, and in Table 1 are showed the results, $T_{g,DMTA}$. The transition around this temperature was ascribed to the glass transition of the hard segment phase[13,14]. Therefore, as increase the hard segment content higher is the glass transition temperature, which is in agreement with the DSC results.

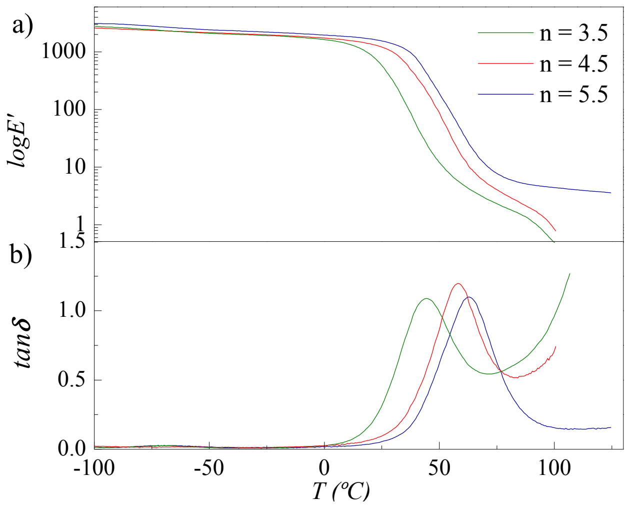


Figure 4. DMTA curves for synthetized polyurethanes: a) storage modulus (*logE'*); and b) loss factor (*tanδ*).

2.4. Shape Memory Behavior

Shape memory effect can be measured qualitatively and quantitatively. In Figure 5, it can be observed that synthetized polyurethanes have shape memory effect. Regarding thermally-activated shape-memory properties, soft segments will be responsible for shape fixity, acting as the switching segments, while hard segments will be responsible for shape recovery, determining the permanent shape[15].

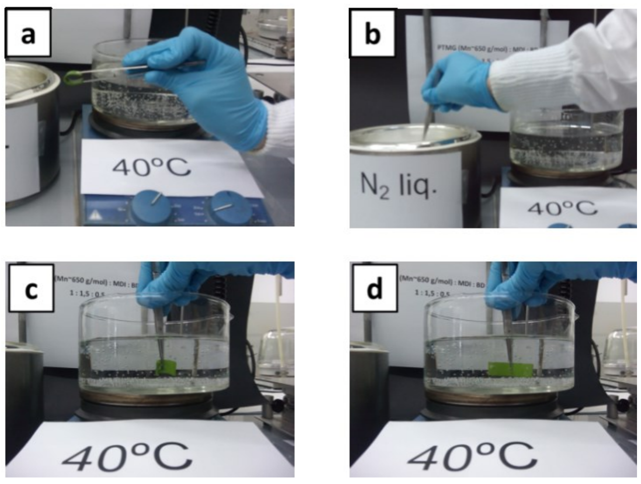


Figure 5. Images of behavior of SMPU. a) Deformation, b) Shape Fixation, c) Recovery, d) Original shape.

The thermally activated shape memory behavior of SMPUs is shown in Figure 6. Table 2 summarized the values obtained of fixity ratio (R_f) and shape recovery ratio (R_r) which were calculated employing Eqs 1 and 2. The shape memory behavior of SMPUs represents almost complete strain fixing (more than 87%) or recovery (99%).

Table 2. Shape fixing and recovery efficiencies of SMPUs.

Samples	R_f (%)	R_r (%)
SMPU-3.5	89.9	99.8
SMPU-4.5	98.9	99.7
SMPU-5.5	87.2	100

3. Materials and Methods

3.1. Materials

Polytetramethylene glycol (PTMG, M_n = 650 g·mol^{-1}) used as polyol, and toluene 2,4-diisocyanate (TDI), were purchased by Sigma Aldrich and were used as received. Moreover, 1,4-butanediol (BD, Sigma Aldrich), chain extender, was dried under vaccum for 3 h at 65ºC before use.

3.2. Synthesis of shape memory polyurethanes

All the SMPUs were synthesized by a two-step method varying the hard-segment content. The SMPUs were prepared by reaction of stoichiometric amount of polyol / diisocyanate / chain extender with block ratios of 1:n+1:n (where *n* is between 3.5 and 5.5), view Table 3. The reaction scheme for the synthesis is shown according to the route outlined in Figure 7.

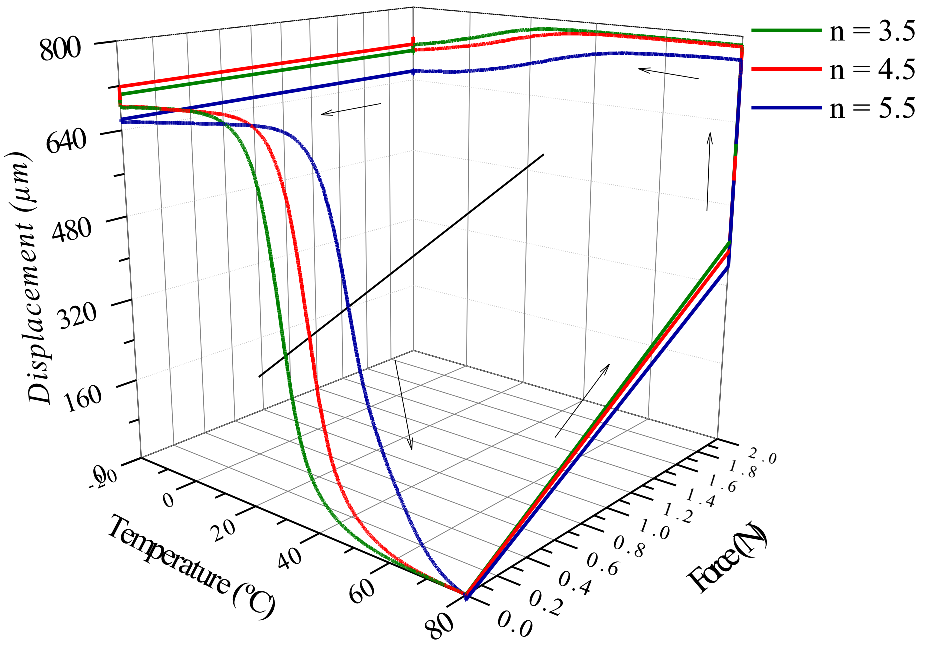


Figure 6. Three-dimensional thermomechanical cycle for synthetized polyurethanes.

Table 3. Polyurethane composition.

Samples	Composition of polyurethane		
	PTMG	TDI	BD
SMPU-3.5	1	4.5	3.5
SMPU-4.5	1	5.5	4.5
SMPU-5.5	1	6.5	5.5

On the one hand, the hard segments of SMPU were composed of toluene 2,4-diisocyanate (TDI) and a chain extender, 1,4-butanediol (BD). On the other hand, the soft segment was prepared by polytetramethylene glycol (PTMG), M_n = 650 g·mol^{-1}.

The synthesis was carried out in a 150 mL 5-neck round-bottom flask heated at 70ºC in an oil bath, equipped with a mechanical stirrer and a nitrogen inlet. In the first step, the polyol, polytetramethylene glycol, was added into the dry reactor. After 30 min with nitrogen atmosphere, TDI was added dropwise. The reaction continued at 70ºC for 2h to obtain -NCO terminated prepolymer, under a vigorous flow of nitrogen to prevent the reaction of the isocyanate groups with air moisture. In the second step, the chain extender, BD, was added dropwise into the reaction system. The reaction mixture was continuously stirred during approximately 2 minutes until a significant increase in viscosity was detected. The viscous mixture was poured into a preheated stainless steel mold and placed in a hydraulic press under pressure at 100ºC overnight to obtain the final polymer.

3.3. Characterization techniques

Thermal properties of all samples were measured by Differential Scanning Calorimetry (DSC 822e from Mettler Toledo) to identify thermal actuation temperatures. The transition temperature of shape memory effect (T_{trans}) was defined from glass transition temperature measured in the second heating cycle ($T_{g,\ DSC}$ in Table 1). Samples in aluminium pans were characterized under constant nitrogen flow (50 mL·min^{-1}). First, the PU samples were equilibrated at -100°C, and then heated at a rate of 10ºC·min^{-1} from -100 to 250°C. In this first cycle, the thermal history of the sample was erased. It was then cooled down to -100°C at a cooling rate of 10ºC·min^{-1}. Subsequently, a second heating scan to 250ºC was conducted at the same heating rate. In all cases samples around 10–15 mg were used.

Thermal degradation behaviour was studied by Thermogravimetric Analysis (TGA METTLER TOLEDO 822e) in alumina pans under nitrogen atmosphere by heating the samples (10-15 mg) from room temperature to 800°C at 10°C·min^{-1}.

Dynamic Mechanical Thermal Analysis (DMTA) was performed on a Mettler-Toledo DMA1 analyzer in tensile mode. 1.5 mm thick, 6 mm wide and 10 mm long specimens were used. Curves displaying storage modulus (E') and the loss factor ($tan\delta$) were recorded in the range of −100 to 150°C at a heating rate of 3°C·min^{-1} and at a deformation frequency of 10 Hz and displacement of 20μm, which is found within the Linear Viscoelastic Region (LVR) of synthetized SMPUs.

For Thermo-Mechanical Analysis (TMA) measurements, samples were conducted on a Mettler Toledo DMA1 in the temperature range of -20–80ºC at a heating rate of 4ºC·min^{-1}. Rectangular samples of about 10 mm x 6 mm x 1.5 mm were used in shape memory performances. First, the sample is heated to programming temperature T_{prog} (80ºC) and deformed applying 2 N force. Once the sample has been stretched, ε_m, the next stage is to cool it to below transition temperature T_{low} (-20ºC) in order to fix the temporary shape. Once the sample is unloaded, the deformation of the sample is ε_u. The shape-memory effect is triggered by heating the sample to a temperature above the transition temperature. The heating rate during shape recovery was 4ºC·min^{-1}. The amount of non-recoverable deformation at the end of programming is ε_p. The fixing (R_f) and recovery (R_r) ratios were calculated for each sample using Eqs. (1) and (2)[16].

$$R_f(\%) = \frac{\varepsilon_u}{\varepsilon_m} \qquad (1)$$

$$R_r(\%) = \frac{\varepsilon_m - \varepsilon_p}{\varepsilon_m} \qquad (2)$$

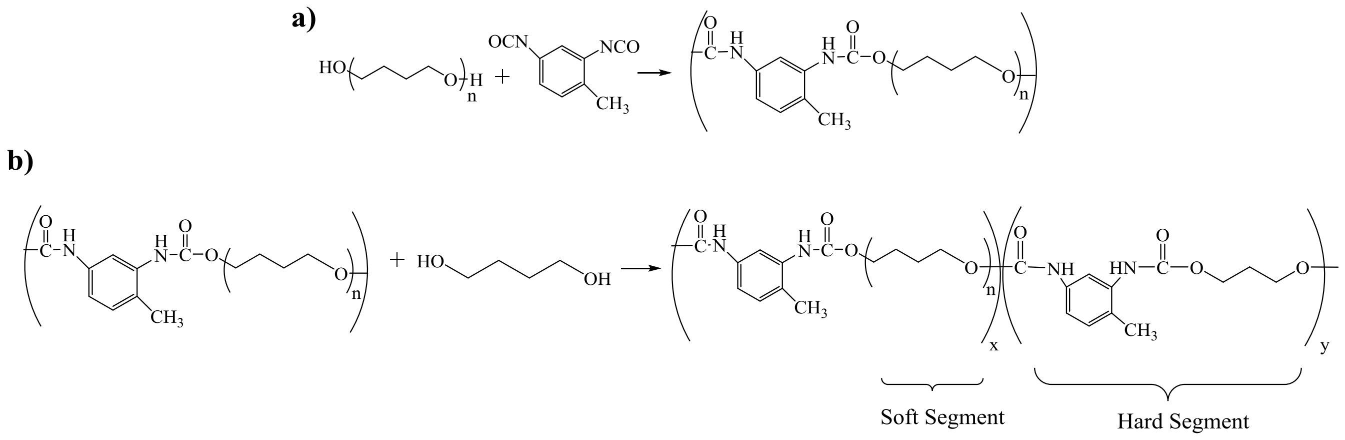


Figure 7. Synthesis route for toluene 2,4-diisocyanate (TDI) based polyurethanes: a) first step, b) second step.

4. Conclusions

In this work, shape memory polyurethanes (SMPUs) have been successfully synthetized by a two-step polymerization. Polyurethane samples display two-stage degradation showing good thermal stability with initial decomposition temperatures higher than 280ºC. Moreover, the glass transition temperatures were measured by DSC and DMTA. Both methods indicate that the glass transition temperature increase with the hard segment content, suggesting that reaction procedure was appropriate.

Finally, shape memory behavior of the synthetized polyurethanes was qualitatively and quantitatively evaluated. The qualitative evaluation demonstrates promoted shape memory response for all samples. At the same time, the quantitative analysis using TMA show that all the SMPUs samples are characterized by fixity ratios higher than 87% and recovery ratios near 99%.

Acknowledgments

The authors would like to acknowledge Government of La Rioja and the Footwear Technology Center of La Rioja for the financial support.

Author Contributions

M.Sáenz-Pérez realized the experiments and wrote the main paper. J.M. Laza, J. García-Barrasa, L.M. León and J.L. Vilas supervised the project.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Lendlein, A. & Kelch, S. Shape-Memory Polymers. *Angew. Chemie Int. Ed.* 2002, 41, 2034–2057.
2. Zhang, W., Chen, L. & Zhang, Y. Surprising shape-memory effect of polylactide resulted from toughening by polyamide elastomer. *Polymer (Guildf).* 2009, 50,1311–1315.
3. Hu, J., Meng, H., Li, G. & Ibekwe, S. I. A review of stimuli-responsive polymers for smart textile applications. *Smart Mater. Struct.* 2012, 21, 53001.
4. Peponi, L. *et al.* Synthesis and characterization of PCL–PLLA polyurethane

with shape memory behavior. *Eur. Polym. J.* 2013, 49, 893–903.

5. Serrano, M. C. & Ameer, G. A. Recent insights into the biomedical applications of shape-memory polymers. *Macromol. Biosci.* 2012, 12, 1156–71.
6. Huang, W. M. *et al.* Shaping tissue with shape memory materials. *Adv. Drug Deliv. Rev.* 2013, 65, 515–535.
7. Chang, Z. *et al.* Synthesis and properties of segmented polyurethanes with triptycene units in the hard segment. *Polymer (Guildf).* 2013, 54, 6910–6917.
8. Prisacariu, C., Scortanu, E., Coseri, S. & Agapie, B. Effect of Soft Segment Polydispersity on the Elasticity of Polyurethane Elastomers. *Ind. Eng. Chem. Res.* 2013, 52, 2316–2322.
9. Trovati, G., Sanches, E. A., Neto, S. C., Mascarenhas, Y. P. & Chierice, G. O. Characterization of polyurethane resins by FTIR, TGA, and XRD. *J. Appl. Polym. Sci.* 2010, 115, 263–268.
10. Liu, N. *et al.* The effects of the molecular weight and structure of polycarbonatediols on the properties of waterborne polyurethanes. *Prog. Org. Coatings.* 2015, 82, 46–56 (2015).
11. Yang, J. H., Chun, B. C., Chung, Y.-C. & Cho, J. H. Comparison of thermal/mechanical properties and shape memory effect of polyurethane block-copolymers with planar or bent shape of hard segment. *Polymer (Guildf).* 2003, 44, 3251–3258.
12. Lin, J. R. & Chen, L. W. Study on shape-memory behavior of polyether-based polyurethanes. II. Influence of soft-segment molecular weight. *J. Appl. Polym. Sci.* 1998, 69, 1575–1586.
13. Fritzsche, N. & Pretsch, T. Programming of Temperature-Memory Onsets in a Semicrystalline Polyurethane Elastomer. *Macromolecules.* 2014, 47, 5952–5959.
14. Pieczyska, E. A. *et al.* Thermomechanical properties of polyurethane shape memory polymer–experiment and modelling. *Smart Mater. Struct.* 2015, 24, 045043.
15. Saralegi, A., Gonzalez, M. L., Valea, A., Eceiza, A. & Corcuera, M. A. The role of cellulose nanocrystals in the improvement of the shape-memory properties of castor oil-based segmented thermoplastic polyurethanes. *Compos. Sci. Technol.* 2014, 92, 27–33.
16. Axpe, E. *et al.* Connecting free volume with shape memory properties in noncytotoxic gamma-irradiated polycyclooctene. *J. Polym. Sci. Part B Polym. Phys.* 2015, 53, 1080–1088.

SciForum
Mol2Net

Study of Dried Blood Spot Reliability for Quantitative Drug Analysis by UHPLC-PDA-FLUO

Beatriz Uribe[1], Oihane E. Alboniga[1], Oskar González*,[1] and Rosa M. Alonso[1]

[1] Analytical Chemistry Department, Science and Technology Faculty, the Basque Country University/EHU, P.O. Box 644, Bilbao, Basque Country 48080, Spain

* oskar.gonzalezm@ehu.eus Tel.: +946012294 Fax: +946013500.

Published: 4 December 2015

Abstract: In this work, the reliability of Dried Blood Spot (DBS) as a sampling technique for drug analysis was studied by Ultra High Performance Liquid Chromatography coupled to Photodiode-Array and Fluorescence Detection (UHPLC-PDA-FLUO). DBS microsampling, a technique based on placing a drop of blood in a cotton support that is allowed to air dry, has lately noticed an increase in use in bioanalysis. Even thought it offers several advantages compared to common blood sampling methods, it also shows some limitations for quantitative analysis due to the dependence on different factors. In this study, the influence of some of them (hematocrit, blood volume and sampling position) has been investigated, using amiloride, propranolol and valsartan drugs as model compounds. According to the results, it has been concluded that the sampling position and the hematocrit have influence in the accuracy and precision of the quantitative results, therefore limiting the use of this technique. On the other hand, dispersion of the analytes in the blood drop depends on their physicochemical properties which implies that the distribution of each analyte must be carefully studied during method development.

Keywords: DBS; UHPLC-PDA-FLUO; Bioanalysis

1. Introduction

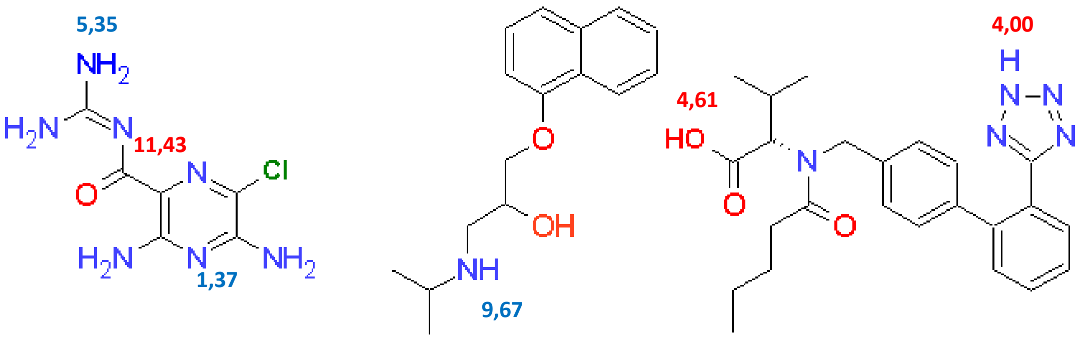
Dried blood spot (DBS) sampling method was first used with human blood in 1963 by Roberth Gurthrie to detect metabolomic diseases (phenylketonuria) in newborn infants. Since then, it has been used in many different areas such as toxicokinetics and pharmacokinetic studies, diagnostic screening, and therapeutic drug monitoring [1,2]. Due to the simplicity of this technique and thanks to the improvement in terms of sensitivity of the analytical instrumentation, its use has been notably increased in the last years.

DBS microsampling is a simple technique in which a drop of blood is placed in a support that is left to air dry prior to analysis. The support is usually cellulose paper chemically treated to prevent the growth of bacteria and other microorganisms [3] and sampling is done by putting a drop of blood from a heel, ear or finger prick in the support [2,4].

Once the blood is placed and dried, the analytes are extracted from the support usually by a liquid extraction of a punch done in the drop, and then analyzed [3]. DBS technique is used as a sampling method for a wide range of analytical techniques such as DNA-based assays, enzyme activity assays, immunoassays, direct mass spectrometry, and liquid chromatography coupled to different detectors [1].

DBS technique shows many advantages compared with conventional blood, plasma or serum collection procedures [5]. Among them, the stability of the cellulose-fixed analytes, the little blood volume required, the possibility of automation of the sample processing [6], the easy storage and transportation [4], or the lower biological risk in comparison with liquid blood samples. Despite these advantages there are also some disadvantages to consider. For example, when they are used in quantitative analysis, some factors can affect the reliability of the results[4,7,8], such as the type of blood (venous or capillary) [9,10], the type of support [11,12], the hematocrit [3,9,10,13-15] or the blood volume placed [16,17].

The aim of this work is the study of the influence of the blood volume, the hematocrit and the punching position on the quantitative results of a DBS based method in order to increase the reliability of the analysis and better understand the dispersion of the analytes in the support. For this purpose three drugs with different physico-chemical properties (amiloride, propranolol and valsartan) were selected as model compounds (Figure 1). For the analysis of these compounds a robust quantitative method was optimized using Ultra High Performance Liquid Chromatography coupled to Photodiode-Array and Fluorescence Detection (UHPLC-PDA-FLUO).



Amiloride
$C_6H_8ClN_7O$
229.62 g/mol
λ_{Ex}= 363 nm
λ_{Em} = 415 nm

Propranolol
$C_{16}H_{21}NO_2$
259.34 g/mol
λ_{Ex}= 287 nm
λ_{Em} = 340 nm

Valsartan
$C_{24}H_{29}N_5O_3$
435.51 g/mol
λ_{Ex}= 237 nm
λ_{Em} = 371 nm

Figure 1. Amiloride, propranolol and valsartan. Chemical structure, molecular formula and weight, and fluorescence excitation and emission wavelengths. In blue, the pKa for basic protons and in red for acidic protons.

2. Results and Discussion

2.1. Study of the influence of hematocrit and sample volume in DBS analysis

The influence of hematocrit in the quantitative application of DBS has been already reported by some authors [3,5,9,18]. In this work seven hematocrit values (25%, 30%, 35%, 40%, 45%, 50% and 55%) were studied in order to cover a wide interval of the hematocrit values of the population. Three sample volumes (15µL, 25µL and 35µL) were studied for each hematocrit level.

Using ANOVA, at 95% confidence level, it was observed for amiloride and propranolol that the results are influenced by the hematocrit value but not by the sample volume. In the case of valsartan, both variables have influence, being the effect of the hematocrit much more significant. In Figure 2 the increase of the chromatographic response of propranolol with the hematocrit value can be noticed. The extreme values (25% and 55%) were compared using a t-test (95% confidence level) and a significant difference was confirmed. The reason of this difference is probably the change in the blood drop area. As it decreases when the hematocrit value increases (due to the density/viscosity difference), the blood volume present in the sample punch is higher for the high hematocrit value samples, and also the amount of analyte.

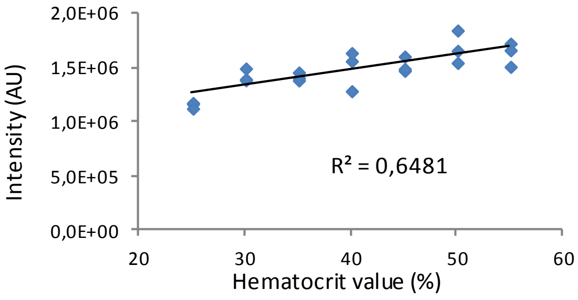


Figure 2. Chromatographic responses obtained for propranolol samples with different hematocrit values, using a 15 µL blood drop.

2.2. Study of the influence of the sampling position

Due to a heterogeneous dispersion of the blood or to chromatographic processes involving the analytes, the distribution of the drugs in the blood spot may not be homogeneous. In those cases the position where the punch that will be analysed is done becomes a critical factor.

To study the influence of the dispersion of the different analytes, the blood drop area has been divided in three zones: central, upper and lower peripheral. In this way a radial distribution of the analytes could be studied as well as the effect of the paper position during the drying step. In order to obtain a sample above the quantitation limit of the method 15 mini punches (1.2 mm diameter) were collected in each zone.

The results of the analysis can be observed in Figure 3. The statistical analysis did not show significant differences in terms of precision but a clear difference in analytes distribution behavior was noticed. On the one hand, amiloride concentration is higher in the central zone, being this behavior more obvious/perceptible at lower hematocrit values. On the other hand, propranolol and valsartan behave differently, with higher concentration values at the peripheral zone (phenomena known as volcano effect and already observed by other authors [11,12]).

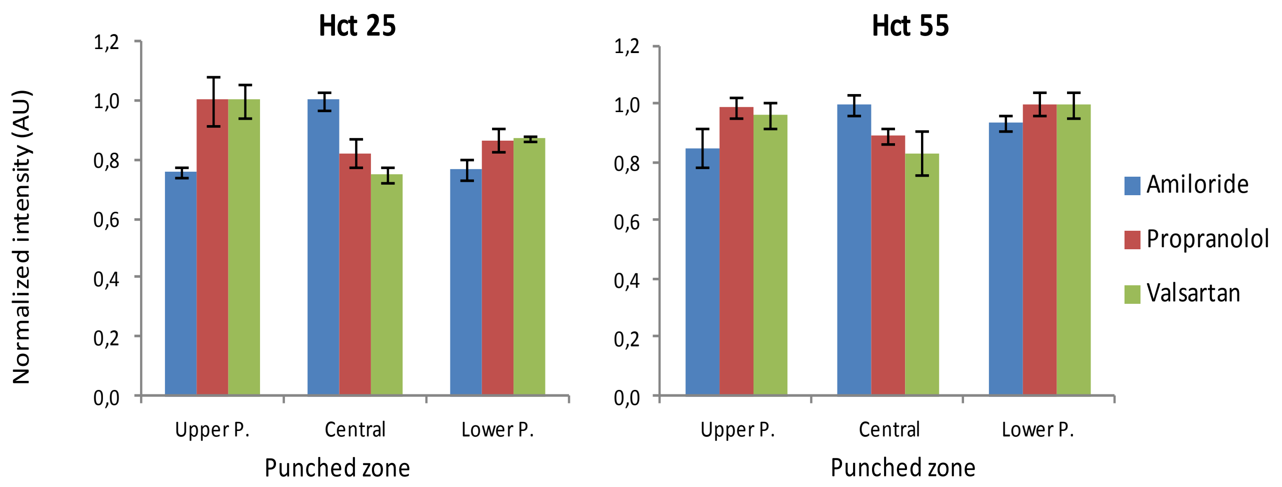


Figure 3. Normalized results for the chromatographic responses of the analytes belonging to central, upper and lower peripheral zones of the blood spot.

3. Materials and Methods

3.1. Chemicals

Amiloride hydrochloride was purchased from Sigma-Aldrich (St. Louis, USA), propranolol hydrochloride from Fluka (Burch, Switzerland) and valsartan from Novartis Pharma AG (Basel, Switzerland). Methanol used for the preparation of stock and working solutions was provided by Romil (Cambridge, England). For the preparation of the different pH solutions dipotassium phosphate (>99%), acetic acid (LC-MS grade), sodium dihydrate citrate anhydrous (>99%) and disodium hydrogen citrate sesquihydrate (>99%) were purchased from Fluka. Potassium dihydrogen phosphate, sodium acetate, trisodium dihydrate citrate, ammonium (25%) and ammonium chloride all of Pro-Analysi grade, and sodium hydroxide, EMSURE (Millipore's premium grade), were obtained from Merck (Darmstad, Germany). Phosphoric acid (85%) and citric acid (PA-ACS-ISO) were supplied by Panreac (Barcelona, Spain). Acetonitrile used for the preparation of the mobile phases was obtained from Romil. LC-MS quality grade formic acid was purchased from Fluka. Ultrapure analytical water was obtained from a Milli-Q Element A10 sysrem (Millipore, Milford, USA).

3.2. Instrumentation

The DBS cards used were Protein saver 903 of Whatman (NJ, USA). The puncher used was a regular office puncher with a diameter of 5.9 mm and the mini puncher was purchased from Harris (CA, USA) with a diameter of 1.2 mm. Analyses were performed on an Acquity UPLC system (Waters, Mildford, USA), coupled to a PDA detector and FLUO detector with a scheduled excitation and emission wavelength method. System control, data collection and data processing were accomplished using Empower 2 software. The chromatographic column used was Acquity BEH C18 (2.1x50, 1.7 µm) of Waters with a filter pre-column. The pH was measured with a Crison GPL22 pH-meter (Barcelona, Spain). To draw the molecules of the analytes ChemSketch 10.02 has been used and for the statistical treatment of the results Microsoft Excel 2010 and Unscrambler softwares have been used.

3.3. Standard solutions and blood samples

Standard stock solutions of 1000 mg/L were prepared in methanol for each analyte separately. With those solutions 50 mg/L working solutions were prepared. 15 blood samples covering the range of 22%-55% hematocrit value were generously provided by the University Hospital of Basurto (Bilbao, Spain), Samples were stored at -80ºC until analysis. The blood samples were spiked from working solutions to concentration of 1 mg/L.

3.4. Chromatographic conditions

The mobile phase consisted of solvent A (0.01% formic acid) and solvent B (acetonitrile). The gradient applied was the following: 0-0.5 min, 1% B; 0-3 min, 1 to 99% B; 3-3.5 min, 95% B; 3.5-3.6 min, 95 to 1% B; 3.6-4.5 min 1% B. Flow was kept at 0.55mL/min. During the chromatographic analysis the column was thermostated at 35 °C and samples were kept at 10±1°C in the autosampler. The PDA wavelength range was set from 190 to 400 nm and the excitation and emission wavelengths of the FLUO detector were as follows: from 0.00-1.39 min, 363/415 nm; 1.40-1.59 min, 287/340 nm; 2.00-4.50 min, 237/371nm).

3.5. Extraction conditions

The extraction conditions were carefully optimized. Extraction solution was methanol: pH 2 phosphate buffer (75:25). 200 µL of this solution was added to a 5.9 mm punch and after sonication and centrifugation the supernatant was transferred to a chromatographic vial.

4. Conclusions

In this work, some limitations of DBS as a sampling technique for small molecule quantitative analysis have been observed. On the one hand, it has been demonstrated that the chromatographic response of samples with the same analyte concentration but different hematocrit values are significantly different. This effect is attached to the difference in blood drop area, which decreases when hematocrit value increases. On the other hand, due the heterogeneous dispersion of the analytes in the blood drop the punching position is a critical parameter that affects the quantitative results. Furthermore, it has been observed that this dispersion is different depending on the analyte which means that during method development the behavior of each analyte must be carefully studied since, currently, anticipating the distribution of the analyte in the blood drop is not possible.

To prevent these problems the analysis of the whole blood drop can be an alternative approach. Nevertheless, this would require a more time- and solvent-consuming method and would make the automation of the sample treatment more complicated. In addition, from the point of view of a reliable quantitation all the blood drops (calibration and study samples) should have the same volume. Although that is not a problem for samples prepared in a laboratory it would make sampling more complicated under other conditions (hospitals, third-world countries, etc.) which is one of the main advantage of the DBS technique. Therefore, analytical methods based on DBS and punching should be studied more in depth in order to guarantee a reliable quantitation.

Acknowledgments

The authors thank the Spanish Ministry of Economy and Competitivity (Project CTQ2013-46179-R), University of the Basque Country (UFI 11/23) and Basque Country Government (Project IT789/13) for financial support. Oskar Gonzalez also thank Basque Country Government for his postdoctoral grants.

Author Contributions

Conceived and designed the experiments: BU, OG, RA. Performed the experiments: BU, OA.
Analyzed the data: BU, OG. Wrote the paper: BU, OG

Conflicts of Interest

The authors declare no conflict of interest

References and Notes

1. Demirev, P.A. Dried blood spots: Analysis and applications. *Analytical chemistry* **2013**, *85*, 779-789.
2. Odoardi, S.; Anzillotti, L.; Strano-Rossi, S. Simplifying sample pretreatment: Application of dried blood spot (dbs) method to blood samples, including postmortem, for uhplc-ms/ms analysis of drugs of abuse. *Forensic science international* **2014**, *243*, 61-67.
3. Li, W.; Tse, F.L. Dried blood spot sampling in combination with lc-ms/ms for quantitative analysis of small molecules. *Biomedical chromatography : BMC* **2010**, *24*, 49-65.
4. Tretzel, L.; Thomas, A.; Geyer, H.; Gmeiner, G.; Forsdahl, G.; Pop, V.; Schanzer, W.; Thevis, M. Use of dried blood spots in doping control analysis of anabolic steroid esters. *Journal of pharmaceutical and biomedical analysis* **2014**, *96*, 21-30.
5. Hawwa, A.F.; Albawab, A.; Rooney, M.; Wedderburn, L.R.; Beresford, M.W.; McElnay, J.C. A novel dried blood spot-lcms method for the quantification of methotrexate polyglutamates as a potential marker for methotrexate use in children. *PloS one* **2014**, *9*, e89908.
6. Deglon, J.; Thomas, A.; Mangin, P.; Staub, C. Direct analysis of dried blood spots coupled with mass spectrometry: Concepts and biomedical applications. *Analytical and bioanalytical chemistry* **2012**, *402*, 2485-2498.
7. Berm, E.J.; Paardekooper, J.; Brummel-Mulder, E.; Hak, E.; Wilffert, B.; Maring, J.G. A simple dried blood spot method for therapeutic drug monitoring of the tricyclic antidepressants amitriptyline, nortriptyline, imipramine, clomipramine, and their active metabolites using lc-ms/ms. *Talanta* **2015**, *134*, 165-172.
8. Thomas, A.; Geyer, H.; Schanzer, W.; Crone, C.; Kellmann, M.; Moehring, T.; Thevis, M. Sensitive determination of prohibited drugs in dried blood spots (dbs) for doping controls by means of a benchtop quadrupole/orbitrap mass spectrometer. *Analytical and bioanalytical chemistry* **2012**, *403*, 1279-1289.
9. De Kesel, P.M.; Capiau, S.; Lambert, W.E.; Stove, C.P. Current strategies for coping with the hematocrit problem in dried blood spot analysis. *Bioanalysis* **2014**, *6*, 1871-1874.
10. Capiau, S.; Stove, V.V.; Lambert, W.E.; Stove, C.P. Prediction of the hematocrit of dried blood spots via potassium measurement on a routine clinical chemistry analyzer. *Analytical chemistry* **2013**, *85*, 404-410.
11. Li, W.; Lee, M.S. *Dried blood spots: Applications and techniques*. 2014; p 376.
12. Cobb, Z.; de Vries, R.; Spooner, N.; Williams, S.; Staelens, L.; Doig, M.; Broadhurst, R.; Barfield, M.; van de Merbel, N.; Schmid, B.*, et al.* In-depth study of homogeneity in dbs using two different techniques: Results from the ebf dbs-microsampling consortium. *Bioanalysis* **2013**, *5*, 2161-2169.

13. Svensson, L.D.; Sennbro, C.J.; Svanstrom, C.; Hansson, G.P. Applying dried blood spot sampling with lcms quantification in the clinical development phase of tasquinimod. *Bioanalysis* **2015**, *7*, 179-191.
14. Wilhelm, A.J.; den Burger, J.C.; Vos, R.M.; Chahbouni, A.; Sinjewel, A. Analysis of cyclosporin a in dried blood spots using liquid chromatography tandem mass spectrometry. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2009**, *877*, 1595-1598.
15. Li, Y.; Henion, J.; Abbott, R.; Wang, P. The use of a membrane filtration device to form dried plasma spots for the quantitative determination of guanfacine in whole blood. *Rapid communications in mass spectrometry : RCM* **2012**, *26*, 1208-1212.
16. ter Heine, R.; Rosing, H.; van Gorp, E.C.; Mulder, J.W.; van der Steeg, W.A.; Beijnen, J.H.; Huitema, A.D. Quantification of protease inhibitors and non-nucleoside reverse transcriptase inhibitors in dried blood spots by liquid chromatography-triple quadrupole mass spectrometry. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2008**, *867*, 205-212.
17. Vu, D.H.; Koster, R.A.; Alffenaar, J.W.; Brouwers, J.R.; Uges, D.R. Determination of moxifloxacin in dried blood spots using lc-ms/ms and the impact of the hematocrit and blood volume. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2011**, *879*, 1063-1070.
18. Holub, M.; Tuschl, K.; Ratschmann, R.; Strnadova, K.A.; Muhl, A.; Heinze, G.; Sperl, W.; Bodamer, O.A. Influence of hematocrit and localisation of punch in dried blood spots on levels of amino acids and acylcarnitines measured by tandem mass spectrometry. *Clinica chimica acta; international journal of clinical chemistry* **2006**, *373*, 27-31.
19. Wiklund, T. Acquity uplc injection technigues fixed loop and flow through needle. https://www.waters.com/webassets/cms/library/docs/local_seminar_presentations/DA_NUT2013_G4_Tony_Wiklund.pdf (13/07/2015),

Mol2Net, **2015**, 1(*Section C*), pages 1-8, *Proceedings*
http://sciforum.net/conference/mol2net-1

SciForum
Mol2Net

Shape Memory Behavior of a Commercial Gamma-Irradiated Polycyclooctene

Nuria García-Huete [1,*], José Manuel Laza [2], José María Cuevas [3], José Luis Vilas [1,2] and Luis Manuel León [1,2]

[1] Basque Center for Materials, Applications and Nanostructures (BCMaterials), Parque Tecnológico de Bizkaia, Ed. 500, Derio 48160, Spain;

[2] Departamento de Química Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco/EHU, Apdo.644, Bilbao E-48080, Spain; E-Mails: josemanuel.laza@ehu.eus (J.M.L.); joseluis.vilas@ehu.eus (J.L.V); luismanuel.leon@ehu.eus (L.M.L.);

[3] Gaiker Technology Centre, Parque Tecnológico de Bizkaia, Ed. 202, Zamudio 48170, Spain; E-Mail: cuevas@gaiker.es

* Author to whom correspondence should be addressed; E-Mail: nuria.garcia@bcmaterials.net; Tel.: +34 94 612 88 11.

Published: 4 December 2015

Abstract: Gamma radiation process for modification of commercial polymers is a widely applied technique to promote new physical, chemical and mechanical properties. Gamma irradiation originates free radicals which can induce chain scission or crosslinking in the polymer backbone. The aim of this work is to research the structural, thermal and mechanical changes induced on a commercial polycyclooctene (PCO) when it is irradiated with a gamma source of ^{60}Co. After gamma irradiation, gel content was determined by Soxhlet extraction in cyclohexane, and thermal properties were evaluated before Soxhlet extraction by means of Thermogravimetric Analysis (TGA) and Differential Scanning Calorimetry (DSC). Finally, the shape memory properties were evaluated both qualitatively and quantitatively, the last one by Thermo-Mechanical Analysis (TMA).

Keywords: Gamma irradiation, Crosslinking, Thermo-Mechanical Analysis, Shape Memory behavior

1. Introduction

Shape memory polymers (SMPs) are those capable of recovering their original shape after having been deformed into a different one (temporary shape), i.e. they ‘remember’ the shape they were given when processed (permanent shape). The shape memory effect can be induced under appropriate stimulus such as temperature [1,2].

The irradiation of polymeric materials with ionizing radiation, like gamma rays, leads to the formation of free radicals [3], which can produce crosslinking and/or scission of the macromolecular chains [4]. Different polymers have been crosslinked employing gamma rays, like polyethylene [5,6], polyamides [7] and poly(vinylidene fluoride) [8].

Under the controlled crosslinking process, polycyclooctene (PCO) has shown excellent shape memory properties [9]. Polycyclooctene has been previously crosslinked using a peroxide [10–12] and employing a gamma rays source [13,14], showing interesting properties.

Here we present the thermal behavior of a selection of polycylooctene samples irradiated at different dosages of gamma rays and we compare them with a non-radiated polycylooctene sample. Additionally, the shape memory behavior of the irradiated samples is analyzed, showing promising results.

2. Results and Discussion

The measurement of gel content by Soxhlet extraction is considered a simple way to determine the degree of crosslinking reached after the gamma irradiation process. The gel fraction (wt%) indicates the insoluble fraction of the irradiated PCO samples. The results are summarized in Table 1. The gel fraction is higher as the radiation dose increases. This increase in the gel content values may be attributed to the radiation crosslinking in the PCO chains.

Thermogravimetric analysis (TGA) provides quantitative information of the weight loss process. All PCO samples decompose in one main breakdown stage (Figure 1). The degradation temperatures (T_d) are listed in Table 1, and there are not significant differences between the samples.

DSC measurements were performed in order to know the transition temperature (T_{trans}) of the shape memory effect, which corresponds with melting temperature (Figure 2).

From the first scan, it appears that radiation has no pronounced influence on T_m, whereas in the second scan T_m decrease with increasing radiation dose. It is well known that during polymer irradiation, both chain scission and crosslinking processes occur, and they take place, primarily, in the amorphous region, while some may take place in the interphase between the crystalline and amorphous regions [15], so it is logic not to observe changes in melting temperature in the first heating. The diminution of this value in the second heating scan can be explained as follows: the samples are recrystallized in the presence of the crosslinks formed in the irradiation process, which act as defect centres, restricting chain mobility of PCO chains, so T_m diminishes. The degree of crystallinity thus decreases with increased density of crosslinks due to more restricted mobility and conformational rearrangement of polymeric chains to form crystals.

The shape memory behavior of the samples has been studied both qualitatively and quantitatively.

In Figure 3 it can be observed the qualitatively studio for PCO-25 sample, whereas Figure 4 shows the same procedure for PCO-200. Both samples recover the original shape from an elbow-shaped temporary state, but in the case of the sample irradiated at 25 kGy this recovery is not total.

Thermomechanical Analysis was performed in order to quantify the shape memory behavior. The results are represented in Figure 5. Looking at Figure 5, recovery ratios near 100% can be appreciate (at the end of the experiment, the length of the samples are similar to the initial ones). This result to PCO-25 seems to be in discrepancy with the qualitatively analysis (Figure 3), but it is necessary to take into account that the thermo-mechanical cycle is completely different considering the different deformation mode and thermal treatment from the qualitative analysis made employing a hot and a cold bath, and a digital camera.

Table 1. Gel fraction, crystallinity values and characteristic temperatures for the studied samples.

Sample	Gel Fraction (wt%)	T_d (ºC)	T_{m1} (ºC)	T_c (ºC)	T_{m2} (ºC)	Crystallinity (%)
PCO-0	0	461.8	60.7	37.0	57.5	30.3
PCO-25	21.3	463.7	60.0	34.1	57.6	30.3
PCO-100	86.9	463.4	60.1	30.0	54.6	25.9
PCO-200	94.6	458.1	59.6	27.7	52.5	22.8

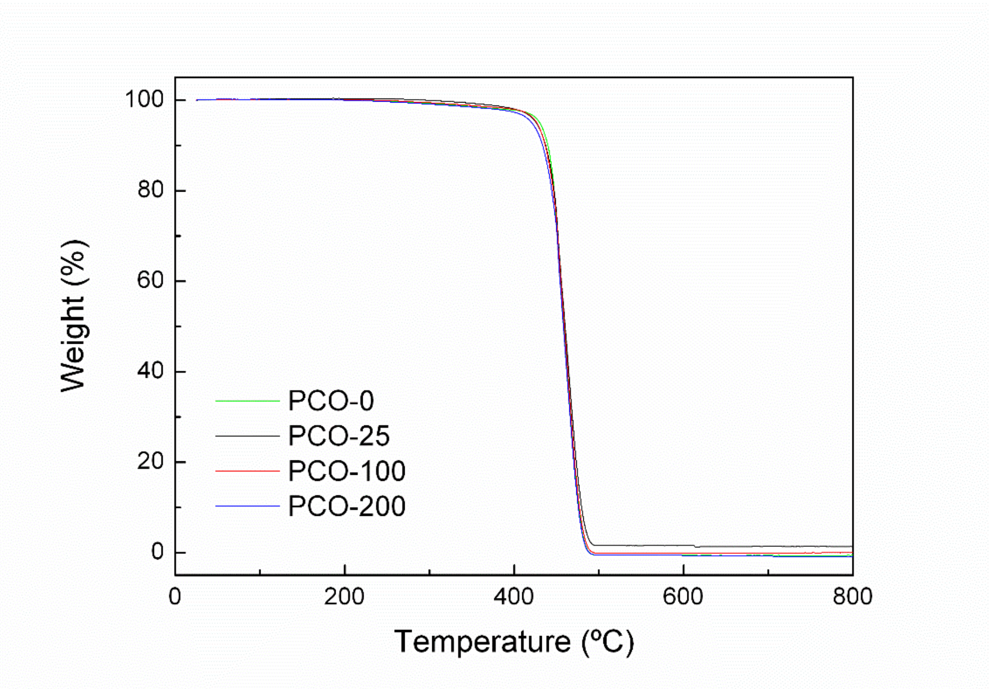


Figure 1. TGA curves for the studied PCO samples.

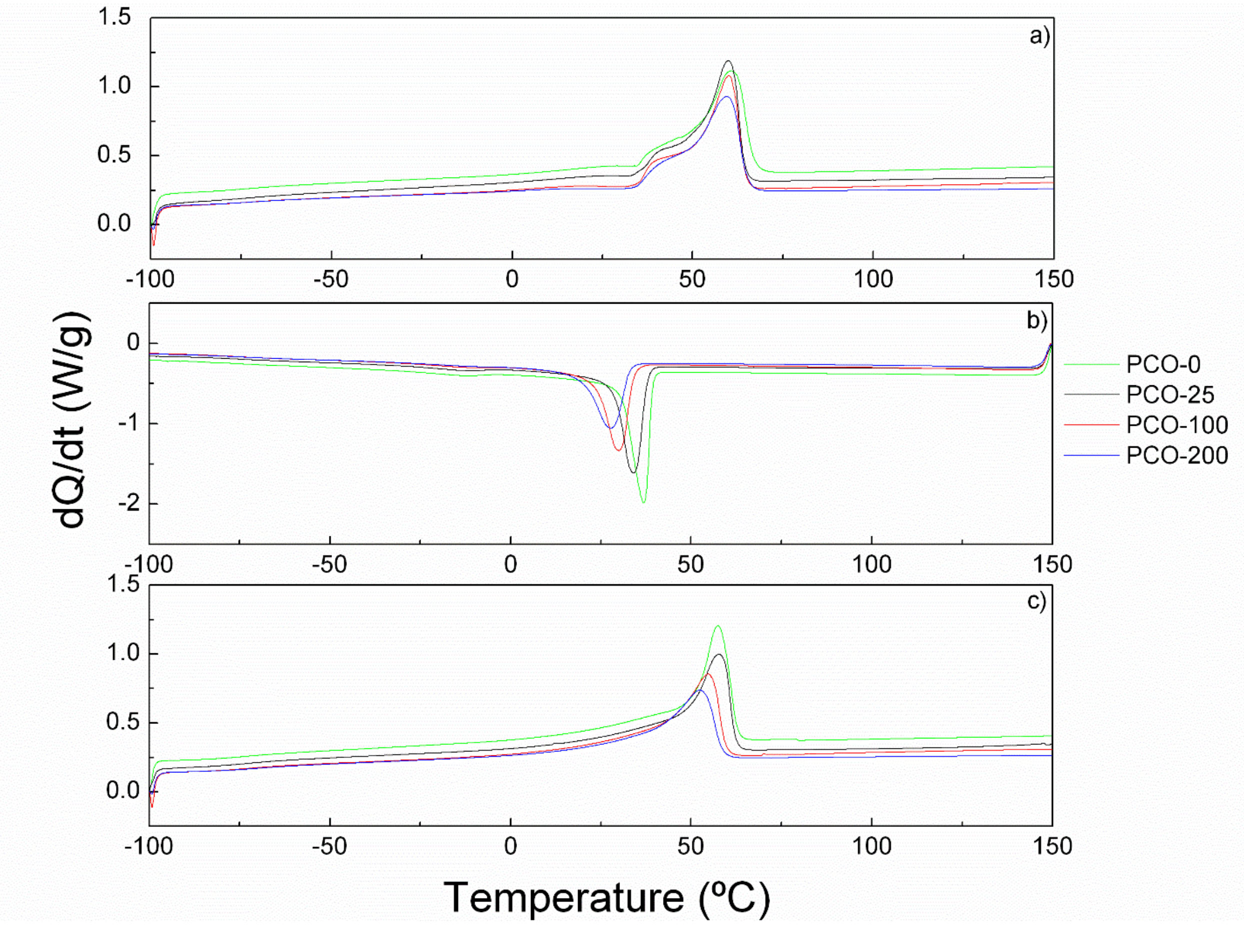


Figure 2. DSC curves for PCO samples: a) first heating, b) cooling and c) second heating scan.

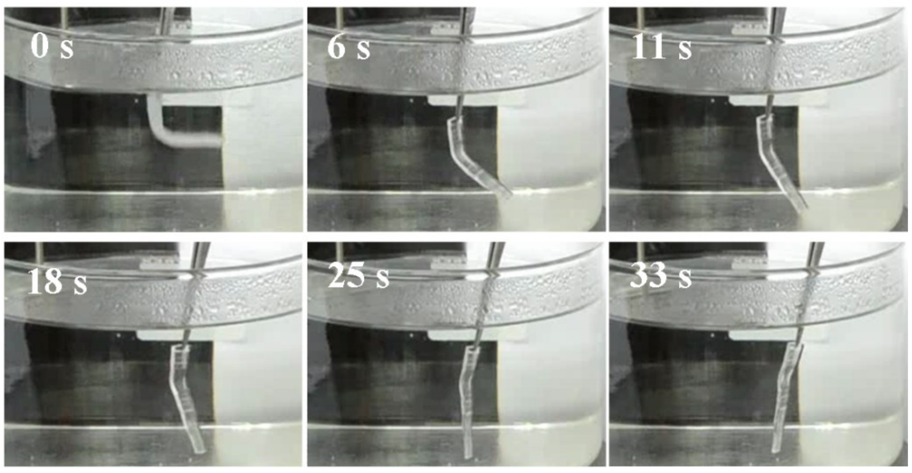


Figure 3. Shape memory recovery of elbow-shaped bent strip for 25 kGy irradiated PCO sample.

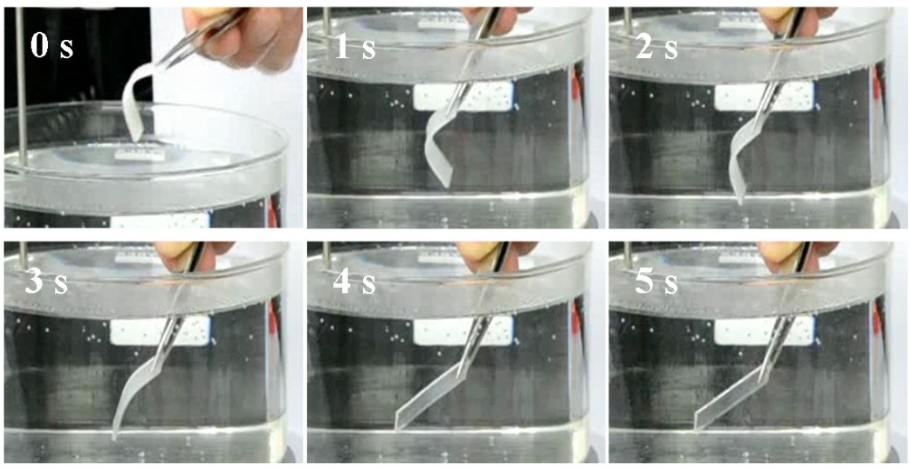


Figure 4. Shape memory recovery of elbow-shaped bent strip for 200 kGy irradiated PCO sample.

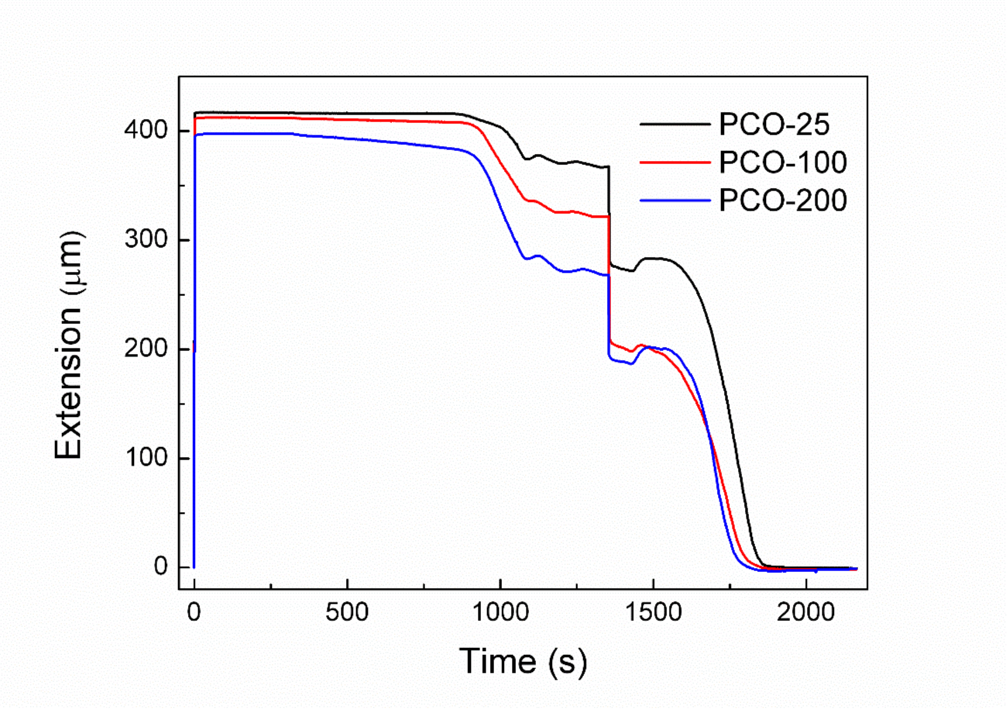


Figure 5. Thermo-mechanical cycle for PCO irradiated samples.

3. Materials and Methods

Polycyclooctene (PCO) Vestenamer® 8012 sheets were made by compression molding employing a hydraulic press with thermostatically controlled platens. PCO sheets were irradiated by gamma rays in a ^{60}Co radiation facility (NÁYADE Irradiation Plant of CIEMAT, Madrid, Spain) at the dosages of 25, 100 and 200 kGy [14].

The gel content of PCO samples was determined gravimetrically using 10 Soxhlet extraction cycles with boiling cyclohexane as solvent. After extraction, the samples were washed and vacuum dried at 50ºC to constant weight. The gel fraction was calculated using the following equation, where W_0 is the initial weight of sample and W_1 is the weight of sample after extraction.

$$gel\,fraction\;(wt\%) = \left({}^{W_1}/_{W_0}\right) \times 100$$

Thermal stability of the samples was evaluated by Thermogravimetric Analysis with a Mettler Toledo TGA/SDTA 851e thermobalance. The measurements were carried out from 25 to 800ºC with a heating rate of 10ºC·min^{-1} under nitrogen atmosphere (Figure 1).

Thermal properties of all samples were measured by Differential Scanning Calorimetry (DSC 822e from Mettler Toledo) to identify thermal actuation temperatures. The transition temperature of shape memory effect (T_{trans}) was defined from melting temperature measured in the second heating cycle (T_{m2}). Employing a constant nitrogen flow (50 mL·min^{-1}), samples were heated from −100 to 150ºC at a rate of 10ºC·min^{-1}, followed by a cooling scan from 150 to −100ºC at a rate of −10ºC·min^{-1}. Subsequently, a second heating scan to 150ºC was conducted at the same heating rate (Figure 2). The crystallinity of the samples was calculated using the next equation, employing the melting temperature in the second heating scan for each sample, where the enthalpy of a 100% crystalline polycyclooctene was 230 J·g^{-1} [16].

$$\%\,cryst = \left\{{}^{\Delta H_{m2}\;sample}/_{\Delta H_m(PCO)}\right\} \times 100$$

Thermally-induced shape memory behavior of irradiated PCO samples was qualitatively evaluated by digitally monitoring the shape recovery process. Rectangular strip samples were deformed in elbow-shaped strips at temperatures 10ºC above its transition temperature of shape memory effect. The temporary shape was fixed cooling down the temperature 20ºC below T_{trans}, and finally, the samples were heated-up above the transition temperature, so the thermal-induced recovery process was observed (Figures 3 and 4).

The quantitative evaluation of the shape memory behavior was performed using thermomechanical analysis (TMA). For that, samples shaped as strips with a cross-section area of 4 mm x 1.5 mm and initial clamps distance of 20 mm was employed. Taking into account the melting temperatures of the samples (Table 1), we thought appropriately to perform the thermo-mechanical experiments in the temperature range of 30-80ºC [17]. The analysis were conducted on a Mettler Toledo DMA-1 at a heating rate of 4ºC min^{-1}, recording the increase of the sample length as a function of temperature (Figure 5).

4. Conclusions

A commercial polycyclooctene (PCO) was irradiated with a gamma source of ^{60}Co at different doses (25-200 kGy). Soxhlet extraction allowed to measure the gel fraction for each sample. This insoluble fraction of each irradiated PCO sample is directly related to the degree of crosslinking reached for each sample in the radiation process, and it is higher as the radiation dose increases.

Thermal properties were evaluated by means of TGA and DSC, showing that the thermal stability of all PCO samples is quite similar independently of radiation dose.

The investigation of the thermal properties by DSC shows minor changes in the melting temperature with irradiation doses in the first heating scan (T_{m1}), which could be attributed to the immobilization of the generated free radicals in the crystalline region with hindered chain mobility. However, for the second heating scan, a decrease in T_{m2} with the increasing radiation dose was observed. During the recrystallization process, crosslinks between polymer chains act as defect centers, which restrict chain mobility of PCO chains and, therefore, the T_m values determined from the second heating scan are lower.

Finally, shape memory behavior of irradiated PCO samples was qualitatively and quantitatively evaluated. Except for the sample irradiated at 25 kGy, where the recoverability is not total, the qualitative evaluation demonstrate promoted shape memory response for the irradiated samples. At the same time, the quantitative analysis using TMA show that all the irradiated samples are characterized by recovery ratios near 100%.

Acknowledgments

Authors thank the Basque Country Government for financial support (Ayudas para apoyar las actividades de los grupos de investigación del sistema universitario vasco, IT718-13).

Author Contributions

Nuria García-Huete conducted the experimental work and the processing of data and wrote the paper. José María Cuevas, José Manuel Laza, José Luis Vilas and Luis Manuel León contributed to discussion and interpretation of results.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Liu, C.; Chun, S. B.; Mather, P. T.; Zheng, L.; Haley, E. H.; Coughlin, E. B. Chemically cross-linked polycyclooctene: synthesis, characterization, and shape memory behavior. *Macromolecules* **2002**, *35*, 9868–9874.

2. Lendlein, A.; Kelch, S. Shape-memory polymers. *Angew. Chemie Int. Ed.* **2002**, *41*, 2034–2057.

3. Chmielewski, A. G.; Haji-Saeid, M.; Ahmed, S. Progress in radiation processing of polymers. *Nucl. Instruments Methods Phys. Res. Sect. B Beam Interact. with Mater. Atoms* **2005**, *236*, 44–54.

4. Zahran, A. R. R.; Kandeil, A. Y.; Higazy, A. A.; Kassem, M. E. Ultrasonic and thermal properties of γ-irradiated low-density polyethylene. *J. Appl. Polym. Sci.* **1993**, *49*, 1291–1297.

5. Basfar, A. A. Flammability of radiation cross-linked low density polyethylene as an insulating material for wire and cable. *Radiat. Phys. Chem.* **2002**, *63*, 505–508.

6. Vachon, C.; Gendron, R. Effect of gamma-irradiation on the foaming behavior of ethylene-co-octene polymers. *Radiat. Phys. Chem.* **2003**, *66*, 415–425.

7. Feng, W.; Hu, F. M.; Yuan, L. H.; Zhou, Y.; Zhou, Y. Y. Radiation crosslinking of polyamide 610. *Radiat. Phys. Chem.* **2002**, *63*, 493–496.

8. Medeiros, A. S.; Gual, M. R.; Pereira, C.; Faria, L. O. Thermal analysis for study of the gamma radiation effects in poly(vinylidene fluoride). *Radiat. Phys. Chem.* **2015**.

9. Alonso-Villanueva, J.; Cuevas, J. M.; Laza, J. M.; Vilas, J. L.; León, L. M. Synthesis of poly(cyclooctene) by ring-opening metathesis polymerization: Characterization and shape memory properties. *J. Appl. Polym. Sci.* **2010**, *115*, 2440–2447.

10. Cuevas, J. M.; Laza, J. M.; Rubio, R.; German, L.; Vilas, J. L.; León, L. M. Development and characterization of semi-crystalline polyalkenamer based shape memory polymers. *Smart Mater. Struct.* **2011**, *20*, 035003.

11. Cuevas, J. M.; Rubio, R.; Laza, J. M.; Vilas, J. L.; Rodriguez, M.; León, L. M. Shape memory composites based on glass-fibre-reinforced poly(ethylene)-like polymers. *Smart Mater. Struct.* **2012**, *21*, 035004.

12. Cuevas, J. M.; Rubio, R.; German, L.; Laza, J. M.; Vilas, J. L.; Rodriguez, M.; Leon, L. M. Triple-shape memory effect of covalently

crosslinked polyalkenamer based semicrystalline polymer blends. *Soft Matter* **2012**, *8*, 4928–4935.

13. Abdel-Aziz, M. M.; Basfar, A. A. Aging of ethylene-propylene diene rubber (EPDM) vulcanized by γ-radiation. *Polym. Test.* **2000**, *19*, 591–602.

14. García-Huete, N.; Laza, J. M.; Cuevas, J. M.; Vilas, J. L.; Bilbao, E.; León, L. M. Study of the effect of gamma irradiation on a commercial polycyclooctene I. Thermal and mechanical properties. *Radiat. Phys. Chem.* **2014**, *102*, 108–116.

15. Luo, S.; Netravali, A. N. Effect of 60Co γ-radiation on the properties of poly (hydroxybutyrate- co- hydroxyvalerate). *J. Appl. Polym. Sci.* **1999**, *73*, 1059–1067.

16. Schneider, W. A.; Müller, M. F. Crystallinity of trans-polyoctenamer: characterization and influence of sample history. *J. Mol. Catal.* **1988**, *46*, 395–403.

17. Axpe, E.; García-Huete, N.; Cuevas, J. M.; Ribeiro, C.; Mérida, D.; Laza, J. M.; García, J. Á.; Vilas, J. L.; Lanceros-Méndez, S.; Plazaola, F.; León, L. M. Connecting free volume with shape memory properties in noncytotoxic gamma-irradiated polycyclooctene. *J. Polym. Sci. Part B Polym. Phys.* **2015**, *53*, 1080–1088.

SciForum
Mol2Net

Performance of the NOF Theory in the Description of the Four-Electron Harmonium Atom in the Singlet State

Mario Piris [1,2,3]*

[1] Kimika Fakultatea, Euskal Herriko Unibertsitatea (UPV/EHU),20018 Donostia, Euskadi, Spain.
[2] Donostia International Physics Center (DIPC), 20018 Donostia, Euskadi, Spain.
[3] IKERBASQUE, Basque Foundation for Science, 48013 Bilbao, Euskadi, Spain.
* E-Mail: mario.piris@ehu.eus; Tel.: +34-943-018-328

Published: 4 December 2015

Abstract: In recent years, the natural orbital functional (NOF) theory has emerged as an alternative approach to both density functional theory (DFT) and wave-function (WFN) methods for electronic structure investigations. Several NOFs have been proposed, for which validation is necessary. A well-known tool for calibration, testing, and benchmarking of an approximate electronic structure method is the harmonium atom. In this model system, the electron-nucleus potential is replaced by a harmonic confinement, but the electron-electron Coulomb interaction remains. By varying the strength of the harmonic potential, the correlation regime can be tuned, making possible the transition from the weakly to the strongly correlated regime. Accordingly, the harmonium stands as an adequate system for studying the behavior of approximate NOFs, since it is possible to contrast them with their exact counterparts obtained from the analytic solution. In this presentation, the comparison between the quasi-exact and approximate electron-electron repulsion energy provided by eight known NOFs, in the singlet state of the four-electron harmonium atom with varying confinements, is analyzed in some detail. The present approach, which will appear soon in the Journal of Chemical Physics 143 (arXiv:1511.06564 [physics.chem-ph]), not only reveals the failures of the functionals but also pinpoints the causes. In general, the functional PNOF6 shows the most consistent behavior, with decent accuracy, along all confinement regimes studied.

Keywords: 4e$^-$ Harmonium atom; Electron Correlation; Natural Orbital Functional Theory;
PNOF6; 1-Matrix Functionals, Reduced Density Matrices.

1. Introduction

The improvements in computer hardware and software have recently allowed the simulation of molecules with an increasing number of atoms. Unfortunately, the most accurate electronic structure methods based on N-particle wave functions (WFN) remain computationally too expensive to be applied to large systems. On the other hand, the most efficient method is the density functional theory (DFT). However, current implementations of DFT suffer from several problems, as for instance in the description of multireference systems.

An alternative to both DFT and WFN methods lies in the development of a functional theory based on the first-order reduced density matrix (1-RDM) [1] in its spectral expansion, known as natural orbital functional (NOF) theory [2]. Like the density, the 1-RDM is a much simpler object than the WFN. The ensemble N-representability conditions that have to be imposed on variations of the 1-RDM are well known [3]. The existence and properties of the 1-RDM energy functional are also well established [4]. The major advantage of a 1-RDM formulation is that the kinetic energy is explicitly defined and it does not require the construction of a functional. The unknown functional only needs to incorporate electron correlation. Several correlation functionals of this type have been proposed, for which validation is necessary.

A well-known tool for calibration, testing, and benchmarking of an approximate electronic structure method is the harmonium atom [5]. In this model, the replacement of the Coulombic central confining electron-nucleus potential occurring in real systems by a harmonic potential makes the problem separable in terms of center-of-mass and relative coordinates. Accordingly, the harmonium stands as an adequate system for studying the behavior of approximate NOFs, since it is possible to contrast them with their exact counterparts obtained from the analytic solution. Moreover, by varying the strength of the harmonic potential, ω, the correlation regime can be tuned, making possible the transition from the weakly to the strongly correlated regime.

The energies of several electronic states of the four-electron harmonium atom are presently known within ca. 1 μhartree for arbitrary values of ω [6]. The respective 1-RDMs and individual energy components are also available from such calculations. Finally, exact asymptotics of these electronic properties are available at both the weak- and strong-correlation limits. Recently [7], the comparison between the exact and approximate electron-electron repulsion energy (U) obtained with several approximate NOFs, for several states of few-electron harmonium atoms, has been reported. The aim of this presentation is to analyze in some detail the performance of eight well-known NOFs in the singlet state of the four-electron harmonium atom with confinements that range from ω=0.001 to ω=1000.

2. Results and Discussion

The NOFs assessed in [7] fall into two broad categories. The first of them encompasses expressions for U that involve only the exchange integrals (K). This category includes:

1. The NOF introduced by Müller [8].
2. The NOF of Goedecker and Umrigar (GU) [9].
3. The NOF of Csányi and Arias (CA) [10].
4. The NOF of Csányi, Goedecker and Arias (CGA) [11].
5. The BBC2 NOF [12].

6. The NOF of Marques-Lathiotakis (ML) [13].
7. The NOF of Marques and Lathiotakis corrected for self-interaction (ML-SIC) [13].

The last NOF included in this report is PNOF6 [14], which corresponds to a JKL functional, that is, involves only the Coulomb (J), exchange (K) and time-inversion-exchange (L) integrals. This NOF is the last member of the family of functionals [15,16] obtained from the reconstruction of the two-particle cumulant in order to satisfy known necessary N-representability conditions for the second-order reduced density matrix.

In Figure 1, the electron-electron repulsion energy U = V_{ee} - V_{ee}(HF) is depicted for the four-electron harmonium atom in the $^1D_+$ state. These values are taken from Tables VI and X of Ref. [7]. Due to its multireference character, this singlet state is challenging for K-only functionals 1-7. A rough agreement with the exact data is observed only for the GU, ML, and ML-SIC when the confinements are weak.

Among the functionals included, only PNOF6 goes parallel and above the exact solution for all confinements considered. Moreover, it is the only one capable of describing the ω→∞ limit of this state with decent accuracy, however its perfor-mance deteriorates upon weakening of the confi-nement since the error represents a higher percentage of the total repulsion energy U. It is worth to note that the PNOF family understimates the electron correlation effects in high-spin states [7] of the four-electron harmonium atom. These results provide clear clues for the future work.

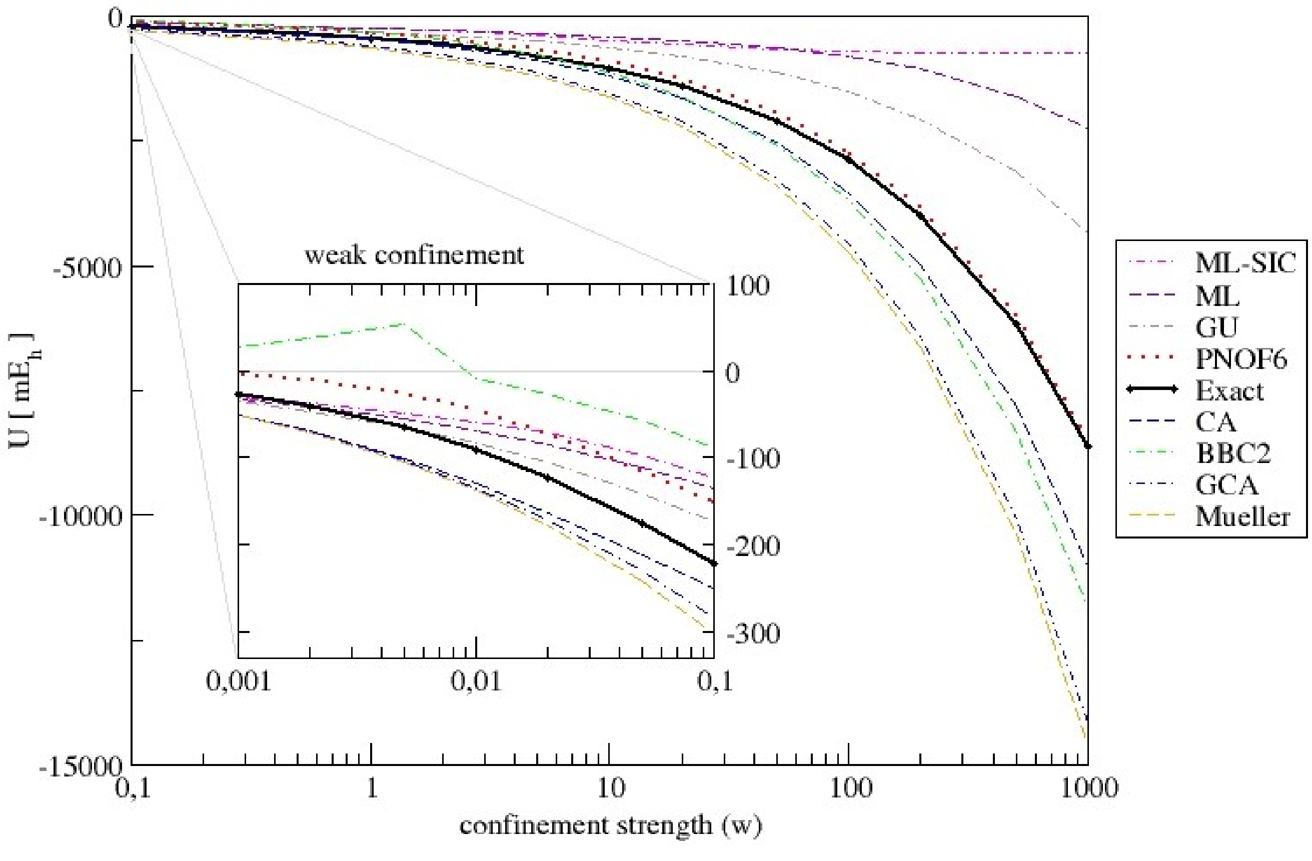


Figure 1. The electron-electron repulsion energy U = V_{ee} - V_{ee}(HF) of the four-electron harmonium atom in the $^1D_+$ state (values taken from Tables VI and X of reference [7])
.

Conflicts of Interest. The author declares no conflict of interest.

References and Notes

1. Gilbert, T. L. *Phys. Rev. B* **1975**, 12, 2111.
2. Piris, M. Natural Orbital Functional Theory. In Reduced-Density-Matrix Mechanics: With Applications to Many-Electron Atoms and Molecules; D. A. Mazziotti, Ed.; Wiley: Hoboken, New Jersey, 2007; Chapter 14, pp. 387–427.
3. Coleman, A. J. *Rev. Mod. Phys.* **1963**, 35, 668.
4. Cioslowski, J. *J. Chem. Phys.* **2005**, 123, 164106.
5. Cioslowski, J.; Matito, E. *J. Chem. Theory Comput.* **2011**, 7, 915; and the references cited therein.
6. Cioslowski, J.; Strasburger, K., Matito, E. *J. Chem. Phys.* **2014**, 141, 044128.
7. Cioslowski, J.; Piris, M.; Matito, E. *J. Chem. Phys.* **2015**. (arXiv:1511.06564 [physics.chem-ph])
8. Müller, A.M.K. *Phys. Lett. A* **1984,** 105, 446.
9. Goedecker S.; Umrigar, C. J. *Phys. Rev. Lett.* **1998**, 81, 866.
10. Csányi, G.; Arias, T. A. *Phys. Rev. B* **2000**, 61,7348.
11. Csányi,G.; Goedecker, S.; Arias, T. A. *Phys. Rev. A* **2002,** 65, 032510.
12. Gritsenko, O.; Pernal, K.; Baerends, E. *J. Chem. Phys.* **2005**, 122, 204102.
13. Marques, M. A. L.; Lathiotakis, N. N. *Phys. Rev. A* **2008***,* 77, 032509..
14. Piris, M. *J. Chem. Phys.* **2014**, 141, 044107.
15. Piris, M. *Int. J. Quantum Chem.* **2013**, 113, 620.
16. Piris, M; Ugalde, J. M. *Int. J. Quantum Chem.* **2014**, 114, 1169.

SciForum
Mol2Net

An Unprecedented Revolution in Medicinal Science

Kuo-Chen Chou[1,2]

[1] Gordon Life Science Institute, Boston, Massachusetts 02478, USA

[2] Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah 21589, Saudi Arabia

Published: 4 December 2015

Abstract: With the explosive growth of biological sequences in this century, medicinal science has been undergoing an unprecedented revolution. Driven by the avalanche of biological sequences generated in the postgenomic age, medicinal science is currently undergoing an unprecedented revolution [1,2], as indicated by, but not limited to, the following four aspects.

I. Post Biological Sequence Modification

Cancer and many other major diseases are often caused by a variety of subtle protein modifications, typically by various different post-translational modifications (PTMs or PTLM) in proteins [3,4]. In order to reveal their pathological mechanisms and find new and revolutionary strategies to treat them, many efforts have been made with the aim to predict the possible modified sites in various proteins concerned (see, e.g., [5-12] and a recent review paper [13]). Meanwhile, to provide efficient tools for the aforementioned researches, the pseudo amino acid composition (PseAAC) [14] and general PseAAC [15] were proposed. And the corresponding web-servers have been established [16-19] that can be used by researchers to generate various different modes of PseAAC to catch the key features of protein/peptide sequences according to their needs. The concept of PseAAC and its approaches to deal with protein/peptide sequences have been widely used in medical science and related areas (see, e.g., [11,20-46] as well as the Editorial [47] for a special Molecular Science issue focused on "Drug Development and Biomedicine"). Similar subtle modifications, the post-replication modification (PTRM) and post-transcription modification (PTCM), also occur respectively in DNA [48] and RNA sequences [49], causing many major diseases as well. Actually, considerable endeavors to develop high-throughput tools for predicting the possible modified sites of DNA/RNA sequences have also been underway (see, e.g., [1,13,50,51].

.

.

.

II. Genome Analysis

Genetic disorders are illnesses caused by one or more abnormalities in the genome, or by changes or variants in genes. To help understand these kinds of genetic or genomic diseases and find revolutionary therapeutic strategies to treat them, many in-depth genome analyses have been carried out recently (see, e.g., [52-69] and a review paper [70]). During the process of the aforementioned studies, inspired by the successes of using PseAAC in dealing with protein/peptide sequences, the pseudo K-tuple nucleotide composition or PseKNC was proposed to formulate various feature vectors for DNA/RNA sequences [71]. And the corresponding web-servers have been established as well [72-75], which can be used by researchers to do the same in computational genomics and genome analysis as done in computational proteomics and proteome analysis.

.

III. Personalized Medicine

Different patients may have different responses to the same drug in clinical treatments even they have the same plasma concentration. The reports from World Health Organization (WHO) indicates that, the percentages of patients having bad responses by using the marketed drugs are 10% ~20%, of which 5% are dead [76]. These kinds of bad responses are mainly due to individual differences, a formidable barrier against clinical treatments and pharmaceutical industry. Therefore, it is an inevitable direction to develop personalized medicines [77-80]. To realize this, it is important to establish personalized medical profile including personal genome, genetic, or genomic data. Personal genomics is the branch of genomics concerned with the sequencing and analysis of the genome of an individual. Again, the aforementioned tools are very useful in this regard..

.

.

IV. Drug-Target Interactions within Cellular Networking

Identifying drug-target interaction is one of the key steps for developing new medicines [81]. A typical tool used for this is molecular docking simulation (see, e.g., [52,82-95]). To conduct molecular docking, however, an indispensable entity is a reliable 3D (three dimensional) structure of the target protein. Although X-ray crystallography is a powerful tool in determining protein 3D structures, not all proteins can be successfully and timely crystallized, particularly for membrane proteins. Although NMR is a very powerful tool for determining the 3D structures of membrane proteins as reported by a series of recent publications (see, e.g., [96-102]), it is time-consuming as well. To timely acquire the 3D structural information, one has to resort to various structural bioinformatics tools [90], including the homologous modelling approach as utilized for a series of protein receptors urgently needed during the drug-developing process [103-110] and other computational modelling methods [111-113]. Unfortunately the number of dependable templates for developing high quality 3D protein structures via homology modelling is very limited [90], while the 3D structure developed purely by the approach of energy minimization or molecular dynamics without a good initial template might be far from the true structure owing to the local minimization problems. To pre-exclude those compounds, which are not likely to interact with the target proteins concerned, some sequence-based methods (see, e.g., [28,33,35,114,115] and a review paper [116]) have been developed that can serve to predict the interaction of the drug compounds with various kinds of target proteins in cellular networking. These newly developed methods can help us enormously in reducing the search scope and speeding up the pace of developing new drugs [117].

References and Notes

Chou, K. C. Impacts of bioinformatics to medicinal chemistry, *Medicinal Chemistry*, **2015**, *11*, 218-234.

[2] Zhou, G. P. *et al.* Perspectives in Medicinal Chemistry, *Current Topics in Medicinal Chemistry*, **2016**, *16*, 381-382.

[3] Foster, M. W. *et al.* Protein S-nitrosylation in health and disease: a current perspective, *Trends Mol Med*, **2009**, *15*, 391-404.

[4] Uehara, T. *et al.* S-nitrosylated protein-disulphide isomerase links protein misfolding to neurodegeneration, *Nature*, **2006**, *441*, 513-517.

[5] Xu, Y. *et al.* iSNO-PseAAC: Predict cysteine S-nitrosylation sites in proteins by incorporating position specific amino acid propensity into pseudo amino acid composition *PLoS ONE*, **2013**, *8*, e55844.

[6] Xu, Y. *et al.* iSNO-AAPair: incorporating amino acid pairwise coupling into PseAAC for predicting cysteine S-nitrosylation sites in proteins, *PeerJ*, **2013**, *1*, e171.

[7] Qiu, W. R. *et al.* iMethyl-PseAAC: Identification of Protein Methylation Sites via a Pseudo Amino Acid Composition Approach, *Biomed Res Int (BMRI)*, **2014**, *2014*, 947416.

[8] Xu, Y. *et al.* iHyd-PseAAC: Predicting hydroxyproline and hydroxylysine in proteins by incorporating dipeptide position-specific propensity into pseudo amino acid composition, *Int. J. Mol. Sci.*, **2014**, *15*, 7594-7610.

[9] Xu, Y. *et al.* iNitro-Tyr: Prediction of nitrotyrosine sites in proteins with general pseudo amino acid composition, *PLoS ONE*, **2014**, *9*, e105018.

[10] Qiu, W. R. *et al.* iUbiq-Lys: Prediction of lysine ubiquitination sites in proteins by extracting sequence evolution information via a grey system model *Journal of Biomolecular Structure and Dynamics (JBSD)* **2015**, *33*, 1731-1742.

[11] Jia, C. *et al.* Prediction of Protein S-Nitrosylation Sites Based on Adapted Normal Distribution Bi-Profile Bayes and Chou's Pseudo Amino Acid Composition, *Int J Mol Sci*, **2014**, *15*, 10410-10423.

[12] Zhang, J. *et al.* PSNO: Predicting Cysteine S-Nitrosylation Sites by Incorporating Various Sequence-Derived Features into the General Form of Chou's PseAAC, *Int J Mol Sci*, **2014**, *15*, 11204-11219.

[13] Xu, Y. *et al.* Recent progress in predicting posttranslational modification sites in proteins, *Curr Top Med Chem*, **2016**, *16*, 591-603.

[14] Chou, K. C. Prediction of protein cellular attributes using pseudo amino acid composition, *PROTEINS: Structure, Function, and Genetics (Erratum: ibid., 2001, Vol.44, 60)*, **2001**, *43*, 246-255.

[15] Chou, K. C. Some remarks on protein attribute prediction and pseudo amino acid composition (50th Anniversary Year Review), *J. Theor. Biol.*, **2011**, *273*, 236-247.

[16] Shen, H. B. *et al.* PseAAC: a flexible web-server for generating various kinds of protein pseudo amino acid composition, *Anal. Biochem.*, **2008**, *373*, 386-388.

[17] Du, P. *et al.* PseAAC-Builder: A cross-platform stand-alone program for generating various special Chou's pseudo-amino acid compositions, *Anal. Biochem.*, **2012**, *425*, 117-119.

[18] Cao, D. S. *et al.* propy: a tool to generate various modes of Chou's PseAAC, *Bioinformatics*, **2013**, *29*, 960-962.

[19] Du, P. *et al.* PseAAC-General: Fast building various modes of general form of Chou's pseudo-amino acid composition for large-scale protein datasets, *International Journal of Molecular Sciences*, **2014**, *15*, 3495-3506.

[20] Zhou, X. B. *et al.* Using Chou's amphiphilic pseudo-amino acid composition and support vector machine for prediction of enzyme subfamily classes, *J. Theor. Biol.*, **2007**, *248*, 546–551.

[21] Esmaeili, M. *et al.* Using the concept of Chou's pseudo amino acid composition for risk type prediction of human papillomaviruses, *J. Theor. Biol.*, **2010**, *263*, 203-209.

[22] Yu, L. *et al.* SecretP: Identifying bacterial secreted proteins by fusing new features into Chou's pseudo-amino acid composition, *J. Theor. Biol.*, **2010**, *267*, 1-6.

[23] Mohammad Beigi, M. *et al.* Prediction of metalloproteinase family based on the concept of Chou's pseudo amino acid composition using a machine learning approach, *Journal of Structural and Functional Genomics*, **2011**, *12*, 191-197.

[24] Nanni, L. *et al.* Identifying bacterial virulent proteins by fusing a set of classifiers based on variants of Chou's pseudo amino acid composition and on evolutionary information, *IEEE-ACM Transaction on Computational Biolology and Bioinformatics*, **2012**, *9*, 467-475.

[25] Zia-ur-Rehman *et al.* Identifying GPCRs and their Types with Chou's Pseudo Amino Acid Composition: An Approach from Multi-scale Energy Representation and Position Specific Scoring Matrix, *Protein & Peptide Letters*, **2012**, *19*, 890-903.

[26] Gupta, M. K. *et al.* An alignment-free method to find similarity among protein sequences via the general form of Chou's pseudo amino acid composition, *SAR QSAR Environ Res (SAR AND QSAR IN ENVIRONMENTAL RESEARCH)*, **2013**, *24*, 597-609.

[27] Khosravian, M. *et al.* Predicting Antibacterial Peptides by the Concept of Chou's Pseudo-amino Acid Composition and Machine Learning Methods, *Protein & Peptide Letters*, **2013**, *20*, 180-186.

[28] Min, J. L. *et al.* iEzy-Drug: A web server for identifying the interaction between enzymes and drugs in cellular networking, *BioMed Research International (BMRI)*, **2013**, *2013*, 701317.

[29] Mohabatkar, H. *et al.* Prediction of Allergenic Proteins by Means of the Concept of Chou's Pseudo Amino Acid Composition and a Machine Learning Approach, *Medicinal Chemistry*, **2013**, *9*, 133-137.

[30] Chou, K. C. Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes, *Bioinformatics*, **2005**, *21*, 10-19.

[31] **Pacharawongsakda, E.** *et al.* Predict Subcellular Locations of Singleplex and Multiplex Proteins by Semi-Supervised Learning and Dimension-Reducing General Mode of Chou's PseAAC, *IEEE Transactions on Nanobioscience*, **2013**, *12*, 311-320.

[32] Xiao, X. *et al.* iAMP-2L: A two-level multi-label classifier for identifying antimicrobial peptides and their functional types, *Anal. Biochem.*, **2013**, *436*, 168-177.

[33] Xiao, X. *et al.* iGPCR-Drug: A web server for predicting interaction between GPCRs and drugs in cellular networking, *PLoS ONE*, **2013**, *8*, e72234.

[34] Ding, H. *et al.* iCTX-Type: A sequence-based predictor for identifying the types of conotoxins in targeting ion channels, *BioMed Research International (BMRI)*, **2014**, *2014*, 286419.

[35] Fan, Y. N. *et al.* iNR-Drug: Predicting the interaction of drugs with nuclear receptors in cellular networking, *Intenational Journal of Molecular Sciences (IJMS)*, **2014**, *15*, 4915-4937.

[36] Hajisharifi, Z. *et al.* Predicting anticancer peptides with Chou's pseudo amino acid composition and investigating their mutagenicity via Ames test, *J. Theor. Biol.*, **2014**, *341*, 34-40.

[37] Li, L. *et al.* Prediction of bacterial protein subcellular localization by incorporating various features into Chou's PseAAC and a backward feature selection approach, *Biochimie*, **2014**, *104*, 100-107.

[38] Ahmad, S. *et al.* Identification of heat shock protein families and J-protein types by incorporating dipeptide composition into Chou's general PseAAC, *Computer methods and programs in biomedicine*, **2015**, *122*, 165-174.

[39] Ali, F. *et al.* Classification of membrane protein types using voting feature interval in combination with Chou's pseudo amino acid composition, *J. Theor. Biol.*, **2015**, *384*, 78-83.

[40] Dehzangi, A. *et al.* Gram-positive and Gram-negative protein subcellular localization by incorporating evolutionary-based descriptors into Chou's general PseAAC, *J. Theor. Biol.*, **2015**, *364*, 284-294.

[41] Fan, G. L. *et al.* DSPMP: Discriminating secretory proteins of malaria parasite by hybridizing different descriptors of Chou's pseudo amino acid patterns, *J. Comput. Chem.*, **2015**, *36*, 2317-2327.

[42] Khan, Z. U. *et al.* Discrimination of acidic and alkaline enzyme using Chou's pseudo amino acid composition in conjunction with probabilistic neural network model, *J. Theor. Biol.*, **2015**, *365*, 197-203.

[43] Kumar, R. *et al.* Prediction of beta-lactamase and its class by Chou's pseudo-amino acid composition and support vector machine, *J. Theor. Biol.*, **2015**, *365*, 96-103.

[44] Liu, B. *et al.* PseDNA-Pro: DNA-binding protein identification by combining Chou's PseAAC and physicochemical distance transformation, *Molecular Informatics*, **2015**, *34*, 8-17

[45] Sanchez, V. *et al.* A new signal characterization and signal-based Chou's PseAAC representation of protein sequences, *Journal of bioinformatics and computational biology*, **2015**, 1550024.

[46] Chou, K. C. Pseudo amino acid composition and its applications in bioinformatics, proteomics and system biology, *Current Proteomics*, **2009**, *6*, 262-274.

[47] Zhong, W. Z. *et al.* Molecular science for drug development and biomedicine, *Intenational Journal of Molecular Sciences*, **2014**, *15*, 20072-20078.

[48] Kobayashi, Y. *et al.* DNA methylation profiling reveals novel biomarkers and important roles for DNA methyltransferases in prostate cancer, *Genome Research*, **2011**, *21*, 1017-1027.

[49] Cantara, W. A. *et al.* The RNA Modification Database, RNAMDB: 2011 update, *Nucleic Acids Res.*, **2011**, *39*, D195-201.

[50] Liu, Z. *et al.* iDNA-Methyl: Identifying DNA methylation sites via pseudo trinucleotide composition, *Analytical Biochemistry (also, Data in Brief, 2015, 4: 87-89)*, **2015**, *474*, 69-77.

[51] Chen, W. *et al.* iRNA-Methyl: Identifying N6-methyladenosine sites using pseudo nucleotide composition *Analytical Biochemistry (also, Data in Brief, 2015, 5: 376-378)*, **2015**, *490*, 26-33.

[52] Cai, L. *et al.* Identification of Proteins Interacting with Human SP110 During the Process of Viral Infections, *Medicinal Chemistry*, **2011**, *7*, 121-126.

[53] Chen, W. *et al.* iNuc-PhysChem: A Sequence-Based Predictor for Identifying Nucleosomes via Physicochemical Properties, *PLoS ONE*, **2012**, *7*, e47843.

[54] Cai, L. *et al.* Prostate Cancer with Variants in CYP17 and UGT2B17 Genes: A Meta-Analysis, *Protein & Peptide Letters*, **2012**, *19*, 62-69.

[55] Chen, W. *et al.* iRSpot-PseDNC: identify recombination spots with pseudo dinucleotide composition *Nucleic Acids Res.*, **2013**, *41*, e68.

[56] Chen, W. *et al.* iTIS-PseTNC: a sequence-based predictor for identifying translation initiation site in human genes using pseudo trinucleotide composition, *Anal. Biochem.*, **2014**, *462*, 76-83.

[57] Chu, W. Z. *et al.* Apolipoprotein E gene variants of Alzheimer's disease and vascular dementia patients in a community population of nanking, *Med Chem* **2014**, *10*, 783-788.

[58] Chen, W. *et al.* iSS-PseDNC: identifying splicing sites using pseudo dinucleotide composition, *Biomed Research International (BMRI)*, **2014**, *2014*, 623149.

[59] Guo, S. H. *et al.* iNuc-PseKNC: a sequence-based predictor for predicting nucleosome positioning in genomes with pseudo k-tuple nucleotide composition, *Bioinformatics*, **2014**, *30*, 1522-1529.

[60] Lin, H. *et al.* iPro54-PseKNC: a sequence-based predictor for identifying sigma-54 promoters in prokaryote with pseudo k-tuple nucleotide composition, *Nucleic Acids Res.*, **2014**, *42*, 12961-12972.

[61] Qiu, W. R. *et al.* iRSpot-TNCPseAAC: Identify recombination spots with trinucleotide composition and pseudo amino acid components, *Int J Mol Sci (IJMS)*, **2014**, *15*, 1746-1766.

[62] Cai, L. *et al.* Gestational influenza increases the risk of psychosis in adults, *Medicinal Chememistry*, **2015**, *11*, 676-682.

[63] Liu, B. *et al.* Identification of real microRNA precursors with a pseudo structure status composition approach, *PLoS ONE*, **2015**, *10*, e0121501.

[64] Liu, B. *et al.* iMiRNA-PseDPC: microRNA precursor identification with a pseudo distance-pair composition approach, *Journal of Biomolecular Structure & Dynamics (JBSD), doi:10.1080/07391102.2015.1014422*, **2015**.

[65] Liu, B. *et al.* iEnhancer-2L: a two-layer predictor for identifying enhancers and their strength by pseudo k-tuple nucleotide composition, *Bioinformatics*, **2015**, doi:10.1093/bioinformatics/btv1604.

[66] Liu, B. *et al.* Identification of microRNA precursor with the degenerate K-tuple or Kmer strategy *Journal of Theoretical Biology*, **2015**, *385*, 153-159.

[67] Liu, J. *et al.* Association of EGF rs4444903 and XPD rs13181 polymorphisms with cutaneous melanoma in Caucasians, *Medicinal Chemistry*, **2015**, *11*, 551-559.

[68] Cai, L. *et al.* Modulation of cytokine network in the comorbidity of schizophrenia and tuberculosis, *Curr Top Med Chem*, **2016**, *16*, 655-665.

[69] Zhu, Y. *et al.* Antithrombin is an importantly inhibitory role against blood clots, *Curr Top Med Chem*, **2016**, *16*, 666-674.

[70] Chen, W. *et al.* Pseudo nucleotide composition or PseKNC: an effective formulation for analyzing genomic sequences, *Mol BioSyst*, **2015**, *11*, 2620-2634.

[71] Chen, W. *et al.* PseKNC: a flexible web-server for generating pseudo K-tuple nucleotide composition, *Anal. Biochem.*, **2014**, *456*, 53-60.

[72] Chen, W. *et al.* PseKNC-General: a cross-platform package for generating various modes of pseudo nucleotide compositions, *Bioinformatics*, **2015**, *31*, 119-120.

[73] Liu, B. *et al.* repDNA: a Python package to generate various modes of feature vectors for DNA sequences by incorporating user-defined physicochemical properties and sequence-order effects, *Bioinformatics*, **2015**, *31*, 1307-1309.

[74] Liu, B. *et al.* repRNA: a web server for generating various feature vectors of RNA sequences *Molecular Genetics and Genomics, DOI:10.1007/s00438-015-1078-7*, **2015**.

[75] Liu, B. *et al.* Pse-in-One: a web server for generating various modes of pseudo components of DNA, RNA, and protein sequences *Nucleic Acids Res.*, **2015**, *43*, W65-W71.

[76] Wang, J. F. *et al.* Review: Pharmacogenomics and personalized use of drugs, *Current Topics of Medicinal Chemistry*, **2008**, *8*, 1573-1579.

[77] Wang, J. F. *et al.* 3D structure modeling of cytochrome P450 2C19 and its implication for personalized drug design, *Biochem Biophys Res Commun (BBRC) (Corrigendum: ibid, 2007, Vol.357, 330)*, **2007**, *355*, 513-519.

[78] Wang, J. F. *et al.* Molecular modeling of two CYP2C19 SNPs and its implications for personalized drug design, *Protein & Peptide Letters*, **2008**, *15*, 27-32.

[79] Wang, J. F. *et al.* Review: Structure of cytochrome P450s and personalized drug, *Current Medicinal Chemistry*, **2009**, *16*, 232-244.

[80] Wang, J. F. *et al.* Structure of cytochrome p450s and personalized drug, *Curr Med Chem*, **2009**, *16*, 232-244.

[81] Knowles, J. *et al.* A guide to drug discovery: Target selection in drug discovery, *Nat Rev Drug Discov*, **2003**, *2*, 63-69.

[82] Chou, K. C. *et al.* Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS. (Erratum: ibid., 2003, Vol.310, 675), *Biochem Biophys Res Comm (BBRC)*, **2003**, *308*, 148-151.

[83] Zhou, G. P. *et al.* NMR studies on how the binding complex of polyisoprenol recognition sequence peptides and polyisoprenols can modulate membrane structure, *Current Protein and Peptide Science*, **2005**, *6*, 399-411.

[84] Du, Q. S. *et al.* Molecular modelling and chemical modification for finding peptide inhibitor against SARS CoV Mpro, *Anal. Biochem.*, **2005**, *337*, 262-270.

[85] Huang, R. B. *et al.* An in-depth analysis of the biological functional studies based on the NMR M2 channel structure of influenza A virus, *Biochem. Biophys Res Comm. (BBRC)*, **2008**, *377*, 1243-1247.

[86] Du, Q. S. *et al.* Energetic analysis of the two controversial drug binding sites of the M2 proton channel in influenza A virus, *J. Theor. Biol.*, **2009**, *259*, 159-164.

[87] Wei, H. *et al.* Investigation into adamantane-based M2 inhibitors with FB-QSAR, *Medicinal Chemistry*, **2009**, *5*, 305-317.

[88] Du, Q. S. *et al.* Designing inhibitors of M2 proton channel against H1N1 swine influenza virus, *PLoS ONE*, **2010**, *5*, e9388.

[89] Wang, S. Q. *et al.* Insights from investigating the interaction of oseltamivir (Tamiflu) with neuraminidase of the 2009 H1N1 swine flu virus, *Biochemical and Biophysical Research Communications (BBRC)*, **2009**, *386*, 432-436.

[90] Chou, K. C. Review: Structural bioinformatics and its impact to biomedical science, *Current Medicinal Chemistry*, **2004**, *11*, 2105-2134.

[91] Liao, Q. H. *et al.* Docking and Molecular Dynamics Study on the Inhibitory Activity of Novel Inhibitors on Epidermal Growth Factor Receptor (EGFR), *Medicinal Chemistry*, **2011**, *7*, 24-31.

[92] Li, X. B. *et al.* Novel Inhibitor Design for Hemagglutinin against H1N1 Influenza Virus by Core Hopping Method, *PLoS One*, **2011**, *6*, e28111.

[93] Ma, Y. *et al.* Design novel dual agonists for treating type-2 diabetes by targeting peroxisome proliferator-activated receptors with core hopping approach, *PLoS One*, **2012**, *7*, e38546.

[94] Wang, J. F. *et al.* Insights from modeling the 3D structure of New Delhi metallo-beta-lactamase and its binding interactions with antibiotic drugs, *PLoS ONE* **2011**, *6*, e18414.

[95] Wang, J. F. *et al.* Insights into the Mutation-Induced HHH Syndrome from Modeling Human Mitochondrial Ornithine Transporter-1, *PLoS One*, **2012**, *7*, e31048.

[96] Berardi, M. J. *et al.* Mitochondrial uncoupling protein 2 structure determined by NMR molecular fragment searching, *Nature*, **2011**, *476*, 109-113.

[97] Schnell, J. R. *et al.* Structure and mechanism of the M2 proton channel of influenza A virus, *Nature*, **2008**, *451*, 591-595.

[98] OuYang, B. *et al.* Unusual architecture of the p7 channel from hepatitis C virus *Nature* **2013** *498*, 521-525.

[99] Call, M. E. *et al.* The structural basis for intramembrane assembly of an activating immunoreceptor complex, *Nature Immunology*, **2010**, *11*, 1023-1029.

[100] Wang, J. *et al.* Solution structure and functional analysis of the influenza B proton channel, *Nature Structural and Molecular Biology*, **2009**, *16*, 1267-1271.

[101] Bruschweiler, S. *et al.* Substrate-modulated ADP/ATP-transporter dynamics revealed by NMR relaxation dispersion, *Nat Struct Mol Biol* **2015**, *22*, 636-641.

[102] Berardi, M. J. *et al.* Fatty acid flippase activity of UCP2 is essential for its proton transport in mitochondria, *Cell metabolism*, **2014**, *20*, 541-552.

[103] Chou, K. C. Modelling extracellular domains of GABA-A receptors: subtypes 1, 2, 3, and 5, *Biochemical and Biophysical Research Communications (BBRC)*, **2004**, *316*, 636-642.

[104] Chou, K. C. Insights from modelling the 3D structure of the extracellular domain of alpha7 nicotinic acetylcholine receptor, *Biochemical and Biophysical Research Communication (BBRC)*, **2004**, *319*, 433-438.

[105] Chou, K. C. Molecular therapeutic target for type-2 diabetes, *Journal of Proteome Research*, **2004**, *3*, 1284-1288.

[106] Chou, K. C. Coupling interaction between thromboxane A2 receptor and alpha-13 subunit of guanine nucleotide-binding protein, *Journal of Proteome Research*, **2005**, *4*, 1681-1686.

[107] Chou, K. C. Insights from modelling three-dimensional structures of the human potassium and sodium channels, *Journal of Proteome Research*, **2004**, *3*, 856-861.

[108] Chou, K. C. Insights from modelling the tertiary structure of BACE2, *Journal of Proteome Research*, **2004**, *3*, 1069-1072.

[109] Chou, K. C. Insights from modeling the 3D structure of DNA-CBF3b complex, *Journal of Proteome Research*, **2005**, *4*, 1657-1660.

[110] Chou, K. C. Modeling the tertiary structure of human cathepsin-E, *Biochem. Biophys. Res. Commun. (BBRC)*, **2005**, *331*, 56-60.

[111] Carlacci, L. *et al.* A heuristic approach to predicting the tertiary structure of bovine somatotropin, *Biochemistry*, **1991**, *30*, 4389-4398.

[112] Chou, K. C. Energy-optimized structure of antifreeze protein and its binding mechanism, *J. Mol. Biol.*, **1992**, *223*, 509-517.

[113] Chou, K. C. The convergence-divergence duality in lectin domains of the selectin family and its implications, *FEBS Lett.*, **1995**, *363*, 123-126.

[114] Xiao, X. *et al.* iCDI-PseFpt: Identify the channel-drug interaction in cellular networking with PseAAC and molecular fingerprints, *J. Theor. Biol.*, **2013**, *337C*, 71-79.

[115] Xiao, X. *et al.* iDrug-Target: predicting the interactions between drug compounds and target proteins in cellular networking via the benchmark dataset optimization approach, *Journal of Biomolecular Structure & Dynamics (JBSD)*, **2015**, *33*, 2221-2233.

[116] Xiao, X. *et al.* Predict drug-protein interaction in cellular networking, *Current Topics in Medicinal Chemistry*, **2013**, *13*, 1707-1712.

[117] Sirois, S. *et al.* Assessment of chemical libraries for their druggability, *Computational Biology & Chemistry*, **2005**, *29*, 55-67.

SciForum
Mol2Net

Fluorinated Nucleosides

Mohamed Ibrahim Elzagheid*
Chemical and Processing Engineering Technology Department (Industrial Chemistry), Jubail Industrial College, PO Box 10099, Jubail Industrial City 31961, Kingdom of Saudi Arabia
Email: elzagheid_m@jic.edu.sa, elzagheid66@yahoo.com

Published: 4 December 2015

Abstract: In this mini review the synthesis of base- and sugar-fluorinated nucleosides is discussed and the use of different fluorinating agents is briefly elaborated within the text. Introduction of fluorine substituent into pyrimidine and purine nucleosides has surely led to a change in the overall chemical behavior. In fact, there are many examples of the sugar fluorinated nucleosides that make a great impact on chemistry, biochemistry, and drug discovery.

Keywords: Nucleoside, fluorinated nucleosides, fluorinating agents.

Introduction

Chemists introduced fluoro group into sugars and nucleosides by different methods and this let for the formation of fluorine carbon bond. Incorporation of fluorine atom into nucleosides especially at the sugar moiety added interesting biological activities to the nucleosides. This is may be due to the electronegativity of fluorine and the strength of the carbon-fluorine bond. Fluorinated nucleosides exhibit interesting physical-chemical properties as nucleosides[1] or as monomers in oligonucleotide synthesis.[2] The synthesis of fluorinated nucleosides (**figure 1, 1-9**) can be achieved by selective fluorination of nucleosides or by glycosylation of nucleobases or fluoro-nucleobases with fluoro-sugar derivatives.[3-8]

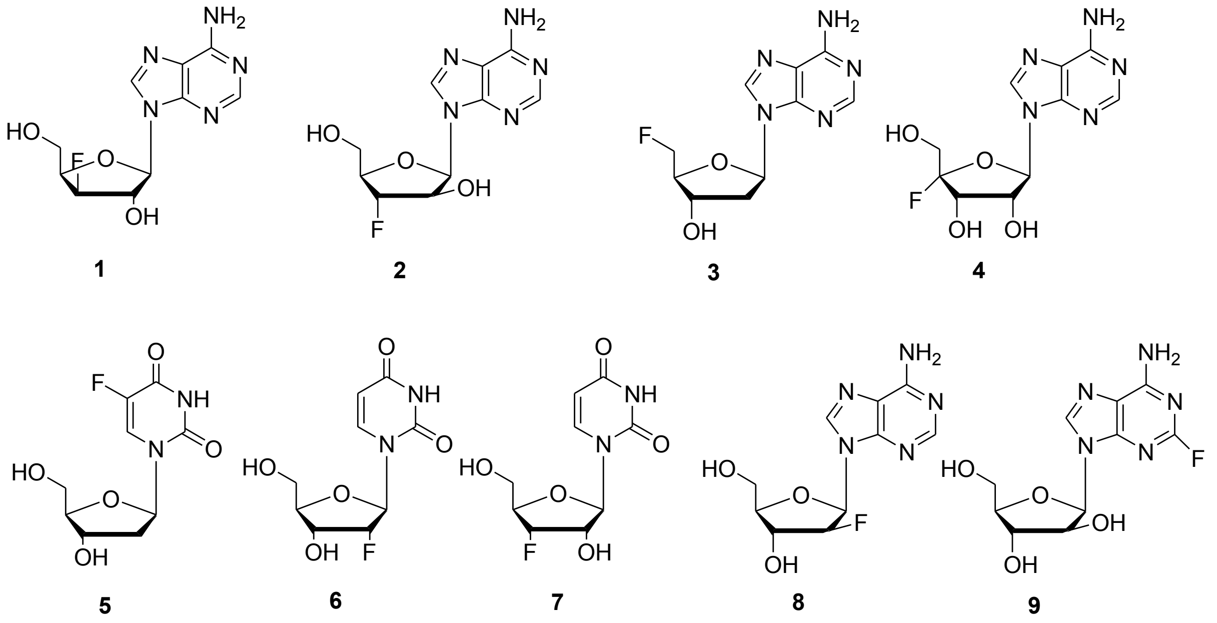
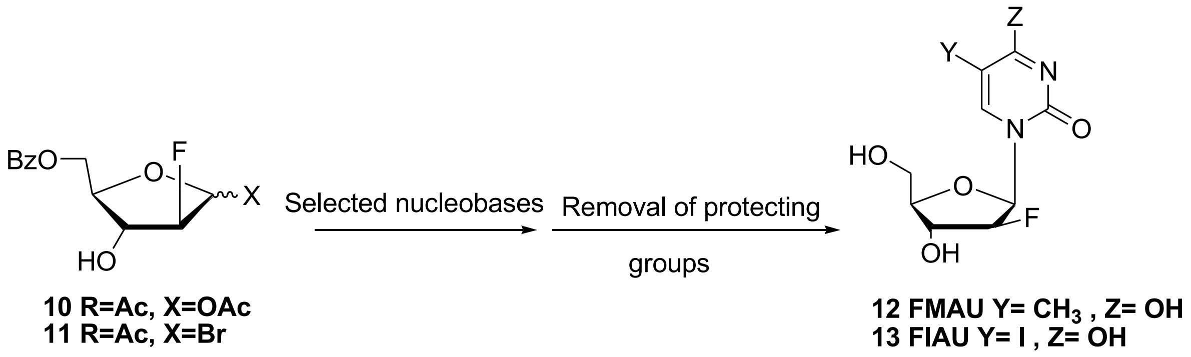


Figure 1

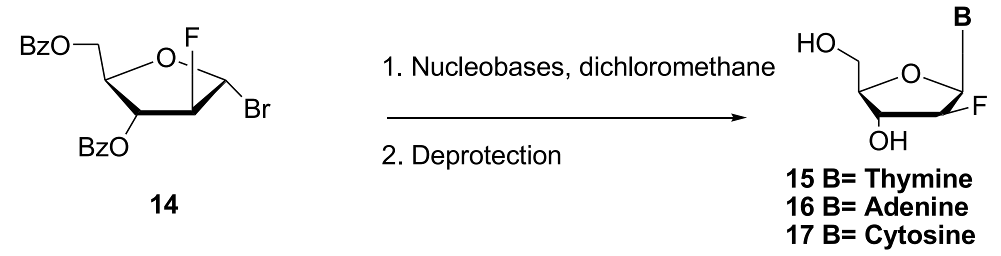
Synthesis of Modified Fluoronucleosides

Different methods were developed for the synthesis of 2'-fluornucleosides. Among those methods the one that was published by Watanabe *etal.*[9] This synthetic method involves the coupling of 2-deoxy-2-fluoro-D-arabinose **10** and 1-bromo-2-deoxy-2-fluoro-**D**-arabinose **11** with selected nucleobases to give fluoronucleosides **12** and **13** (**Scheme 1**).



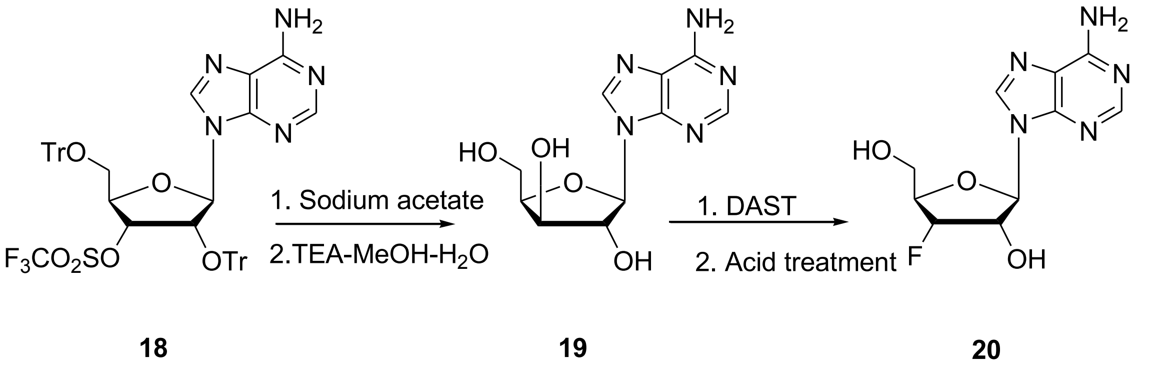
Scheme 1

2'-Fluorinated nucleosides, namely 2'-deoxy-2'-fluoro arabinonucleosides (araF-nucleosides)[10,11] were used as building blocks for the synthesis of 2'-deoxy-2'-fluoro arabinonucleic acid (2'-F'ANA),[12-14] a very promising antisense oligonucleotides. Here the 2'-fluoro nucleosides were prepared via condensation of the silylated nucleic acids bases; silylated *N*-acetyl cytosine or silylated thymine or *N*-benzoylated adenine with 2-deoxy-2-fluoro-3, 5-di-O-benzoyl-α-**D**-arabinofuranosyl bromide **14**. Deprotection of the produced the desired nucleosides **15-17** in good yield (**Scheme 2**).



Scheme 2

Synthesis of 3'-fluorinatednucleosides has been reported by Hansske *et al.*[15] This involves the treatment of the nucleoside triflate **18** with sodium acetate followed by mild hydrolysis of the later nucleoside in triethylamine-methanol-water mixture gave nucleoside **19** in good yield. When nucleoside **19** was treated with DAST [(diethylaminosulfurtrifluoride, Et_2NSF_3] followed by acid treatment nucleoside **20** was obtained in good yield (**Scheme 3**).



Scheme 3

Fluorinating Agents

Fluoride ion is the smallest anion with the largest negative charge density. In general it acts as a hydrogen-bond acceptor rather than as a nucleophilic agent. Depending on the reaction environment, the fluoride ion can act either as a poor nucleophile or as a good nucleophile. Activation of alcohols with good leaving groups, such as mesylate, tosylate or triflate; followed by a S_N2 substitution by a fluoride ion has become a standard method to replace OH with F. Fluorinating agents can be classified into nucleophilic reagents and electrophilic reagents. Fluorinating reagents (**Figure 2**) that are usually used for the fluorination of the hydroxyl groups in sugars or nucleosides are listed below:

i. **Pyridinium poly (hydrogen fluoride, PPHF or Py.nHF) - Olah's reagent[16]**

The Olah reagent is a nucleophilic fluorinating agent. It consists of a mixture of 70 % hydrogen fluoride and 30% pyridine. Secondary and tertiary alcohols can be converted to the corresponding fluorides by this reagent.

ii. **Diethylaminosulfur trifluoride, Et_2NSF_3- DAST reagent**

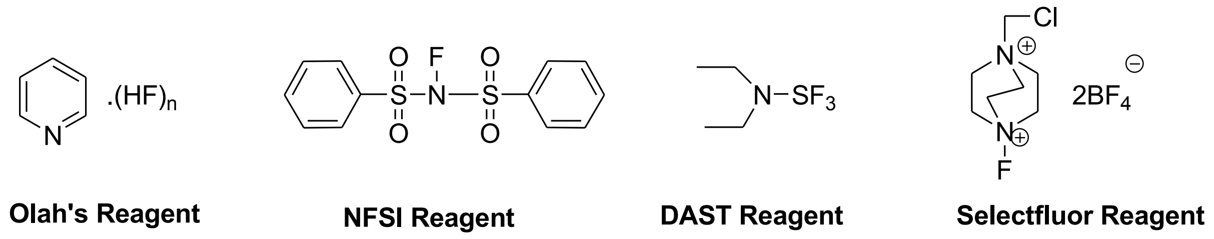
DAST was prepared by the reaction of SF_4 with diethylaminotrimethylsilane.[17]It is the most versatile reagent in nucleoside chemistry for a one-step exchange of the hydroxyl group by fluorine via S_N2 displacement. This reaction occurs with a complete inversion of configuration. It can replace primary, secondary and tertiary hydroxyl groups with fluorine in very good yields.

iii. **1-Chloromethyl-4-fluoro-1, 4-diazoniabicyclo [2.2.2] octane bis (tetrafluoroborate) Selectfluor reagent**

Selectfluor has much higher reactivity than NFSI. It is a stable, easily handled, solid electrophilic fluorinating agent. It is an equivalent of F_2, but much more effective and selective. In nucleoside chemistry, selectfluor is widely used to introduce a fluorine atom into heterocyclic bases *via* electrophilic substitution. Selectfluor can also selectively fluorinate certain sugar moieties, which possess electron-rich double bonds *via* an electrophilic addition.

iv. N-fluorobenzenesulfonimide-NFSI reagent

Although N-fluorosulfonamides are fairly weak fluorinating reagents, N-fluorosulfonimides, such as N-fluorobenzenesulfonimide (NFSI), are very effective and in common use.[18]



Olah's Reagent **NFSI Reagent** **DAST Reagent** **Selectfluor Reagent**

Figure 2

Conclusions

In this review, synthetic methods of certain fluorinated nucleosides have been covered. Introduction of fluorine atom, as a mimic of hydrogen or hydroxyl group into the sugar moiety of the nucleoside dramatically changes the physical-chemical properties of the nucleosides. The two main tactics that have been employed for the synthesis of fluorinated nucleosides are the installation of fluorine atom(s) into pre-modified precursor sugars before the introduction of nucleic bases and/or the regio- or stereo-selective introduction of fluorine atom(s) into suitably modified nucleoside derivatives. Among the above mentioned fluorinating agents, DAST seems to be the most commonly used for the fluorination of nucleosides. I hope that, with this mini review, I have provided an appropriate short description of the synthesis of fluorinated nucleosides and the most versatile fluorinating agents.

Acknowledgment

Author is thankful to Jubail Industrial College management for their support.

References

1. Ma T., Lin J.-S. Newton M.G., Chang Y.-C., Chu C.K. **1997**, *J. Med. Chem.*, **40**, 2750-2754.

2. Damha M.J., Wilds C.J., Novonha A., Brunker I., Brokow G., Arion D., Parniak M. A. **1998**, *J. Am. Chem. Soc.*, **120**, 12976-12977.

3. Elzagheid M.I., Viazovkina E., Damha M.J. **2002**, *Current protocols in Nucleic Acids Chemistry*, 1.7.1-1.7.19.

4. Wilds C.J., Damha M.J. **2000**, *Nucl. Acids Res*., **28**, 3625-3635

5. Pankiewicz K.W. **2000**, *Carbohydr. Res*., **327**, 87-105.

6. Kodama T., Matsuda A., Shuto S. **2006**, *Nucleic Acids symposium series*, **50**, 3-4.

7. Herdwjn P., Van Aerschot A., Kerremans L. **1989**, *Nucleosides, Nucleotides*, **8(1),** 65-96.

8. Takamatsa S., Katayama S., Hirose N., Delock E., Schelkens G., Demillequand M., Brepoels J., Izawa K. **2002**. *Nucleosides, Nucleotides, Nucleic Acids*, **21 (11, 12),** 849-861.

9. Watanabe K.A, Su T.-L, Klein R.S., Chu C.K., Matsuda A., Chun M.W., Lopez C., Fox J. **1983**, *J. Med. Chem.*, **26**, 152-156.

10. Elzagheid M.I., Viazovkina E., Damha M.J. **2002**, *Current Protocols in nucleic acids chemistry*, 1.7.1-1.7.19.

11. Elzagheid M.I., Viazovkina E., Damha M.J. **2003**, *Nucleosides, Nucleotides, Nucleic Acids*, **22**, 1339-1342.

12. Damha M.J., Wilds C.J., Novonha A., Brunker I., Brokow G., Arion D., Parniak M.A. **1998***, J. Am. Chem. Soc.,* **120**, 12976-12977.

13. Wilds C.J., Damha M.J. **2000**, *Nucl. Acids Res*., **28**, 3625-3635.

14. Lok C.N., Viazovkina E., Min K.L., Nagy E., Wilds C.J., Damha M.J. Parniak M.A. **2002**, *Biochemistry*, **41**, 3457-3467.

15. Hansske F., Madej D., Robins M. J. **1984**, *Tetrahedron*, **40**, 125-135.

16. Olah G. A., Welch J. T., Yankar Y. D., Nojima M., Kerekes I., Olah J. A. **1979**, *J. Org. Chem*.,**44**, 3872-3881.

17. Middleton W. J., Bingham E. M. **1979**, Org. Synth., **57**, 50.

18. Liu P., Sharon A., Chu C.K. **2008**, *J. Fluor. Chem*., **129 (9)**, 743-766.

SciForum
Mol2Net

Perturbation Theory Modeling of Intramolecular Carbolithiation Reactions

Asier Gómez-SanJuan [1], Sonia Arrasate [1], Humberto González-Díaz [1,2,*], Nuria Sotomayor [1], Esther Lete [1]

[1] Department of Organic Chemistry II, University of the Basque Country UPV/EHU, 48940, Bilbao, Spain; E-Mails: asgo1982@hotmail.com; sonia.arrasate@ehu.eus; humberto.gonzalezdiaz@ehu.eus; nuria.sotomayor@ehu.es; esther.lete@ehu.eus.

[2] IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain;

* Author to whom correspondence should be addressed; E-Mail: humberto.gonzalezdiaz@ehu.eus; Tel.: +34 94 601 3547; Fax: +34 94 601 2748.

Published: 4 December 2015

Abstract: PT-QSRR models are Quantitative Structure-Reactivity Relationship (QSRR) models based on Perturbation Theory (PT) that may be useful for multi-objective optimization in organic synthesis. In this communication, we summarize some of the more important results and conclusions obtained in our previous research / review paper about PT-QSRR models published in *Curr. Top. Med. Chem.*, **2013**, *13*, (5), 1713-1741. I this previous work, firstly we reviewed general aspects and applications of both perturbation theory and QSPR models. Secondly, we formulate a general-purpose perturbation theory for multiple-boundary QSPR problems. In this previous work, we developed a new QSPR-Perturbation theory model that classify correctly >100,000 pairs of intra-molecular carbolithiations with 75-95% of Accuracy (Ac), Sensitivity (Sn), and Specificity (Sp). The model predicts probabilities of variations in the yield and enantiomeric excess of reactions due to at least one perturbation in boundary conditions (solvent, temperature, temperature of addition, or time of reaction). The model also account for changes in chemical structure (connectivity structure and/or chirality patterns in substrate, product, electrophile agent, organolithium, and ligand of the asymmetric catalyst).

Keywords: Organometallic addition; Carbolithiation reactions; Asymmetric synthesis; Perturbation theory, QSRR models

1. Introduction

In the wider sense Perturbation Theory (PT) methods starts with a known exact solution of a problem and continue adding "small" terms to the mathematical description in order to approach a solution to a related problem without known exact solution. On the other hand, solving Quantitative Structure-Property Relationships (QSPR) problems may be important to study the chemical reactivity of compounds in organic synthesis. QSRR techniques involve: (1) numerical codification of molecular structure, and stage (2) search of a quantitative connection betw*ee*n the structure and reactivity. QM and/or Graph theory are computational techniques typical of stage (1); while Statistical and/or Machine Learning (ML) techniques are commonly used in stage (2). In this work, we give one example of the new theory to predict enantiomeric excess and yield in a set of intramolecular carbolithiation reactions.

PT-QSRR of intra-molecular carbolithiations

A particular class of reaction of high interest in organic synthesis is the so-called carbolithiation. The carbolithiation reaction offers an attractive pathway for the efficient construction of new carbon-carbon bonds by addition of an organo-lithium reagent to non-activated alkenes or alkynes, with the possibility of introducing further functionalization on the molecule by trapping the reactive organolithium intermediates with electrophiles. The intramolecular variant of this reaction has been applied mainly with alkyl- and alkenyllithiums, though there are also some examples of cycloisomerization of alkenyl substituted aryllithiums, generated by metal-halogen exchange. In particular, this type of intramolecular carbolithiation reaction has found application in the synthesis of both carbocycles and heterocycles, with a high degr*ee* of regio- and stereoselectivity in the formation of five-membered rings, although its application to larger rings is still not general. When alkenes are used, up to two contiguous stereogenic centers may be generated, which may be controlled by using chiral ligands for lithium, and so opening new opportunities for application of this methodology to asymmetric synthesis. Bailey and Groth reported independently the intramolecular carbolithiation of *N*-allyl substituted *o*-haloanilines in the enantioselective synthesis of indolines, Barluenga described the preparation of dihydrobenzofurans in high *ee*, starting from allyl *o*-haloaryl ethers. Lete *et. al.* previously reported the preparation of 4-substituted 2-phenyltetrahydroquinolines from *N*-alkenylsubstituted 2-iodoanilines via intramolecular carbolithiation reactions. We developed a QSPR model for a very large set of input-output perturbations in the enantioselective intramolecular carbolithiation reactions via aryllithiums generated by halogen-lithium exchange (*see* **Figure 1**).

We propose herein, for the first time, a PT-QSRR model able to predict both the change in yield and enantiomeric excess for two pair of reactions after at least one change in chemical structure and/or perturbation of reaction variables of at least one of the molecules involved. After optimization of coefficients the best model found with LDA was:

$$\begin{aligned}\varepsilon_s(p_i) = {} & \mathit{-0.00105976} \cdot \mathrm{V}(\mathrm{product}_i)_{icr} \qquad (1)\\ & + \mathit{0.13496123} \cdot \Delta\mathrm{V}(\mathrm{susbstrate}_i)\\ & - \mathit{0.00000065} \cdot \Delta\mathrm{V}(\mathrm{solvent}_i)\\ & + \mathit{0.08477068} \cdot \Delta\mathrm{V}(\mathrm{electrophile}_i)\\ & - \mathit{0.00029331} \cdot \Delta\mathrm{V}(\mathrm{ligand}_i)\\ & - \mathit{0.00029331} \cdot \Delta\mathrm{V}(\mathrm{org.lithium}_i)\\ & + \mathit{0.01087415}\end{aligned}$$

Here, $\varepsilon_s(p_i)_{rr}$ is called efficiency of reference and may be either the yield of this reaction $\varepsilon_s(p_i)_{rr} = \mathrm{yld}(p_i)_{rr}$ or the enantiomeric excess of the same reaction $\varepsilon_0(p_i)_{rr} = ee(p_i)_{rr}$. The variables in the right side of the equation quantify multiple structural and physicochemical factors driven the yield and stereoselectivity of the reaction (see details on the original work). We consider the following six roles or types of molecules substrates, solvents, ligands of chiral catalysts, organolithium reactants, electrophiles, and products. As a results, we can calculate the effect of different perturbations $\Delta V(m_{qi})$ on the previous over reactivity for a given set of initial and final conditions (reaction of reference and query reaction). This is a very strong result because it determines that this is probably the first multi-objective optimization (MOOP) model for the effect of structural or condition perturbations in reactions.

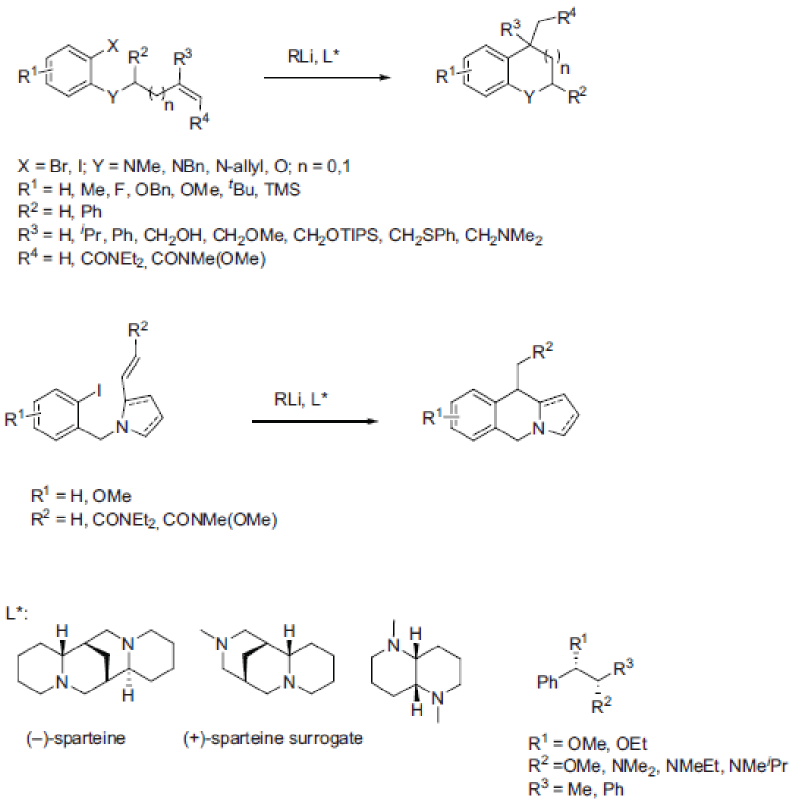

RLi, L*
X = Br, I; Y = NMe, NBn, N-allyl, O; n = 0,1
R1 = H, Me, F, OBn, OMe, tBu, TMS
R2 = H, Ph
R3 = H, iPr, Ph, CH2OH, CH2OMe, CH2OTIPS, CH2SPh, CH2NMe2
R4 = H, CONEt2, CONMe(OMe)
RLi, L*
R1 = H, OMe
R2 = H, CONEt2, CONMe(OMe)
L*:
(–)-sparteine
(+)-sparteine surrogate
R1 = OMe, OEt
R2 =OMe, NMe2, NMeEt, NMeiPr
R3 = Me, Ph

Figure 1. General scheme for intramolecular carbolithiations studied here.

4. Conclusions

It is possible to develop general models to predict the results of multiple input-output perturbations using ideas of QSPR analysis and perturbation theory. The new QSPR-Perturbation models may be used to study complex molecular systems. One the properties we can study are the yield and enantiomeric excess of reactions. These models may include perturbations in a very high number of input-output variables like (time, temperature, solvent, catalyst, assay, pharmacological experimental measures, molecular and cellular targets, and many others). The electronegativity values calculated with MARCH-INSIDE s*ee*ms to be good molecular descriptors for QSPR-Perturbation theory.

Acknowledgments

The Ministerio de Economía y Competitividad (CTQ2013-41229-P), Gobierno Vasco (IT-623-13) and Universidad del País Vasco/Euskal Herriko Unibertsitatea UPV/EHU are gratefully acknowledged for their financial support.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Cropper, W.H. *Great Physicists: The Life and Times of Leading Physicists from Galileo to Hawking*. Oxford University Press, USA, **2004**.
2. González-Díaz, H.; Arrasate, S.; Gómez-SanJuan, A.; Sotomayor, N.; Lete, E.; Besada-Porto, L.; Ruso, J.M., General Theory for Multiple Input-Output Perturbations in Complex Molecular Systems. 1. Linear QSPR Electronegativity Models in Physical, Organic, and Medicinal Chemistry. *Curr. Top. Med. Chem.,* **2013**, *13*, (5), 1713-1741.
3. Mealy, M. J.; Luderer, M. R.; Bailey, W. F.; Sommer, M. B. Effect of Ligand Structure on the Asymmetric Cyclization of Achiral Olefinic Organolithiums. *J. Org. Chem.* **2004,** *69,* 6042–6049.
4. Groth, U.; Koettgen, P.; Langenbach, P.; Lindenmaier, A.; Schütz, T.; Wiegand, M. Enantioselective Synthesis of 3,3-Disubstituted Indolines via Asymmetric Intramolecular Carbolithiation in the Presence of (-)-Sparteine. *Synlett* **2008,** 1301–1304.
5. Barluenga, J.; Fañanás, F. J.; Sanz, R.; Marcos, C. Intramolecular Carbolithiation of Allyl *o*-Lithioaryl Ethers: A New Enantioselective Synthesis of Functionalized 2,3-Dihydrobenzofurans. *Chem.–Eur. J.* **2005,** *11,* 5397–5407.
6. Arrasate, S.; Lete, S.; Sotomayor, N., Synthesis of enantiomerically enriched amines by chiral ligand mediated addition of organolithium reagents to imines *Tetrahedron: Asymmetry,* **2001**, *12*, (14), 2077-2082.
7. Martínez-Estíbalez, U.; Gómez-SanJuan, A.; García-Calvo, O.; Arrasate, S.; Sotomayor, N.; Lete, E. Intramolecular carbolithiation reactions of aryllithiums in the synthesis of carbocyclic and heterocyclic compounds. In *Targets in Heterocyclic Systems;* Attanasi, O.; Spinelli, D., Eds.; Italian Society of Chemistry: Rome, 2010; Vol. 14, pp 124–149.
8. Sotomayor, N.; Lete, E. Carbolithiation of carbon-carbon multiple bonds. In *Science of Synthesis. Knowledge Updates 2011/4;* Hall, D. G.; Ishikara, K.; Li, J. J.; Marek, I.; North, M.; Schaumann, E.; Weinreb, S. M.; Yus, M., Eds.; Thieme: Stuttgart, 2011; pp 191–251.
9. Gómez-SanJuan, A.; Sotomayor, N.; Lete, E., Inter- and intramolecular enantioselective carbolithiation reactions. *Beilstein J. Org. Chem.,* **2013**, *9*, 313-322.
10. O'Brien, P., Basic instinct: design, synthesis and evaluation of (+)-sparteine surrogates for asymmetric synthesis. *Chem Commun (Camb),* **2008**, (6), 655-667.

SciForum
MOL2NET

Editorial: MOL2NET International Conference on Multidisciplinary Sciences.

2015, 05–15 Dec., MDPI Sciforum, HQ UPV/EHU, Bilbao, http://sciforum.net/conference/mol2net-1

Published: 5 December 2015

The full title of this conference is the MOL2NET International Conference on Multidisciplinary Sciences. This is an International Conference to Foster Interdisciplinary Collaborations in Experimental Chemistry (all branches), Medicine, Nanotechnology, Data Analysis, Bioinformatics, and Networks Sciences. MOL2NET (the conference's running title) is the acronym of the lemma of the conference: **From Molecules to Networks**. This running title is inspired by the possibility of multidisciplinary collaborations in science between experimentalists and theoretical scientists; represented disciplines will encompass the molecular and biomedical sciences, social networks analysis, and beyond. More specifically, this conference aims to promote scientific synergies between groups of experimental molecular and bio-medical scientists. Relevant fields include Chemistry, all areas (Inorganic Chemistry, Organic Chemistry, Medicinal Chemistry, Analytical Chemistry, Chemical Engineering), Pharmaceutical Sciences, Pharmacology, Cancer Research, OMICS, Neurosciences, Nanosciences, Materials Science, Medicine, and Biomedical Engineering, Cancer Research. Moreover, the conference welcomes computational and social sciences experts from different areas, such as Computational Chemistry, Bioinformatics, Networks Science, Social Networks analysis, Data and Computer Sciences, Predictive analytics, Biostatistics, *etc*. The Scientific Headquarters (HQs) are in the Faculty of Science and Technology, University of Basque Country (UPV/EHU), Bizkaia. The participation and publication of communications is online via the platform SciForum of the Editorial Molecular Diversity Preservation International (MDPI), with HQ in Basel, Switzerland, and Beijing - Wuhan, China.

The conference per se is the result of the synergy between the Department of Organic Chemistry II, UPV/EHU, and IKERBASQUE, Basque Foundation for Sciences, with the Faculty of Informatics, University of Coruña (UDC). MDPI Sciforum platform (http://sciforum.net/) will publish accepted communications online. In parallel, we are editing one special issue for International Journal of Molecular Sciences (IJMS) journal of editorial MDPI (http://www.mdpi.com/). The revision process is totally independent, please contact the editorial if you are interested.

NOTES:

* The conference is Totally Online; no physical presence is needed saving traveling costs. We accept experimental works, theoretical works, or experimental-theoretic works in the areas mentioned.

* Proceedings will be Published Online, Open Access, and **Totally Free of Charges** (no cost). Online submission system is ready and submissions will be open until **Nov 30, 2015**. Please, see the following instructions:

(1) Read call for papers:
http://sciforum.net/conference/mol2net-1/page/call

(2) Read instructions and download template .doc file:
http://sciforum.net/conference/mol2net-1/page/instructions

(3) Submit short communications (2-3 pages), reviews, papers, and videos:
http://sciforum.net/user/submission_for_conference/83

(4) See our social networks. We also invite all colleagues to share the conference website through social media. We are uploading flyers, conferences, and promotional videos in different languages to the MOL2NET accounts in different social networks.

- GOOGLE+ account with +30000 viewers: http://bit.do/gmol2net
- FACEBOOK group with +8000 followers: http://bit.do/fbmol2net
- TWITTER account: @mol2net

Conference Chairman

ORGANIZERS

Conference Chairman
Prof. Humberto Gonzalez-Diaz
IKERBASQUE Professor
Department of Organic Chemistry II, UPV/EHU, 48940, Leioa, Bizkaia.

President (Computer Sciences)
Prof. Alejandro Pazos, Ph.D., D.M.
Chair and Director of Dpt. of Computer Sciences, Faculty of Informatics, University of Coruña (UDC), Coruña, Spain.

ADVISORY COMMITTEE (Members of Honor)
Prof. Fernando Cossío (IKERBASQUE PRES.)
Dept. of Organic Chemistry I, UPV/EHU, San Sebastián, Gipuzkoa. President of IKERBASQUE, Basque Foundation for Science.
http://www.ikerbasque.net/

President (Experimental Sciences)
Prof. Esther Lete
Coordinator PhD Program, Department of Organic Chemistry II, UPV/EHU, 48940, Leioa, Bizkaia.

President (Law, Ethics, Biosciences)
Full Prof CM Romeo Casabona
Full Professor of Law, Director of Law & Human Genome Chair, UPV/EHU-UDEUSTO, Bilbao.

Prof. James J Chou
Department of Biological Chemistry & Molecular Pharmacology (BCMP), Harvard Medical School, Boston, USA.

President (Experimental Sciences)
Prof. Nuria Sotomayor
Coordinator MSc Program, Department of Organic Chemistry II, UPV/EHU, 48940, Leioa, Bizkaia.

Prof. Allen B. Reitz, PhD.
CEO Fox Chase Chemical Diversity Center, PA, USA. Editor-In-Chief of Current Topics in Medicinal Chemistry.

Prof. Danail Bonchev
Director of Research, Center for the Study of Biological Complexity, Dpt. of Mathematics, Virginia Commonwealth University (VCU), Virginia, USA.

Prof. Esther Domínguez Pérez
Dept. of Organic Chemistry II, Dean of Faculty of Science and Technology, UPV/EHU, Bizkaia.

Prof. Victor M. Preciado
Raj and Neera Singh Term Assistant Professor Electrical and Systems Engineering (ESE), Penn Engineering, University of Pennsilvania, USA.

Prof. Jose María Pitarke,
Professor of Condesed Matter Physics, UPV/EHU, Director of Nanomaterials Cooperative Research Center (CICNanoGune), San Sebastian, Gipuzkoa.

Prof. Roberto I Vazquez Padron
Research Associate Professor of Surgery, Molecular and Cellular Pharmacology, Miller School of Medicine, University of Miami, USA.

Prof. Daniel Graham
Department of Chemistry, Loyola University of Chicago, USA.

Prof. (C4) Prof. Dr. Peter Langer,
Vice Director Universität Rostock Institut für Chemie Abteilung für Organische Chemie, 18059 Rostock, Germany.

Prof. Yiyu Cheng
Director of Pharmaceutical Informatics Institute, Zhejiang University (ZJU), China

Prof. Néstor Etxebarria Loizate,
Director Department of Analytical Chemistry, University of Basque Country (UPV/EHU), Leioa, Bizkaia.

Prof. Jorge Galvez,
Department of Physical Chemistry, Faculty of Pharmacy, Universitat de Valencia, Spain.

Prof. Fernando Martin Sanchez, PhD.
Director and Foundational Chair of Health Informatics at Melbourne Medical School, Professor Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Australia.

Prof. Mario Piris,
IKERBASQUE Professor
Faculty of Chemistry, University of the Basque Country UPV/EHU, Donostia International Physics Center (DIPC), P.K. 1072, 20080 Donostia, Euskadi, Spain.

Prof. Claudio Palomo Nicolau
Director Dept. of Organic Chemistry I, UPV/EHU, San Sebastián, Gipuzkoa.

Prof. Ernesto Estrada
Dept. of Mathematics and Statistics, Physics, Chair in Complexity Sciences, Institute of Complexity Systems, University of Strathclyde.

Prof. Jesús Jiménez Barbero
IKERBASQUE Professor
Scientific Director of Center for Cooperative Research in Biosciences (CICBiogune), Bizkaia.

Prof. Jose A. Irabien Gulias,
Department of Chemical and Biomolecular Engineering ETSIIT, University of Cantabria, Spain.

Prof. Roberto Todeschini
Professor of Chemometrics, Department of Environmental Sciences of the University of Milano-Bicocca, Milano, Italy.

Prof. Mabel Loza,
Department of Pharmacology, University of Santiago de Compostela (USC), Santiago, Spain.

Prof. Pilar Goya,
Institute of Medicinal Chemistry (IQM), CSIC, Juan de la Cierva st., Madrid.

Prof. Carmen Cadarso-Suarez,
Unit of Biostatistics, Department of Statistics and Operations Research, School of Medicine, University of Santiago de Compostela (USC), Spain.

Prof. Yagamare Fall,
Department of Organic Chemistry, University of Vigo (UVIGO), Vigo, Spain. Africa Editor of Current Topics Medicinal Chemistry.

Prof. Ramon García Domenech,
Department of Physicalchemistry, Faculty of Pharmacy, Universitat de Valencia, Spain.

Prof. Bairong Shen
Director of Center of Complex Systems, University of Soochow (SUDA), PRC, China. China Coordinator of SUDA-UVP/EHU Collaboration Agreement.

Prof. Ramón J. Estévez Cabanas
Director Dept. of Organic Chemistry, USC, Santiago de Compostela, Spain.

Prof. Subhash Chandra Basak,
Senior Scientist and Adjunct Professor, Department of Chemistry & Biochemistry, University of Minnesota Duluth, MN, USA.

Prof. Luis M. Liz-Marzán
IKERBASQUE Professor
Scientific Director of Center for Cooperative Research in Biomaterials (CICbiomaGUNE), Gipuzkoa.

Prof. Francesc Illas Riera
Director of Institute of Theoretical and Computational Chemistry (IQTCUB), Dpt. Physical Chemistry, UB, Barcelona.

Prof. Javier Luque Garriga,
Department of Physical Chemistry, Faculty of Pharmacy and Institut of Biomedicine (IBUB), Universitat de Barcelona.

Prof. F.M. Ubeira, M.D., Ph.D.
Department of Microbiology and Parasitology, University of Santiago de Compostela (USC), Santiago, Spain.

Prof. Luis Manuel Leon Isidro,
Department of Physical Chemistry, University of Basque Country (UPV/EHU), Leioa, Bizkaia

Prof. Luis Lezama,
Department of Inorganic Chemistry, University of Basque Country (UPV/EHU), Leioa, Bizkaia.

Prof. Eugenio Uriarte,
Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, USC, Spain.

Prof. Javier Meana, M.D., Ph.D.
Department of Pharmacology, Faculty of Medicine, University of the Basque Country (UPV/EHU), Bizkaia.

Prof. Victor Maojo, M.D. Ph.D.,
Head of Biomedical Informatics Group, Polytechnic University of Madrid, Spain.

HEADS OF HOSPITAL MEDICAL SERVICES, ADVISORY COMMITTEE (Members of Honor)

Prof. Ana Gonzalez-Pinto Arrillaga, M.D., Ph.D.
Department of Neurosciences, University of the Basque Country (UPV/EHU), Head of Psychiatry Research of Osakidetza (Basque Public Health System), Head of Medical Psychiatry Service, University Hospital Santiago Apostol de Vitoria-Gasteiz, Vitoria.

Assoc. Prof. Jose Ramon Rey Caeiro, M.D.,
Head of Traumatology and Orthopedic Surgery, University Hospital of Santiago de Compostela (USC). SERGAS, Xunta de Galicia, University of Santiago de Compostela (USC).

Prof. A Rodríguez-Antigüedad Zarrans, MD. PhD., Head of Neurology Hospital of Basurto, Bilbao. Prof. Dept. of Neuroscience, Faculty of Medicine UPV/EHU, Leioa, Bizkaia. President Spain Society of Neurology (SEN).

M.D. Javier Castro Alvariño,
Head of Department of Gastroenterology, Ferrol University Hospital Complex, SERGAS, Xunta de Galicia, Ferrol, Spain.

Prof. Mariano Provencio, PhD., D.M.
Head of Medical Oncology Service, Universitary Hospital Puerta de Hierro (HUPH), Professor of Autonomous University of Madrid (UAM), Madrid, Spain.

CHAIRMANS OF SISTER CONFERENCES, ADVISORY COMMITTEE (Members of Honor)

Prof. Kuo-Chen Chou
Head and Founder of Gordon Life Science Institue, USA.
Chairman of The 10th International Conference on Bioinformatics and Biomedical Engineering (iCBBE 2016)

Prof. Julio Seijas Vasquez,
Department of Organic Chemistry, University of Santiago de Compostela (USC), Campus Lugo, Lugo, Spain.
Chairman of Electronic Conference on Synthetic Organic Chemistry, SciForum, MDPI Switzerland. (ECSOC 2016)

Professor Ignacio Rojas Ruiz
CITIC Director, and Professor of Department of Computer Architecture and Technology, University of Granada (UGR).
Chairman of 4th International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO 2016).

ORGANIZING COMMITTEE

Coordinator (Theoretical Studies)
Assoc. Prof. Cristian Robert Munteanu,
FIC, University of Coruña (UDC), Spain.

Coordinator (Experimental Studies)
Adjunct Prof. Uxue Uria Pujana,
Department of Organic Chemistry II, University of Basque Country (UPV/EHU), Leioa, Bizkaia.

Coordinator (TICs, Legal & Biosciences)
PhD. Aliuska Duardo-Sánchez,
Department of Public Law, Faculty of Law, University of Santiago de Compostela (USC), Santiago de Compostela, 15782, Spain.

Committee Members

PhD. Jose A. Seoane,
Stanford Cancer Institute, Stanford School of Medicine, Stanford University, Stanford, 94305, USA.
Ph.D. Vanessa Aguiar-Pulido,
Florida International University (FIU), School of Computing and Information Sciences, Miami, FL, USA.
Ph.D. Yasset Perez-Riverol,
Post-Doc Researcher, Proteomics Services Team, PRIDE Group, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, UK.
Ph.D. Santiago Vilar,
Research Associate, Department of Systems Biology, Columbia University, USA.
PhD. Advait Apte
Scientific programmer, Department of Biology, City College of New York, NY, USA.

PhD. Hugo Gutierrez de Terán
Department of Cell and Molecular Biology, Computational and Systems Biology, Uppsala Biomedicinska Centrum BMC, Uppsala Universitet, Uppsala, Sweden.
PhD. Maykel Cruz-Monteagudo,
CIQUP/Departamento de Química e Bioquímica, Universidade do Porto, Portugal. Instituto de Investigaciones Biomédicas (IIB), Universidad de Las Américas, Quito, Ecuador.
PhD. Juan Alberto Castillo Garit,
Visting Prof. of Bioinformatics Research Carleton University, Ottawa, Canda.

PhD. Carlos Fernandez Lozano.
RNASA-IMEDIR Group, University of Coruña (UDC), Coruña University Hospital (CHUAC), Spain.
Ph.D. Oana Chis,
Technological Institute for Industrial Mathematics (ITMATI), University of Santiago de Compostela (USC), Spain.

M.D. Paula Peleteiro Higuero,
Oncology and Radiotherapy, University Hospital of Santiago de Compostela, SERGAS, Xunta de Galicia, Santiago de Compostela

Adj. Prof. Eider Aranzamendi, Ph.D.
Department of Organic Chemistry II, University of Basque Country (UPV/EHU), Bizkaia, Spain.
Ph.D. Jesus vicente De Julián Ortiz,
Staff Researcher I+D+i at Foundation Center of Innovation and Technologic Demonstration, Valencia, Spain.
Ph.D. Verónica Ortiz de Elguea,
Department of Organic Chemistry II, University of Basque Country (UPV/EHU), Leioa, Bizkaia, Spain.

Ph.D. Ricardo Riccardo Concu,
FCUP-Faculty of Sciences, University of Porto (UPORTO), Porto, Portugal.
M.D. Xavier Romero Duran,
Service of Neurology, IMQ Zorrotzaure Clinic, Bilbao, Bizkaia, Spain.

MSc. Maria Galvez-Llompart,
Research Associate, Department of Physicalchemistry, Faculty of Pharmacy, Universitat de Valencia, Spain.

M.D. Berkis Turiño Guerra,
Aralia Servicios Sociosanitarios SA, Madrid

M.D. JoséLuis Ulla-Rocha,
Department of Gastroenterology, University Hospital of Pontevedra. SERGAS. Xunta de Galicia
PhD. JoséManuel Brea Floriani,
BioPharma, REGID- Innopharma, CIMUS, University of Santiago de Compostela, Spain.
PhD. Juan Ramon Rabuñal Dopico.
RNASA-IMEDIR Group, University of Coruña (UDC), Coruña University Hospital (CHUAC), Spain.

PhD. Daniel Torrecilla,
BioPharma, REGID- Innopharma, CIMUS, University of Santiago de Compostela, Spain.

Ph.D. Diana Herrera Ibatá,
Department of Computer Sciences, FIC, University of Coruña (UDC), Spain.

Adjunct Prof. Irantzu Barrio, Ph.D.
Department of Applied Mathematics and Statistics, University of Basque Country (UPV/EHU), Leioa, Bizkaia.
PhD. Francisco M. Ortuño Guzmán.
Postdoctoral Researcher, University of Granada (UGR)

Ph.D. Yong Liu,
Fellow Computer Science Faculty, University of A Coruna (UDC), Spain.
M.Sc. Enrrique Barreiro,
Department of Computer Sciences, FIC, University of Coruña (UDC), Spain.

MSc. Riccardo Zanni,
Research Associate, Department of Physicalchemistry, Faculty of Pharmacy, Universitat de Valencia, Spain.

Ph.D. Virginia Mato Abad,
LAIMBIO, University Rey Juan Carlos, Madrid, Spain.

PhD. Enrique Fernández Blanco,
RNASA-IMEDIR Group, University of Coruña (UDC), Coruña University Hospital (CHUAC), Spain.
PhD. MD. Javier Saavedra.
Family doctor. EOXI - SERGAS. Galician Government, Institute of Biomedical Research (INIBIC), Coruña, Spain.
PhD. Javier Pereira Loureiro.
RNASA-IMEDIR Group. University of Coruña (UDC), Institute of Biomedical Research of Coruña (INIBIC), Coruña University Hospital (CHUAC), Spain.

SPONSORS:

SciForum
Mol2Net

Pro-ChInt: Machine Learning Methods for Identifying Dual-/Multi- Protein Chains Interactions with Python

Yong Liu [*a]

[a] Information and Communication Technologies Department, Faculty of Computer Science, University of A Coruna, 15071 A Coruña, Spain

[*] Correspondent author: Yong Liu; e-mail: y.liu86@outlook.com; Tel.: +34 981 167 000; Fax: +34 981 167 160.

Published: 7 December 2015

Abstract: In nature, protein chain interactions (Pro-ChInt) of single- / multi-protein, a common but complex system, refer to physical contacts established between two or more protein chains depending on the amino acid sequences, which contains tremendous information. Decoding amino acid sequence information of protein using complex networks or graphs of the peptides is a grateful solution to discover the communication information between different Pro-ChInt. We first constructed some python codes to directly download the specify protein sequences from the RCSB protein data bank (PDB). Then, we changed the FASTA format to S2SNet format to calculate the embedded / non-embedded parameters of protein chains according to the star graph topological indices of peptide sequences. Meanwhile, we numbered all protein chains, then used the chain numbers to get a random number for a given set of chain number or case number used for each protein. Then, we replaced all the random numbers with the corresponding parameters of each protein chain calculated with S2SNet application. After that, a machine learning classification model was constructed based on the combinatorial / combining interaction of different chains. This new method can be used to identify two or more protein chain interactions combined with machine learning technique.

Keywords: Machine Learning, Protein Interaction, Protein Sequence, Python Scripts, Classification Model.

1. Introduction

Proteins are the main components of the biological metabolic pathways in living organisms. In nature, it could be one individual chain, or more than two chains to constitute a functional complex organic whole. Generally, the communications among different protein chains are very complicated, how to decode the communicational "language" is an important research topic in current chemoinformatics, bioinformatics, and pharmaceutics.

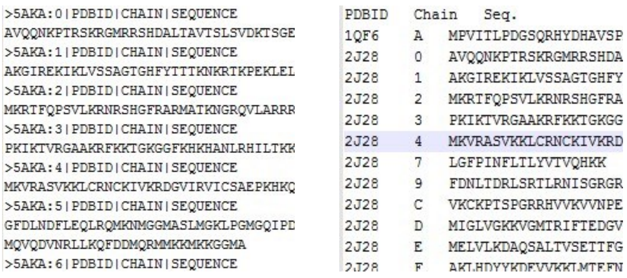
The biological systems are very complicated, therefore, a lot of scientists try to account for the biological complex problems with the techniques of genomics, transcriptomics and proteomics. However, proteomics are more complicated than genomics as genome is generally constant, whereas the proteome differs lie on cell and time. Proteins are subjected to a wide variety of chemical modifications after translation. It called as post-translational modifications, such as phosphorylation, ubiquitination, methylation, oxidation, *etc*.

Well understood of protein molecular information is helpful to disease control or prevention. This is because structure decides function for proteins. Whereas, proteins are the "practitioner" directly participating in the complex biological life cycle. In nature, protein – protein interactions refer to the physical contacts established between two or more proteins by the electrostatic forces and/or biochemical events. Whereas, the functional domains are generally formed by two or more protein chains but not only one chain or one protein. Decoding amino acid sequence information of protein, using complex networks or graphs of the peptide, is a grateful solution to uncover protein chain – chain interaction (Pro-ChInt).

Some sequence to structure graphs are used to calculate the numeric descriptors of molecular structure, for instance, MARCH-INSIDE [1] and S2SNet [2, 3]. These tools can transform the characters and numeric sequences into Star network graph. And then to calculate Star Graph Topological Indices.

2. Results and Discussion

In present work, we first searched the target PDB-ID with some special performance, and save all PDB-ID in a text file. Then we got all the FASTA profile of protein chain by using the python module "urllib2". We transformed the FASTA to S2SNet format, some examples of FASTA and S2SNet profile was presented in **figure 1**. S2SNet format is easier for further work.



```
>5AKA:0|PDBID|CHAIN|SEQUENCE
AVQQNKPTRSKRGMRRSHDALTAVTSLSVDKTSGE
>5AKA:1|PDBID|CHAIN|SEQUENCE
AKGIREKIKLVSSAGTGHFYTTTKNKRTKPEKLEL
>5AKA:2|PDBID|CHAIN|SEQUENCE
MKRTFQPSVLKRNRSHGFRARMATKNGRQVLARRR
>5AKA:3|PDBID|CHAIN|SEQUENCE
PKIKTVRGAAKRFKKTGKGGFKHKHANLRHILTKK
>5AKA:4|PDBID|CHAIN|SEQUENCE
MKVRASVKKLCRNCKIVKRDGVIRVICSAEPKHKQ
>5AKA:5|PDBID|CHAIN|SEQUENCE
GFDLNDFLEQLRQMKNMGGMASLMGKLPGMGQIPD
MQVQDVNRLLKQFDDMQRMMKKMKKGGMA
>5AKA:6|PDBID|CHAIN|SEQUENCE
```

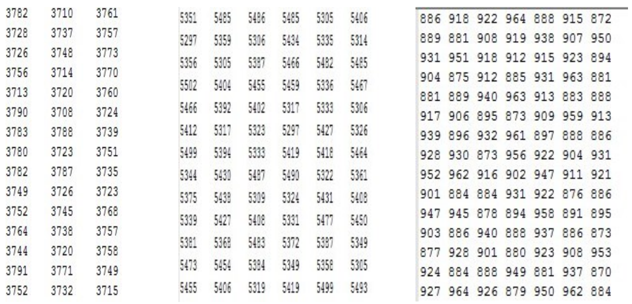
```
PDBID   Chain   Seq.
1QF6    A    MPVITLPDGSQRHYDHAVSP
2J28    0    AVQQNKPTRSKRGMRRSHDA
2J28    1    AKGIREKIKLVSSAGTGHFY
2J28    2    MKRTFQPSVLKRNRSHGFRA
2J28    3    PKIKTVRGAAKRFKKTGKGG
2J28    4    MKVRASVKKLCRNCKIVKRD
2J28    7    LGFPINFLTLYVTVQHKK
2J28    9    FDNLTDRLSRTLRNISGRGR
2J28    C    VKCKPTSPGRRHVVKVVNPE
2J28    D    MIGLVGKKVGMTRIFTEDGV
2J28    E    MELVLKDAQSALTVSETTFG
2J28    F    AKLHDYYKDEVVKKLMTEFN
```

FASTA format S2SNet format

Figure 1. The FASTA and S2SNet profiles of some protein chain

After to get the S2SNet format, the TIs parameters of each protein chain was calculated by S2SNet application. In here, we also can use others methods to calculate the molecular descriptors for protein sequences. For instance, we are trying to divide all the amino acids into four different types, polar or non-polar, charged or uncharged amino acids. We can count the number of polar-polar amino acids, polar-x-polar amino acids, and polar-x…x-polar amino acids. Or other types of connection between the different amino acids. However, this part of work, we have not yet finished. So in here, we used S2SNet, one of previous work in our group.

On the other hand, we numbered all the chain in a given file, and to select the corresponding numbers of each protein to run a random selection among the given chain numbers (n). For example, the first protein has 9 chains, these chains have the number from 1 to 9. We let users to put the number (m-fold), we can get the random cases = n × m. However, we have to remove the duplicates before we get all the final cases. The more important part of present work is to define how many of chain will be assigned to run the interaction between one to others. For example, with our new codes, we can perform the interaction among two or more (depending on the users, **Figure 2**).



```
3782   3710   3761
3728   3737   3757
3726   3748   3773
3756   3714   3770
3713   3720   3760
3790   3708   3724
3783   3788   3739
3780   3723   3751
3782   3787   3735
3749   3726   3723
3752   3745   3768
3764   3738   3757
3744   3720   3758
3791   3771   3749
3752   3732   3715
```

```
5351  5485  5486  5485  5305  5406
5297  5359  5306  5434  5335  5314
5356  5305  5387  5466  5482  5485
5502  5404  5455  5459  5336  5467
5466  5392  5402  5317  5333  5306
5412  5317  5323  5297  5427  5326
5499  5394  5333  5419  5418  5464
5344  5430  5487  5490  5322  5361
5375  5438  5309  5324  5431  5408
5339  5427  5408  5331  5477  5450
5381  5368  5483  5372  5387  5349
5473  5454  5384  5349  5358  5305
5455  5406  5319  5419  5499  5493
```

```
886 918 922 964 888 915 872
889 881 908 919 938 907 950
931 951 918 912 915 923 894
904 875 912 885 931 963 881
881 889 940 963 913 883 888
917 906 895 873 909 959 913
939 896 932 961 897 888 886
928 930 873 956 922 904 931
952 962 916 902 947 911 921
901 884 884 931 922 876 886
947 945 878 894 958 891 895
903 886 940 888 937 886 873
877 928 901 880 923 908 953
924 884 888 949 881 937 870
927 964 926 879 950 962 884
```

3 chains 6 chains 7 chains

Figure 2. The examples of random numbers for the chain-chain interaction

In addition, each random number refers to the corresponding chain sequence, and each sequence would be calculated into 42 TIs. If all the sequences (numbers) in the combination are from the same protein, we defined this case as the "positive" or "1", whereas, if not all numbers from the same protein, we consider this case as the "negative" or "0".

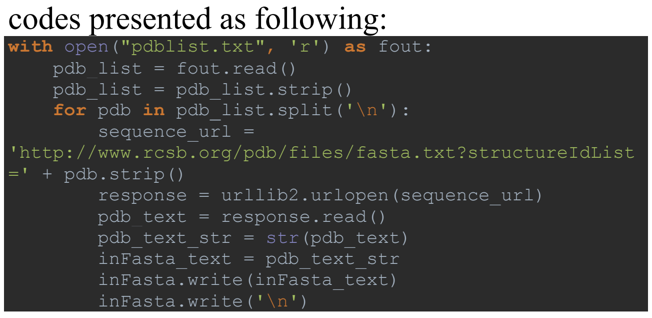
After that, we obtained and calculated each combination character based on the average values of each combination (42 TIs average values). Using this data to run a classification model to identify if there are interactions among the two or more chains.

3. Material and Methods

All codes were programmed in the platform of PyCharm 3.5 version under the environmental of python 2.7 version. There are different steps to establish Pro-ChInt. They include to obtain the target protein chains sequence, change FASTA to S2SNet format and calculate S2SNet star graph topological indices, to get the serial random number, etc.

3.1. Download FASTA files

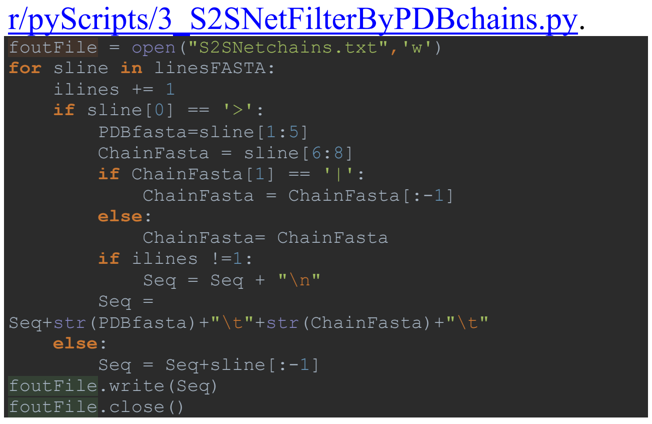
First step, we programed the codes in python to download the FASTA file from the protein data bank (PDB) according to the PDB-ID (serial number). In this part, we used urllib2 module to corresponding website of specify protein ID. The codes presented as following:



```
with open("pdblist.txt", 'r') as fout:
    pdb_list = fout.read()
    pdb_list = pdb_list.strip()
    for pdb in pdb_list.split('\n'):
        sequence_url =
'http://www.rcsb.org/pdb/files/fasta.txt?structureIdList
=' + pdb.strip()
        response = urllib2.urlopen(sequence_url)
        pdb_text = response.read()
        pdb_text_str = str(pdb_text)
        inFasta_text = pdb_text_str
        inFasta.write(inFasta_text)
        inFasta.write('\n')
```

3.2. Transform FASTA to S2SNet format

In this part, we change the FASTA format to S2SNet format for the further calculation. The details python codes presented in GitHub: https://github.com/muntisa/pyS2SNet/blob/master/pyScripts/3_S2SNetFilterByPDBchains.py.



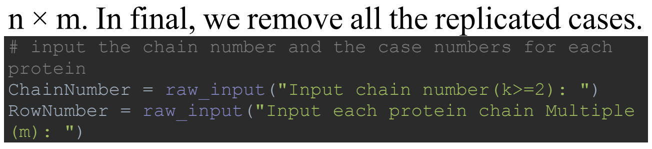
```
foutFile = open("S2SNetchains.txt",'w')
for sline in linesFASTA:
    ilines += 1
    if sline[0] == '>':
        PDBfasta=sline[1:5]
        ChainFasta = sline[6:8]
        if ChainFasta[1] == '|':
            ChainFasta = ChainFasta[:-1]
        else:
            ChainFasta= ChainFasta
        if ilines !=1:
            Seq = Seq + "\n"
        Seq =
Seq+str(PDBfasta)+"\t"+str(ChainFasta)+"\t"
    else:
        Seq = Seq+sline[:-1]
foutFile.write(Seq)
foutFile.close()
```

3.3. Calculate the S2SNet topological indices

We obtained 42 topological indices (TIs) for each protein chain, calculated by S2SNet star graph. There are two types of TIs (Embedded and non-Embedded indices). Each one has 21 TIs. Like Shannon entropies, connectivitymatrices, Harary number, Wiener index, Gutman topological index with different power [4]. S2SNet are widely used in obtaining the molecular information of protein [5].

3.4. Get the random number matrix

In this part, we first numbered all the protein chains according to the order of all chain appeared in the PDB chain file. Each protein chain has the only special number. For example, the first protein has 9 chains (n), these chains have the number from 1 to 9, but for the second protein, if it has 15 chains, the number of these chains are from 10 to 24, and so on. Then, we used two codes to let the users to input the chain number, how many chains (Maximum) will be accounted for Pro-ChInt. Meantime, we let users to put the number (m-fold), we can get the random cases = n × m. In final, we remove all the replicated cases.



```
# input the chain number and the case numbers for each
protein
ChainNumber = raw_input("Input chain number(k>=2): ")
RowNumber = raw_input("Input each protein chain Multiple
(m): ")
```

3.5. Classification modeling

In this step, we replace the random number with the 42 TIs of corresponding protein chain sequence. For one combination, if all the chains are from the same original protein, we consider this combination has the chain-chain interaction (Pro-ChInt) set as “1 or positive”. If the combination is from different protein, we consider this combination has no chain-chain interaction, Pro-ChInt, set as “0 or negative”.

For each combination, we calculate the average value of each parameter in 42 TIs of S2SNet Star Graph. After that, we can use Weka to obtain the best classification model depending on the combination mentioned previous.

4. Conclusions

This short communication is presenting some original python codes for identify the protein chain – chain interactions lie on the S2SNet Star Graph Topological Indices. The ideas of this work are on account of molecular descriptors obtained from Star Graphs. Then to use Machine Learning methods running in Weka to search for the best classification model. We can explain the protein chain – chain interaction based on the molecular information of protein sequences.

Acknowledgments

The authors acknowledge the support provided by the Galician Network of Drugs R+D REGID (Xunta de Galicia R2014/025). This work was partially supported by the Galician Network for Colorectal Cancer Research (Red Gallega de Cáncer Colorrectal - REGICC, Ref.: CN 2012/217), Institute for Biomedical Informatics of A Coruña (INIBIC), and Center for Research of Information and Communication Technologies (CITIC).

Conflicts of Interest

The authors declare no conflict of interest.

References

1. González-Díaz, H.; Torres-Gomez, L. A.; Guevara, Y.; Almeida, M. S.; Molina, R.; Castanedo, N.; Santana, L.; Uriarte, E. Markovian chemicals "in silico" design (MARCH-INSIDE), a promising approach for computer-aided molecular design III: 2.5D indices for the discovery of antibacterials. *J Mol Model* **2005,** *11* (2), 116-23.
2. Munteanu, C. R.; Gonzáles-Díaz, H. *S2SNet - Sequence to Star Network, Reg. No. 03 / 2008 / 1338, Santiago de Compostela, Spain*, Santiago de Compostela, Spain, 2008.
3. Munteanu, C. R.; Magalhaes, A. L.; Duardo-Sanchez, A.; Pazos, A.; Gonzalez-Diaz, H. S2SNet: A Tool for Transforming Characters and Numeric Sequences into Star Network Topological Indices in Chemoinformatics, Bioinformatics, Biomedical, and Social-Legal Sciences. *Current Bioinformatics* **2013,** *8* (4), 429-437.
4. Fernández-Blanco, E.; Aguiar-Pulido, V.; Munteanu, C. R.; Dorado, J. Random Forest classification based on star graph topological indices for antioxidant proteins. *Journal of Theoretical Biology* **2013,** *317*, 331-337.
5. Liu, Y.; Munteanu, C. R.; Fernández Blanco, E.; Tan, Z.; Santos del Riego, A.; Pazos, A. Prediction of Nucleotide Binding Peptides Using Star Graph Topological Indices. *Molecular Informatics* **2015,** *34* (11-12), 736-741.

SciForum
Mol2Net

Synthesis and Platinum (II) Complexes of Different Polyazacyclophane Receptors

Begoña Verdejo[1,*], Estefanía-Delgado-Pinar[1], Javier Pitarch-Jarque[1], Lorena Magraner-Pardo[2], J. Vicente de Julián-Ortiz[2], Enrique García-España[1]

[1] Institut de Ciència Molecular, Universitat de València, Paterna, Valencia, Spain
[2] Innotecno Development SL, Parc Científic, Paterna, Valencia, Spain.
* Author to whom correspondence should be addressed; E-Mail: begona.verdejo@uv.es
Tel.: +34-963544401

Published: 7 December 2015

Abstract:

The interaction of $PtCl_4^{2-}$ with different polyazacyclophanes containing a pyridine unit as aromatic spacer has been studied by ^{1}H and ^{195}Pt NMR spectroscopy, Analysis of the recorded spectra of D_2O solutions containing L and $PtCl_4^{2-}$ in a 1:1 molar ratio at acidic pH shows the evolution with time of the ^{1}H and ^{195}Pt signals. Different crystal structures have been solved by X-ray diffraction analysis. At acidic pHs, the metal ion is coordinated by the central amino group of the macrocyclic cavity and three chloride or bromide atoms, in a square planar geometry. Formation of [Pt(H$_2$**L1**)Br$_3$]Br (**1**) and [Pt(H$_2$**L2**)Br$_3$]Br (**2**) reveals the rapid substitution of chloride ligands in $PtCl_4^{2-}$ by bromide ligands. However, as reveals the crystal structure obtained for [Pt^{IV}**L3**Br$_2$](PtBr$_4$)(H$_2$O) (**4**), at slightly higher pH values, the metal ion is coordinated through all nitrogen atoms of the macrocyclic cavity and an oxidation to Pt(IV) occurs.
.

Keywords: platinum complexes, polyazacyclophanes, coordination chemistry

1. Introduction

During the last years, research on coordination chemistry of platinum has aroused great interest due to their potential biological applications in drug design. [1-7]

Here, we communicate some initial results on coordination chemistry of Pt(II) with different azapyridinacyclophanes (see Chart 1), These triaza macrocycles, have been shown to display interesting properties in their coordination to metal ions.[8, 9]

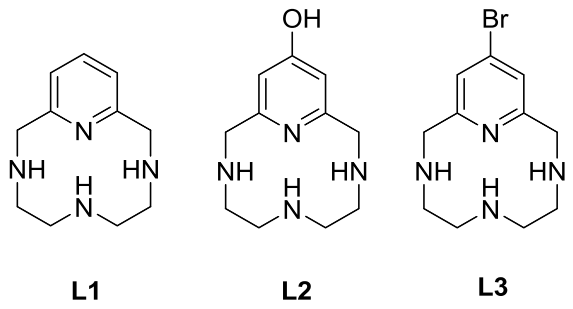


Chart 1

2. Results and Discussion

Crystals suitable for x-ray diffraction were obtained from K_2PtCl_4 and L·3HBr in I:1 molar ratio. In accordance with a behaviour previously reported by García-España *et al.*,[10] the crystal structures obtained for [Pt(H_2**L1**)Br_3]Br (**1**) [Pt(H_2**L2**)Br_3]Br (**2**) and [Pt^{IV}**L3**Br_2] ($PtBr_4$)(H_2O) (**4**) reveal the rapid substitution of chloride ligands in $PtCl_4^{2-}$ by bromide ligands.

For [Pt(H_2**L1**)Br_3]Br (**1**), each unit includes one [Pt(H_2**L1**)Br_3]$^+$ cation and a bromide counterion. Due to the presence of two protonated amino groups (N2 and N4), the metal ion cannot be accommodated in the macrocyclic cavity. Thus, Pt(II) presents the characteristic square planar geometry, coordinated by three bromide lignds and the central amino group of the macrocycle. (see Figure 1). The different [Pt(H_2**L1**)Br_3]$^+$ cations are interconnected, through an intermolecular hydrogen-bond array in which the bromide counterion links through hydrogen bonding the protonated amino groups of two contiguous units (Br4-H···N4=2.280Å and Br4-H···N2=2.238Å). At the same time, another hydrogen bond can be observed between the protonated amino group N2 and one of the bromide ligands coordinated to the metal ion (Br2-H···N2=2.737Å). It is noteworthy that [Pt (H_2**L2**)Br_3]Br (**2**) and [Pt(H_2**L1**)Cl_3]Cl (**3**) present an analogous structure to the described for [Pt(H_2**L1**)Br_3]Br (**1**).

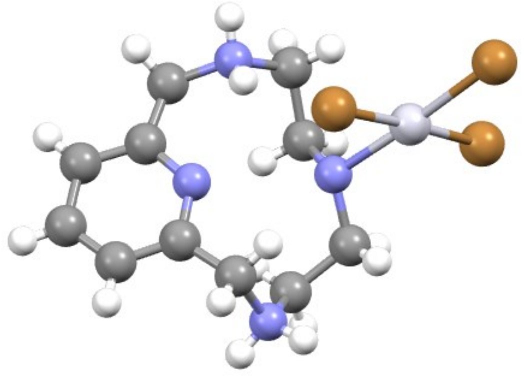


Figure 1. X-Ray crystal structure of the cation [Pt(H_2L1)Br_3]$^+$

However, as reveals the crystal structure obtained for [Pt^{IV}**L3**Br_2]($PtBr_4$)(H_2O) (**4**), at slightly higher pH values, the metal ion is coordinated through all nitrogen atoms of the macrocyclic cavity and two bromide ligands complete the octahedral geometry, indicating that an oxidation to Pt(IV) occurs. As can be seen in Table 1, the Pt-N bond distances in the [Pt^{IV}**L3**Br_2]$^{2+}$ cation are shorter than the obtained for the bromide ligands. Furthermore, the [Pt^{IV}**L3**Br_2]$^{2+}$ cations are interconnected through intermolecular hydrogen-bonds with the [$PtBr_4$]$^{2-}$ anions. This counterion links through hydrogen bonding one of the benzylic amino groups of a unit (Br6-H···N2=2.323Å) with the same amino group of the following unit creating chains that are isolated and do not show any kind of interconnection

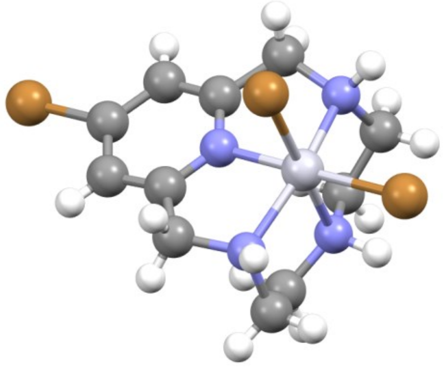


Figure 2. X-Ray crystal structure of the cation [Pt^{IV}**L3**Br_2]$^{2+}$

Figure 3 shows the evolution with time of ^{1}H and ^{195}Pt NMR spectra of D_2O solutions containing K_2PtCl_4 and **L1**·3HCl in 1:1 molar ratio recorded at acidic pH. Initially, the ^{195}Pt spectrum consists of a signal at - 1660 ppm which can be attributed to [$PtCI_4$]$^{2-}$.[11] After 2 days, a new signal at -2030 ppm appears and can be attributed to a platinum(II) ion coordinated to three chlorides and to a nitrogen atom of the macrocycle,, in accordance with the crystal structure obtained for this receptor, [Pt(H_2**L1**)Cl_3]Cl (**3**).

These ^{195}Pt NMR spectral changes are accompanied by significative variations in the ^{1}H NMR spectra. The initial ^{1}H NMR spectrum, which corresponds to the fully protonated free receptor presents, in the aliphatic region, a singlet signal at 4.55 which can be assigned to the benzylic protons, and two triplet signals at 3.20 and 2.95 ppm assigned to the protons of the ethylenic chains. In the aromatic region, a triplet and a doblet signals appear at 8.00 ppm and 7.48 ppm respectively. As the reaction proceeds, although the symmetry is essentially preserved, new signals with more complex spin systems appear.

Table 1. Selected distances and angles

$[Pt(H_2\mathbf{L1})Br_3]Br$ (**1**)				$[Pt^{IV}\mathbf{L3}Br_2](PtBr_4)(H_2O)$ (**4**)			
Distances (Å)		**Angles(°)**		**Distances (Å)**		**Angles(°)**	
Pt1-N1	2.077(5)	Br1-Pt1-Br2	92.22 (2)	Pt1-N1	1.974(5)	N1-Pt1-N2	82.21 (2)
Pt1-Br1	2.374 (3)	Br2-Pt1-Br3	91.30(2)	Pt1-N3	2.055(5)	N2-Pt1-N3	84.35 (2)
Pt1-Br2	2.406(5)	Br1-Pt1-N3	85.40(3)	Pt1-N2	2.080(3)	Br2-Pt1-Br3	91.64(2)
Pt1-Br3	2.410(6)	Br3-Pt1-N3	91.08(2)	Pt1-Br2	2.446(5)	Br2-Pt1-N2	85.57(3)
				Pt1-Br3	2.414(6)	Br2-Pt1-N1	87.86(2)
						Br3-Pt1-N2	97.83(2)
						Br3-Pt1-N3	88.76(2)

3. Materials and Methods

The synthesis of **L1-L3** has been carried out by slightly modifications on the general procedures described in literature.[8,9,12] All reagents and chemicals were obtained from commercial sources and used as received. Solvents used for the chemical synthesis were of analytical grade and used without further purification.

Synthesis of $[Pt(H_2L1)Br_3]Br$ (1). To an aqueous solution (5 mL) of **L1**·3HBr, $K_2[PtCl_4)]$ in water (5 mL) in a 1.1 molar ratio was added dropwise with stirring. After the mixture was stirred for 2 h at room temperature, it was filtered. Orange crystals suitable for X-Ray analysis were obtained by slow evaporation of the solvent.

Synthesis of $[Pt(H_2L2)Br_3]Br$ (2). To an aqueous solution (5 mL) of **L2**·3HBr, $K_2[PtCl_4)]$ in water (5 mL) in a 1.1 molar ratio was added dropwise with stirring. After the mixture was stirred for 2 h at room temperature, it was filtered. Orange crystals suitable for X-Ray analysis were obtained by slow evaporation of the solvent.

Synthesis of $[Pt(H_2L1)Cl_3]Cl$ (3). To an aqueous solution (5 mL) of **L1**·3HCl, $K_2[PtCl_4)]$ in water (5 mL) in a 1.1 molar ratio was added dropwise with stirring. After the mixture was stirred for 2 h at room temperature, it was filtered. Yellow crystals suitable for X-Ray analysis were obtained by slow evaporation of the solvent.

Synthesis of $[Pt^{IV}L3Br_2](PtBr_4)(H_2O)$ (4). To an aqueous solution (5 mL) of **L3**·3HBr, $K_2[PtCl_4)]$ in water (5 mL) in a 1.1 molar ratio was added dropwise with stirring.. After the mixture was stirred for 2 h at room temperature, it was filtered. Orange crystals suitable for X-Ray analysis were obtained by slow evaporation of the solvent.

NMR Measurements. The ^1H and ^{13}C NMR spectra were recorded on a Bruker Avance AC-300 spectrometer operating at 299.95 MHz for ^1H. The chemical shifts are given in parts per million referenced to the solvent signal. Adjustments to the desired pH were made using drops of DCl or NaOD solutions. The pD was calculated from the measured pH values using the correlation, $pH = pD - 0.4$.[13]

Crystallographic analysis Analysis of single crystals was carried out with an Enraf-Nonius KAPPA CCD single-crystal diffractometer (λ =0.71073 Å). The structures were solved with using the program SHELXS-86.[14] Structure refinement was performed by means of the program SHELXL-97.[15] Molecular plots were produced with either the program MERCURY[16] or ORTEP.[17] Crystal data, data collection parameters, and results of analysis are listed in Table 2.

Table 2. Crystal data for **1**, **2**, **3** and **4**.

Compound	**1**	**2**	**3**	**4**
Formula	$C_{11}H_{20}Br_4N_4Pt$	$C_{11}H_{20}Br_4N_4OPt$	$C_{11}H_{20}Cl_4N_4Pt$	$C_{11}H_{21}Br_7N_4O_2Pt_2$
M.W.	723.04	739.04	545.2	1190.87
T (K)	293(2)	293(2)	293(2)	293(2)
Crystal system	monoclinic	monoclinic	monoclinic	orthorhombic
Space group	$P2_1/c$	$P2_1/c$	$P2_1/c$	Pnma
a (Å)	8.0766(3)	8.1303(4)	7.8747(2)	25.8639(8)
b (Å)	19.7244(8)	20.1032(8)	19.5912(4)	11.3142(2)
c (Å)	11.4678(7)	11.0855(4)	11.3248(3)	7.5064(2)
α (°)	90	90	90	90
β (°)	106.707(5)	105.033(4)	105.741(3)	90
γ (°)	90	90	90	90
Volume (Å^3)	1749.77(15)	1749.86(13)	1681.61(8)	2196.59(11)
Z	4	4	4	4
ρ (g/cm^3)	2.745	2.805	2.153	3.601
λ (Å)	0.71073	0.71073	0.71073	0.71073
F(000)	1328	1360	1040	2128
μ(mm^{-1})	17.149	17.156	8.975	25.473
2Θ range	6.694 to 49.998	6.584 to 53.996	6.798 to 49.986	6.514 to 49.986
Ref. collect.	6380	7827	6380	10431
Indep. ref.	3050	3814	2955	2030
R(int)	0.0903	0.0353	0.0336	0.0365
Data/restr/param	3050/67/181	3814/37/192	2955/0/181	2030/2/136
R1 (I>4σ)	0.0984	0.0662	0.02279	0.0451
***wR*^{2}**	0.3075	0.1949	0.0521	0.1136
GOOF (F^2)	1.045	1.059	1.035	1.038

4. Conclusions

Interaction of $K_2[PtCl_4]$ with different pyridine azacyclophanes in aqueous solution leads to a fast replacement of the chloride ligands in $[PtCl_4]^{2-}$. by bromide ligands. The X-ray analysis of the compound shows that interaction of $[PtCl_4]^{2-}$ with L.3HBr in 1:1 molar ratio gives rise to two different complexes as function of the pH, which differ in the location of platinum in the macrocycle. At acidic pH, Pt(ll) binds to the central nitrogen of the macrocycle, while at slightly higher pH values, as the benzylic nitrogens deprotonate, the metal ion can be coordinated by all the nitrogen atoms of the macrocyclic cavity and an oxidation to Pt(IV) occurs.

We are currently trying to characterize better these complexes in solution, analyzing the effect of pH changes, as well as trying to obtain information on interaction with nucleobases and oligonucleotides.

Acknowledgments

Financial support by the Spanish Ministerio de Economía y Competitividad (Projects CONSOLIDER INGENIO CSD-2010-00065 and CTQ2013-48917-C3-1-P), Generalitat Valenciana (Project PROMETEOII2015-002) is gratefully acknowledged.

Conflicts of Interest

The authors declare no conflict of interest

References and Notes

1. Bruhn, S, L; Toney, J H.; Lippard, S J, in S.J Lippard (ed,), Progress in Inorganic Chemistry: Bioinorganic Chemistry, Vol 38, Wiley, 1990 p, 477.
2. Lippard, S,J,; Berg, J.M. Principles of Bioinorganic Chemistry, University Science Books, Mill Valley, CA, 1994;
3. Lippard, S,J ,in Benini, I,; Gray, H,B.; Lippard S J,; Valentine J.S, (Eds), Bioinorganic Chemistry, University Science Books, Mill Valley, CA. 1994, Ch. 9, p 505;
4. Barton, J,K. in Benini, I,; Gray, H,B.; Lippard S J,; Valentine J.S, (Eds) Bioinorganic Chemistry, University Science Books, Mill Valley, CA, 1994, Ch. 8, p. 455;
5. Bloemink, M J.; . R¢cd|jk. I, in Sigel H. and Sigel A. (Eds.). Metal Ions m Biological Systems. Marcel Dckkcr, New York. 1996, p 32.
6. Chu, G. Cellular responses to cisplatin. The roles of DNA-binding proteins and DNA repair. *Journal of Biological Chemistry,* **1994**, *269*, 787-790
7. Giandomenico, G. M.; Abrams M. J.; Murrer. B.A.; Vollano. J. F.; Rheinhelmer. M, I.; Wyer, S. B. Bossard G. E.; Higgins J. D. Carboxylation of Kinetically Inert Platinum(IV) Hydroxy Complexes. An Entrée into Orally Active Platinum(IV) Antitumor Agents. *lnorganic Chemistry* **1995**, *34*, 1015-1021; Reedijk, J.. Improved understanding in platinium antitumour chemistry. *Chemical Communi*cations. **1996**, 801-806, and Refs. Therein
8. Costa, J.; Delgado, R. Metal Complexes of Macrocyclic Ligands Containing Pyridine. *Inorganic Chemistry* **1993**, *32*, 5257-5265
9. Lincoln, K.M.; Offutt, M. E; Hayden, T. D.; Saunders, R. E.; Green, K. N.; Structural, Spectral, and Electrochemical Properties of Nickel(II), Copper(II), and Zinc(II) Complexes Containing 12-Membered Pyridine- and Pyridol-Based Tetra-aza Macrocycles. *Inorganic Chemistry* **2014**, *53*, 1406−1416
10. García-España, E; Latorre, J.; Marcelino, V.; Ramírez, J. A.; Luis, S. V.; Miravet, J. F.; Querol, M. Outer and inner coordination sphere chemistry of polyazacyclophane platinum (II) complexes. Crystal structure of $[PtBr_4]_2(H_4L1)\cdot H_2O$ (L1 = 2,6,9,13-tetraaza [14] paracyclophane) *Inorganica Chimica Acta* **1997**, *265*, 179-186
11. Alei, M.; Vergamini, P.J.; Wageman, E.E.; ^{15}N NMR of cis-diamine-platinum(II) complexes in aqueous solution. *Journal of the American Chemical Society*, **1979**, *101*, 5415-5417. Chikuma M.; Pollock, R.J. The ^{195}Pt chemical shifts and ^{195}Pt-^{15}N coupling constants for *cis*-diammineplatinum(II) complexes. *Journal of Magnetic Resonance* **1982**, *47*, 324-327. Hollis L.S.; Lippard, S.J. Synthesis, structure, and ^{195}Pt NMR studies of binuclear complexes of cis-

diammineplatinum(II) with bridging .alpha.-pyridonate ligands. *Journal of the American Chemical Society*, **1983**, *105*, 3494-3503; Appleton, T. G.; Berry, R. D.; Davis, C. A.; Hall J. R; Kimlin, H.A. Reactions of platinum(II) aqua complexes. 1. Multinuclear (^{195}Pt, ^{15}N, and ^{31}P) NMR study of reactions between the cis-diamminediaquaplatinum(II) cation and the oxygen-donor ligands hydroxide, perchlorate, nitrate, sulfate, phosphate, and acetate. *lnorganic Chemistry*, **1984**, *23*, 3514-3521; Appleton, T.G.; Hall, J.R.; Ralph S. F.; Thompson C. S. M. Reactions of platinum(II) aqua complexes. 2. ^{195}Pt NMR study of reactions between the tetraaquaplatinum(II) cation and chloride, hydroxide, perchlorate, nitrate, sulfate, phosphate, and acetate. *lnorganic Chemistry* **1984**, *23*, 3521-3525; Appleton, T.G.; Hall J.R.; Ralph S,F. ^{15}N and ^{195}Pt NMR spectra of platinum ammine complexes: trans- and cis-influence series based on ^{195}Pt - ^{15}N coupling constants and ^{15}N chemical shifts. *lnorganic Chemistry* **1985**, *24*, 4685-4693; Bales, J. R.; Mazid, M.A.; Sadler, P. J.; Aggarwal, A; Kuroda, R.; Neidle, S.; Gilmour, D. W.; Pearl B. J.; Ramsden C.A. Platinum(II) complexes of nitroimidazoles: synthesis, characterisation, and *X*-ray crystal structures of *cis*-dichlorobis[1-(2′-hydroxyethyl)-2-hydroxymethyl-5-nitroimidazole]platinum(II) and *trans*-dichlorobis[1-(2′-hydroxy-3′-methoxypropyl)-2-nitroimidazole]platinum(II). *Journal of Chemical Society, Dalton Transactions*, **1985**, 795-802. Sundquist, W. I.; Ahmed, K, J.; Hollis L.S.; Lippard S. J. Solvolysis reactions of cis- and trans-diamminedichloroplatinum(II) in dimethyl sulfoxide. Structural characterization and DNA binding of trans-bis(ammine)chloro(DMSO)platinum(1+). *lnorganic Chemistry*, **1987**, *26*, 1524-1528; Habtemariam A.; Sadler, P.J. Design of chelate ring-opening platinum anticancer complexes: reversible binding to guanine. *Chemical Communications*, **1996**, 1785-1786,

12. Takalo, H.; Kankare, J. Preparation of new macrocyclic polyamines containing 4-(phenylethynyl)pyridine subunit. *Journal of Heterocyclic Chemistry* **1990**, *27*, 167–169.
13. Glasoe, P. K.;. Long, F. A. Use of glass electrodes to measure acidities in deuterium oxide. *Journal of Physical Chemistry* **1960**, 64, 188−190. Covington, A. K.; Paabo, M.; Robinson, R. A.; Bates, R. G. Use of the glass electrode in deuterium oxide and the relation between the standardized pD (paD) scale and the operational pH in heavy water. *Analytical Chem*istry,**1968**, 40, 700−706.
14. Sheldrick, G. M.; Kruger, C., Goddard, R. Eds. Crystallographic Computing; Clarendon Press: Oxford, England, 1985; p 1175.
15. Sheldrick, G.M. A short history of *SHELX*. *Acta Crystallographica Section A*, **2008**, *64*, 112-122.
16. Edgington, P. R.; McCabe, P.; Macrae, C. F.; Pidcock, E.; Shields, G. P.; Taylor; R.; Towler M.; Streek, J. v. d Mercury: visualization and analysis of crystal structures. *Journal of Applied Crystallography* **2006**, *39*, 453-457.
17. C. K. Johnson, ORTEP; Report ORNL-3794, Oak Ridge National Laboratory, Oak Ridge, TN, 1971.

SciForum
Mol2Net

Molecular Rearrangement of an Aza-Scorpiand Macrocycle Induced by pH. A Computational Study

J. Vicente de Julián-Ortiz [1,*], Begoña Verdejo [2], Víctor Polo [3], Emili Besalú [4] and Enrique García-España [2]

[1] Fundación Centro de Innovación y Demostración Tecnológica, Paterna, Valencia, Spain; Departamento de Química Física, Universidad de Valencia, Spain; E-Mail: jejuor@uv.es
[2] Institut de Ciència Molecular, Universitat de València, Paterna, Valencia, Spain; E-Mail: begona.verdejo@uv.es (B.V.); enrique.garcia-es@uv.es (E.G.E.)
[3] Departamento de Química Física, Universidad de Zaragoza, Spain, E-Mail: vipolo@unizar.es
[4] Institut de Química Computacional, Universitat de Girona, Girona, Spain; E-Mail: emili.besalu@udg.edu
* Author to whom correspondence should be addressed; E-Mail: jejuor@uv.es;
Tel.: +34-963-543-279; Fax: +34-963-544-892.

Published: 7 December 2015

Abstract: Rearrangements and their control are a hot topic in supramolecular chemistry due to the possibilities that these phenomena open in the design of synthetic receptors and molecular machines. Macrocycle aza-scorpiands constitute an interesting system that can reorganize their spatial structure depending on pH variations or the presence of metal cations. In our case, the conformations change varies between the so called 'open' and 'closed', the last being found at lower pH. In this study, the relative stabilities of these conformations were predicted computationally by the Density Functional Theory approximation and the reorganization from closed to open was simulated by using the Monte Carlo Multiple Minimum method.

Keywords: pH controlled; supramolecular chemistry; synthetic receptors; aza-scorpiands; Density Functional Theory; Monte Carlo Multiple Minimum

1. Introduction

The possibility of controlling the conformation of chemical structures is interesting because opens the possibility of creating molecular machines and synthetic receptors that react to the desired stimuli.1 Among the molecular structures showing such property stand out the aza-macrocycles that show a coordinating tail, known as scorpiands. These systems merit attention due to their potential biological and pharmacological applications, since their ability to recognize hydrophilic and hydrophobic amino acids has been demonstrated.2 Receptors for amino acid sensing3 or drug delivery4 are two fields for their potential applications.

Changes in the protonation state are able to induce conformational reorganizations in scorpiand-like ligands.5,6 It has been observed that the presence of metal centers produce similar effects. Also, the metal coordination is influenced by the protonation state because of electrostatic repulsion.7

In this study, pH variation that make the structure change from the so called 'closed' to 'open' conformations have been simulated.

2. Methods

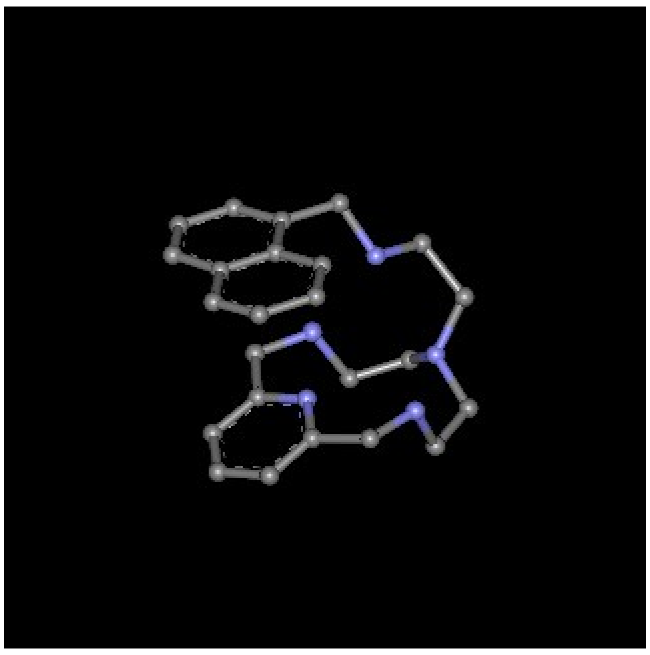
MM Conformational search. The simplified structures presented in Figure 1 were taken from X-ray geometries. Experimental values of the energy required to change the conformations from closed (**A**) to open (**B**) and vice versa, depending of the protonation state, are lacking. To have an understanding of the conformational stabilities involved in this process, the following simulation was undertaken. Conformation **A** is preferred for the mono- and diprotonated species. Preserving this conformation the molecule was triprotonated. Then, a conformational search was performed by Monte Carlo Multiple Minimum (MCMM) method8 with the MM+ force field,[9] by allowing the free change of the dihedral angles comprised in the carbon bridge that joins the two rings (tail). This method allows finding the lowest energy conformations of a molecule by randomly varying specified dihedral angles to generate new starting conformations and then energy minimizing each of those. Low-energy unique conformations are stored while high-energy or duplicate structures are discarded. Rotation is used for acyclic bond dihedral angles. The energy cut-off for discarding repeated structures was set to 6 Kcal/mol. Conversely, conformation **B** was di- or monoprotonated and treated with the same procedure, to see if it was able to reach the conformation **A**. These calculations were performed with the program Hyperchem version 7.5.[10]

The relative stabilities of these conformations were predicted by Density Functional Theory and the rearrangement was simulated by a Molecular Mechanics method.

Relative energies calculation. In order to provide computationally derived relative stability estimates to have an understanding of the conformational energies, density functional theory (DFT) calculations were undertaken. The three protonation states - mono, di and tri - were combined with the two possible conformational states **A** and **B** giving six chemical structures. For each protonation, the conformation energies of optimized **A** and **B** were compared to see which one would be more stable into simulated water environment. It was considered the atomisation energy, this is, the stabilization of the molecule relative to the free constitutive atoms. Several combinations of exchange functionals and basis sets were tested. The method finally chosen was all-electron local density approximation (LDA)[11] with the Vosko Wilk Nusair (VWN) functional and the Slater basis valence quadruple zeta with four polarization functions (ZORA/QZ4P), under the restricted spin formalism. Water environment was simulated by the method COSMO.[12] Some results also presented in this paper were obtained with the "generalized gradient approximation Becke88 Perdew86"[13] (GGA-BP) gradient corrected exchange funtional and the TZ2P basis, with no solvent simulation. All these calculations were performed by means of the program Amsterdam Density Functional.[14]

Table 1. Torsion angles in the tail for different conformations obtained with MCMM, starting with the conformation **A**

Conformation	Torsions allowed	a	b	c	d	e
A	-	-146.8	74.4	-170.9	80.7	77.0
1	12-ring, d	-148.4	59.1	176.8	144.9	66.3
2	12-ring, b, d	68.2	179.2	-91.2	71.3	74.7
3	a, b	-67.9	-175.5	94.2	-67.3	105.3
4	d	-62.6	-167.8	-177.3	-176.9	-814
5	12-ring, a	-154.9	170.8	177.6	172.9	-87.1
B	-	69.4	-176.5	-177.2	-179.2	84.6



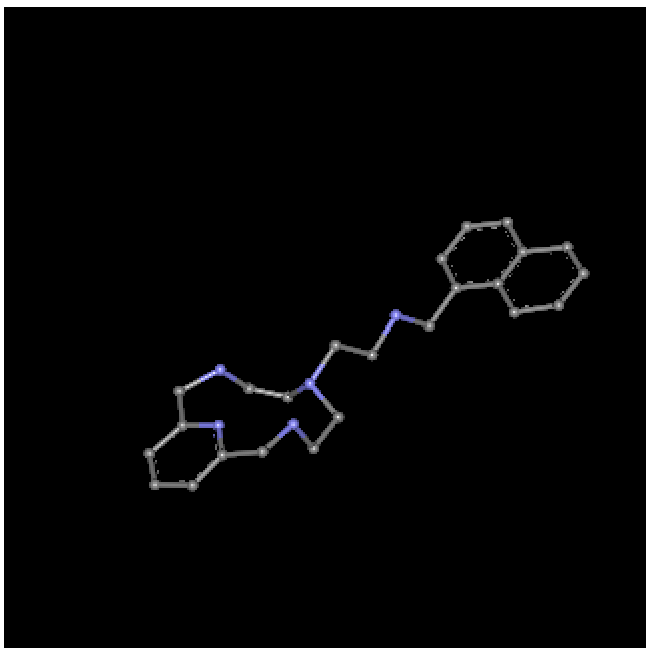


Figure 1. Conformations closed (**A**) and open (**B**) of the aza-scorpiand macrocycle studied

Table 2. Geometric parameters involved in the water box calculations for each conformation obtained from MCMM

Conformation	Smallest Box Enclosing Solute / Å			Cubic Periodic Box Edge / Å	Maximum number of water molecules
	X	Y	Z		
A	6.90	6.19	7.73	18.10	216
1	7.65	5.87	7.80	18.10	216
2	8.08	6.91	9.07	18.10	216
3	7.76	5.71	10.75	21.51	329
4	7.49	5.42	15.48	30.96	980
5	7.30	5.17	19.91	31.81	1064
B	6.90	3.29	17.62	35.25	1447

Table 3. Calculated energies for each protonation state and conformation closed (**A**) and open (**B**)

Starting conformation	Number of H+	Method	Environment simulation method	Total Bond Energy/ kcal/mol
A	1	LDA/QZ4P	COSMO	-9171.05
B	1	LDA/QZ4P	COSMO	-9162.84
A	2	LDA/QZ4P	COSMO	-9178.82
B	2	LDA/QZ4P	COSMO	-9152.72
A	3	LDA/QZ4P	COSMO	-9133.36
B	3	LDA/QZ4P	COSMO	-9134.40
A	3	GGA-BP/TZ2P	vacuum	-8396.54
B	3	GGA-BP/TZ2P	vacuum	-8408.01

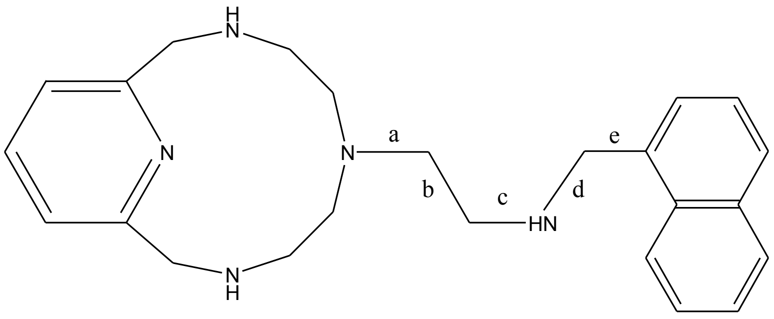


Figure 2. Labels for the dihedral angles

3. Results and Discussion

MM Conformational search. As the pH in the solvent decreases, the two macrocyclic secondary amines are first protonated, and then the bridge's secondary amine. Thus, H-bonds between the bridge and the macrocyclic amines and with the pyridine nitrogen in A are progressively weakening, and this makes conformation B more stable for the triprotonated species. To model these changes, MCMM was preferred to Molecular Dynamics (MD) trajectories because the conformational changes pursued were mainly torsions of the bridge atoms, and only MCMM give these straightforwardly. Furthermore, the groups attached to the extremes of the bridge were too large to switch efficiently under MD. Figure 2 shows the neutral chemical structure with the nomenclature used for the different dihedral angles that have been rotated in different runs. The labels point out the central bond that undergoes torsion for each dihedron. The conformational change between A and B can be done in several ways. It can imply complete torsion of the bond between the two carbon atoms in the bridge, b, and the consecutive bond that joins to the 12-ring amine a (eventually, e can be rotated to obtain the same enantiomer). Also, it can be done by inverting the 12-ring and rotating b and e. Thus, if we track the changes, we see that the sequence of torsion angles is not linear, but even corkscrewing in some cases. These are the changes that must overcome the main rotational barriers, since all the dihedral angles suffer minor changes to obtain B, due to the MM+ optimization.

Beginning with the closed conformation **A**, it was triprotonated and MCMM simulations were performed allowing different freedom degrees in different runs.

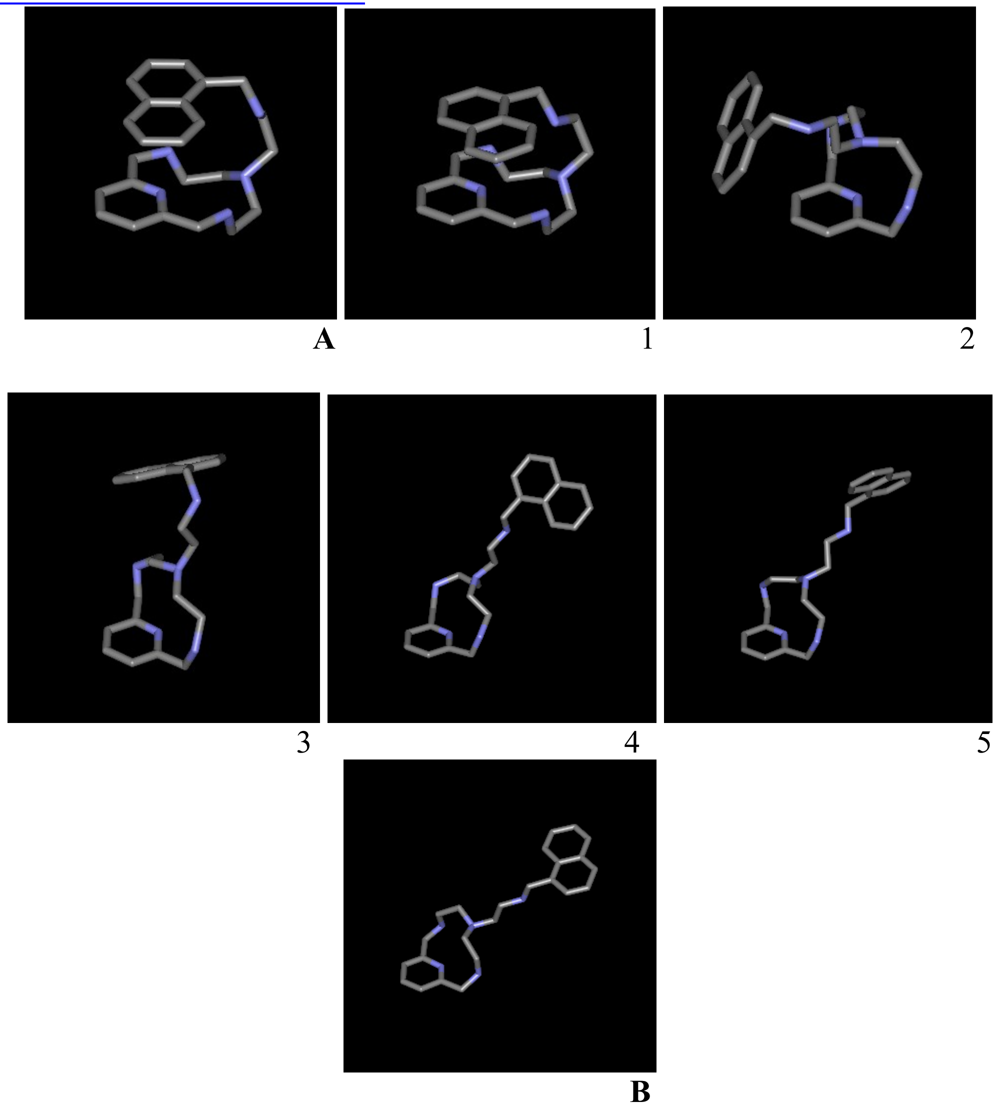


Figure 3. **A** and **B**, conformations from X-ray minimised in the periodic box. 1-5, conformations obtained from MCMM in vacuum and further minimisation in standard water density-constant TI3P box until their respective local minima, for different runs

These calculations were first attempted with the molecule in a water constant-density periodic box (standard water molecules TIP3P, equilibrated at 300 K and 1 atm., minimum distance between solvent and solute atoms: 2.3 Å), but the results did not converge and the program was unstable. For this reason, MCMM simulations were run in vacuum and the final conformations achieved were minimised in the referred periodic box boundary conditions.

It is possible ordering the resulting conformations obtained and figuring a movie in which **B** is obtained from **A** in successive steps, which are relative minima. Table 1 and Figure 3 display different conformations obtained as local minima, all of them from **A**, with their respective values for the torsion angles. As said, these conformations result from MCMM and minimisation within their corresponding water periodic boxes. For clarity, water molecules have been removed from Figure 3. Table 2 shows the parameters involved in the periodic box boundary step.

The conformations for the 12-ring obtained with MM+ are distorted with respect the X-ray diffraction model as well as the DFT simulations. Furthermore, it seems that ring torsion is not well achieved for macrocycles, with the algorithms currently implemented in MCMM for ring inversion. In spite of these drawbacks, conformation 5, near **B,** is obtained and the sequence can be illustrative of the conformation change.

Unfortunately, it was not possible obtaining a conformation near **A** from **B** with MCMM, due probably to the huge number of freedom degrees involved and the underestimation of H-bond interaction in MM+ that implied calculation times beyond our possibilities. The entropy of **B** must be the greatest.

.**Relative energies calculation.** DFT calculations to compare the relative stabilities of conformations for each protonation state were tried by using different exchange functionals and basis sets. Water environment was simulated by the COSMO method. Relativistic contributions were not considered significant because there were no heavy nuclei involved. The determinant importance of hydrogen bonds in the conformational equilibrium made necessary using several polarization functions. In fact, BP86/TZP and GGA-BP/TZ3P did not give good results, and all-electron QZ4P basis was necessary. All the calculations performed assigned the correct energy order for the triprotonated species: the extended conformation **B** was found more stable. But this result was also found with the monoprotonated stage when QZ4P was not used. LDA was useful to estimate properties in big molecules and it was the only method that allowed good results within reasonable calculation time (ca. eight hours with a Pentium 4 3.2GHz running on Windows XP, 1Gb RAM). Although the predicted values for atomization energies estimated are maybe not accurate, the method qualitatively predicted the experimental results. Table 3 shows the results obtained. The integration accuracy was four decimal places. Three decimal places were also tried to see the influence in the final result. It was seen that the dispersion of the result was lower than the differences between the values to be compared. The conclusions were, thus, unchanged.

In order to have a confirmation with a more complex functional of the trends obtained, GGA-BP/TZ2P was used for a simulation without solvent. Only two calculations were performed to compare atomization energies of the triprotonated species both, in its crystal conformation, **B**-open, and in the diprotonated one, **A**-closed. The integration accuracy was fixed in three decimal places. The calculation time for each process was eight days approximately with the same computer. These gave more reliable stability prediction at greater protonation state. The results are displayed in Table 3.

The optimized conformations obtained were compared with the crystal ones. Using the Carbó index measured through Coulomb integrals and the local Newton-Raphson as superposition algorithm, the similarity was quantified. Thus, monoprotonated **A** and its minimized conformation obtained by LDA/QZ4P showed Carbó index equal to 0.932, and triprotonated **B** and its respective optimization gave 0.964.

4. Conclusions

MCMM method was useful in the present study, but some drawbacks must be pointed out. Thus, ring torsions in macrocycles do not seem well simulated. The pH dependent opening of the scorpiand was easy to predict, but the reverse effect was not possible to be simulated.

DFT was a good method to give account of the experimental stability, although all-electron basis was necessary. The LDA approximation with the COSMO method was enough to obtain relative stabilities clearly according to the experiments at lower protonation. At more acidic pH, the more complex functional GGA-BP with a somewhat simple basis gave more reliable results.

The optimized conformations obtained were well-agree with the crystal ones, as verified by Quantum Molecular Similarity.

Acknowledgments

Financial support by the Spanish Ministerio de Economía y Competitividad (Projects CONSOLIDER INGENIO CSD-2010-00065 and CTQ2013-48917-C3-1-P), Generalitat Valenciana (Project PROMETEOII2015-002) is gratefully acknowledged.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Kinbara, K., & Aida, T. (2005). Toward intelligent molecular machines: directed motions of biological and artificial molecules and assemblies. *Chemical reviews*, *105*(4), 1377-1400.
2. Blasco, S., Verdejo, B., Bazzicalupi, C., Bianchi, A., Giorgi, C., Soriano, C., & García-España, E. (2015). A thermodynamic insight into the recognition of hydrophilic and hydrophobic amino acids in pure water by aza-scorpiand type receptors. *Organic & biomolecular chemistry*, *13*(3), 843-850.
3. Bernier, N., Esteves, C. V., & Delgado, R. (2012). Heteroditopic receptor based on crown ether and cyclen units for the recognition of zwitterionic amino acids. *Tetrahedron*, *68*(24), 4860-4868..
4. Aydın, I., Aral, T., Karakaplan, M., & Hoşgören, H. (2009). Chiral lariat ethers as membrane carriers for chiral amino acids and their sodium and potassium salts. *Tetrahedron: Asymmetry*, *20*(2), 179-183.
5. Verdejo, B., Acosta-Rueda, L., Clares, M. P., Aguinaco, A., Basallote, M. G., Soriano, C., ... & García-España, E. (2015). Equilibrium, Kinetic, and Computational Studies on the Formation of Cu2+ and Zn2+ Complexes with an Indazole-Containing Azamacrocyclic Scorpiand: Evidence for Metal-Induced Tautomerism. *Inorganic chemistry*, *54*(4), 1983-1991.
6. Verdejo, B., Ferrer, A., Blasco, S., Castillo, C. E., González, J., Latorre, J., ..& García-España, E. (2007). Hydrogen and copper ion-induced molecular reorganizations in scorpionand-like ligands. A potentiometric, mechanistic, and solid-state study. *Inorganic chemistry*, *46*(14), 5707-5719.
7. Pallavicini, P. S., Perotti, A., Poggi, A., Seghi, B., & Fabbrizzi, L. (1987). N-(aminoethyl) cyclam: a tetraaza macrocycle with a coordinating tail (scorpiand). Acidity controlled coordination of the side chain to nickel (II) and nickel (III) cations. *Journal of the American Chemical Society*, *109*(17), 5139-5144.
8. a) Chang, G., Guida, W. C., & Still, W. C. (1989). An internal-coordinate Monte Carlo method for searching conformational space. *Journal of the American Chemical Society*, *111*(12), 4379-4386; b) Saunders, M., Houk, K. N., Wu, Y. D., Still, W. C., Lipton, M., Chang, G., & Guida, W. C. (1990). Conformations of cycloheptadecane. A comparison of methods for conformational searching. *Journal of the American Chemical Society*, *112*(4), 1419-1427; c) Kolossváry, I., & Guida, W. C. (1993). Torsional flexing: Conformational searching of cyclic molecules in biased internal coordinate space. *Journal of computational chemistry*, *14*(6), 691-698.
9. Allinger, N. L. (1977). Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *Journal of the American Chemical Society*, *99*(25), 8127-8134.
10. Hypercube (2005). HYPERCHEM. Version 7.5. Hypercube Inc., 1115 NW 4th St., Gainsville, FL 32601-4256, USA.
11. Versluis, L., & Ziegler, T. (1988). The determination of molecular structures by density functional theory. The evaluation of analytical energy gradients by numerical integration. *The Journal of chemical physics*, *88*(1), 322-328.
12. a) Klamt, A., & Schüürmann, G. (1993). COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society, Perkin Transactions 2*, (5), 799-805; b) Pye, C. C., & Ziegler, T. (1999). An implementation of the conductor-like screening model of solvation within the Amsterdam density functional package. *Theoretical Chemistry Accounts*, *101*(6), 396-408.
13. Perdew, J. P., & Burke, K. (1996). Comparison shopping for a gradient-corrected density functional. *International journal of quantum chemistry*, *57*(3), 309-319.

14. ADF2006.01, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, http://www.scm.com

SciForum
Mol2Net

A Computer-Aided SAS Macro for the Evaluation of the Simulation Performances in Missingness Settings

Urko Aguirre [1,*], Inmaculada Arostegui[2,4] , Jose M. Quintana[3]

[1] Research Unit, REDISSEC: Red de Investigación en Servicios Sanitarios y Enfermedades Crónicas, Hospital Galdakao-Usansolo, Galdakao, Spain; urko.aguirrelarracoechea@osakidetza.eus.
[2] Department of Applied Mathematics, Statistics and Operational Research; REDISSEC: Red de Investigación en Servicios Sanitarios y Enfermedades Crónicas. Faculty of Science and Technology, Leioa. Spain; inmaculada.arostegui@ehu.es
[3] Research Unit, REDISSEC: Red de Investigación en Servicios Sanitarios y Enfermedades Crónicas, Hospital Galdakao-Usansolo, Galdakao, Spain; josemaria.quintanalopez@osakidetza.eus.
[4] BCAM-Basque Center for Applied Mathematics, Bilbao, Spain.

* Author to whom correspondence should be addressed;
E-Mail: urko.aguirrelarracoechea@osakidetza.eus; Tel.: +34 94 400 71 05; Fax: +34 94 400 71 32.

Published: 7 December 2015

Abstract: Model validation has become a topic of great interest to many fields such as industry, medicine or even to government. Its main challenge is to provide stable and credible tools so that the decision-maker with the information necessary can make high-consequence judgments. This process requires simulation modelling and consequently, some guidelines or evaluation criteria are essential in order to draw meaningful conclusions. A computer-aided SAS® macro is developed using the SAS/IML programming language. Researchers should provide the dataset to be analyzed and the true values to be compared. As a result, the statistical program shows measures (i.e., number of simulations to be performed, bias, accuracy, coverage, etc…) which help investigators to make decisions with a minimal effort of programming. Numerical results of the aforementioned statistical parameters, plots and a report are returned by the statistical tool. Although this macro is focused on the missingness setting, it is applicable to any other discipline. We encourage researchers to use it to make better statistical assessments of the used methods.

Keywords: simulation; missing data; SAS®; macro

Mol2Net YouTube channel: *http://bit.do/mol2net-tube*

1. Introduction

Statistical modelling is a powerful tool which is used to describe mathematically real-life world issues.

One of the most important roles of the statistical modelling is to create stable and robust mathematical expressions: models should be internally consistent and provide the best accurate estimates. Model validation is one of the approaches to address this problem.

Model validation is a straightforward and powerful process. It is concerned with building the right model. As noted, it is utilized to determine that a model is an accurate representation of the real system. Validation is usually achieved through the calibration of the model, an iterative process of comparing the model to actual system behavior and using the discrepancies between the two, and the insights gained, to improve the model. This process is repeated until model accuracy is judged to be acceptable. To this end, it requires simulation modeling.

Simulation is a computational technique that relies on repeating a computation on many different random samples in order to estimate a statistical quantity. These techniques provide empirical estimation of the sampling distribution of the parameters of interest that could not be achieved from a single study and enable the estimation of accuracy measures, such as the bias in the estimates of interest, as the truth is known [1]. Based on several parameters applied to the simulation performances, researchers would be able to determine whether the developed statistical model is acceptable.

The main objective of this manuscript is to provide a statistical tool that assess the performed simulation models.

2. Results and Discussion

Once executed the SAS® macro, results are given for the overall dataset or stratified by a specified group. These numerical outputs are displayed in a Word /PDF file or if desired, stored in a dataset (see Figure 1).

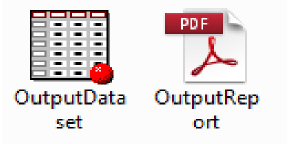


Figure 1. Ouput Delivery System of the SAS® macro.

The advances in computer technology have allowed simulation studies to be more accessible. This statistical tool allows to compute the most used parameters to make a proper simulation assessment. However, performing simulations is not simple; to perform a proper simulation assessment several parameters and designs should be taken into account.

3. Materials and Methods

The main routine is composed of explicit-passed several parameters needed for the assessment of the performed simulations. First of all, as input parameter, researchers should provide a dataset including the obtained simulation results. Moreover, the true values of the parameters to be assessed are also required. Finally, the type of missing data mechanisms (missing completely at random, missing at random and missing not at random) and missingness rate should be provided.

This statistical procedure evaluates the simulation performances according to the following mathematical expressions [2]:

- Raw Bias: It is defined as the difference between the true value and the mean average of the simulated coefficients.
- Relative bias: Relative bias was calculated by dividing the raw bias (difference between the mean value over simulation results and the true parameter) by the true value.
- Standardized bias: We compared the true value of the beta regression coefficient of the considered interaction factor in the model with the corresponding value obtained with each of the analyzed methods, relative to the standard error of the simulated value.
- Coverage: The coverage of a confidence interval is the proportion of times that the obtained confidence interval (using the imsulated values) contains the true specified

parameter value. If the coverage value is below the 90%, the performance of the interval procedure will be troublesome.
- Relative width: If one procedure has a similar or higher rate of coverage than another but yields intervals that are substantially shorter, then it should be preferred. Shorter intervals translate into greater accuracy and higher power.
The computer-aided SAS® macro is internally programmed using the SAS/IML language under SAS 9.4 release. Different SAS statements such as PROC MEANS or DATA steps are used to set up the aforementioned statistical parameters.

4. Conclusions

The SAS® macro is highly accessible and we encourage researchers to use it to make better statistical assessments of the used methods.

Author Contributions

All authors have equally contributed.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Burton A.; Altman,, D.G.; Royston, P.; Holder, R.L. The design of simulation studies in medical statistics. *Statistics in Medicine* **2006,** 25, 4279-4292.
2. Aguirre U.; Arostegui, I; Esteban, C; Quintana, J.M. Assessment of the performance of imputation techniques in observational studies with two measurements. *International Journal of Statistics in Medical Research* **2015**, 4(3), 240-251.

SciForum
Mol2Net

Building a New High-Selective Molecular Imprinted Polymer

Riccardo Concu [*], Maria Natalia DiasSoeiro Cordeiro [*]
REQUIMTE, Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal
* Correspondence: ric.concu@gmail.com, ncordeir@fc.up.pt

Published: 8 December 2015

Abstract: Molecular imprinted polymers (MIP) allows the preparation of tailored and high specific materials able to recognize a specific template. In this work, we simulated the affinity of a new high selective MIP able to specifically bind the isobutylphenylpropanoic acid (ibuprofen, template molecule). We have performed a series of molecular dynamics (MD) simulations of different mixtures in order to undercover the mechanisms occurring during the process of molecular imprinted polymers. The simulations were performed using the GROMACS 5.0 and the the OPLS-AA force field were used to parameterize and verifiy the studied molecules. A single system were simulated representing the pregelification state of the system. The radial distribution function (RDF) analysis and cluster analysis were used to evaluate the affinity of the template molecule, ibuprofen, for the gel backbone. Results confirm that the new material is high-selective and MD simulations are essential to study the molecular imprinting process because can give a deeper knowledge of the mechanism occurring during the imprinting process.

Keywords: Molecular imprinted polymers (MIP), molecular dynamics (MD), GROMACS, xerogel, polymers, isobutylphenylpropanoic acid.

1. Introduction

In recent years a new methodology have been developed to produce new and high selective polymers. Molecular imprintedpolymers (MIT) is a breaking through technology which is growing faster in these last years. For instance, the MITcan be used to prepare molecular imprinted polymers (MIP) that can be used to prepare synthetic receptors able to recognize and bind or release the template molecules, new HPLC matrix for selective detection and/or separation of drugs.In this context, MIPmaterials are gaining day by day a most relevant role due to the growing demand for sensitive, accurate and simple methods and materials able to achieve this goal. In fact, MIPare widely used because they are able to recognize small chemicals or large biological molecules such as, proteins, DNA or RNA. In addition, MIP can be used to create sorbents for specific chiral chromatographic materials or specific sorbents for high-performance liquid chromatography-ultraviolet detection (HPLC-UV). The creation of new drug release materials is also an application of sol-gel MIP.

In this short communication, we present a Molecular Dynamics (MD) simulationof a new MIP, with a high selectivity for the isobutylphenylpropanoic acid (ibuprofen, template molecule). The radial distribution function (RDF) analysis has been used to evaluate the affinity between the MIP and the template molecule. This is a preliminar study to assess the affinity between the polymer and a template molecule which is a modification of a ORMOSIL we have recently published[1].

2. Results and Discussion

The RDF analysis was used to study the affinity between the template and the polymer. As referred in the materials and method section, the RDF was calculated using a specific atom instead of the center of the mass of the molecule. The atoms used for the template are the two oxygens of the carboxylic terminal, while for the polymer the hydrogen of the dehydroimidazolium; these atoms have been circle-marked in the **Figure 1**. In the **Figure 2** we have reported the RDF analysis between IBU$^-$(the template molecule) and a cationic dehydroimidazolium ORMOSIL (DHI$^+$, [$Si_3O_3(CH_3)_4(OH))_2$-$(C_3H_6)]_2$-$C_3H_5N_2^+$). In the image is clear the affinity between the template and the polymer; in fact, there are two sharp and high peaks at a distance of 0.25nm which clearly confirms the affinity. This peaks in fact correspond to the two oxygens of the template interacting with the hydrogen of the dihydroimidazolium of the ORMOSIL. In addition, in the same figure we have reported the RDF of the ORMOSIL with the counter ion, but in this case there is no relevant affinity. Thus, this result can confirms that is likely to happen a successfully imprinting effect.

Figure 1. IBU$^-$ and ORMOSIL structure

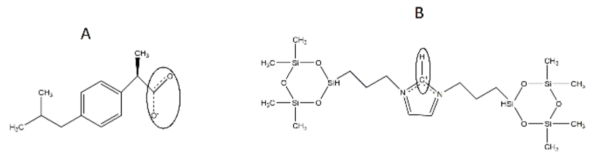
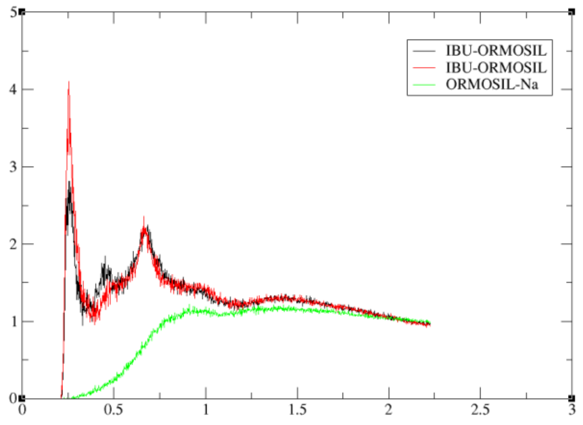


Figure 2. RDF analysis.



3. Materials and Methods

The MD simulations were performed with GROMACS 5.0.4.package applying the OPLS-AA[2,3]. The system under study contained water, methanol, the anyonic form of Ibuprofen (the template, IBU$^-$), the dual cyclic silicate trimer corresponding to a hydrolyzed and condensed species derived from the cationic dehydroimidazolium ORMOSIL (DHI^+, $[Si_3O_3(CH_3)_4(OH))_2-(C_3H_6)]_2-C_3H_5N_2^+$),in the Figure 1 we have reported the ORMOSIL and the IBU^- structures.The initial state of the system was obtained by inserting into the boxes the respective number of units at random positions using the packmol package[4]. The composition of the model is reported in the **Table 1**. After energy minimization using steepest-descent methods included in the GROMACS package, a temperature annealing was performed in the *NVT* ensemble for 1ns, reaching a temperature of 600 K, so as to ensure a proper mixing and gather three random independent initial configuration. Then the system were simulated for a total of 20ns in the *NpT* ensemble for data collection.Observable properties were sampled every 2 ps, from which total averages and standard deviations for each run were computed.The analysis consisted essentially in the calculation of radial distribution functions (RDF). The RDF between different types of molecules has been calculated as:

$$g_{AB}(r) = \frac{\langle\rho_B(r)\rangle}{\langle\rho_B\rangle_{loc}},$$

where $<\rho_B(r)>$ refers to the average density of particle B at a distance *r*, around the particle A, and $<\rho_B>_{loc}$ refers to the density of the particle B averaged over all spheres around particles A with a maximum radius (r_{max}) which was half of the box length.

Table 1. Composition of the model

Molecule	Number
Ibuprofen	10
Na	20
Water	230
Methanol	1130
Iodum	20
Ormosil	10

4. Conclusions

This is only a preliminary work in order to assess the affinity between the template and the ORMOSIL molecule. Considering the reported results, we can affirm that a molecular imprinting process in this system is likely to happen. In addition, this work demonstrates that MD simulations could be useful to undercover atomistic basis of a imprinted process.

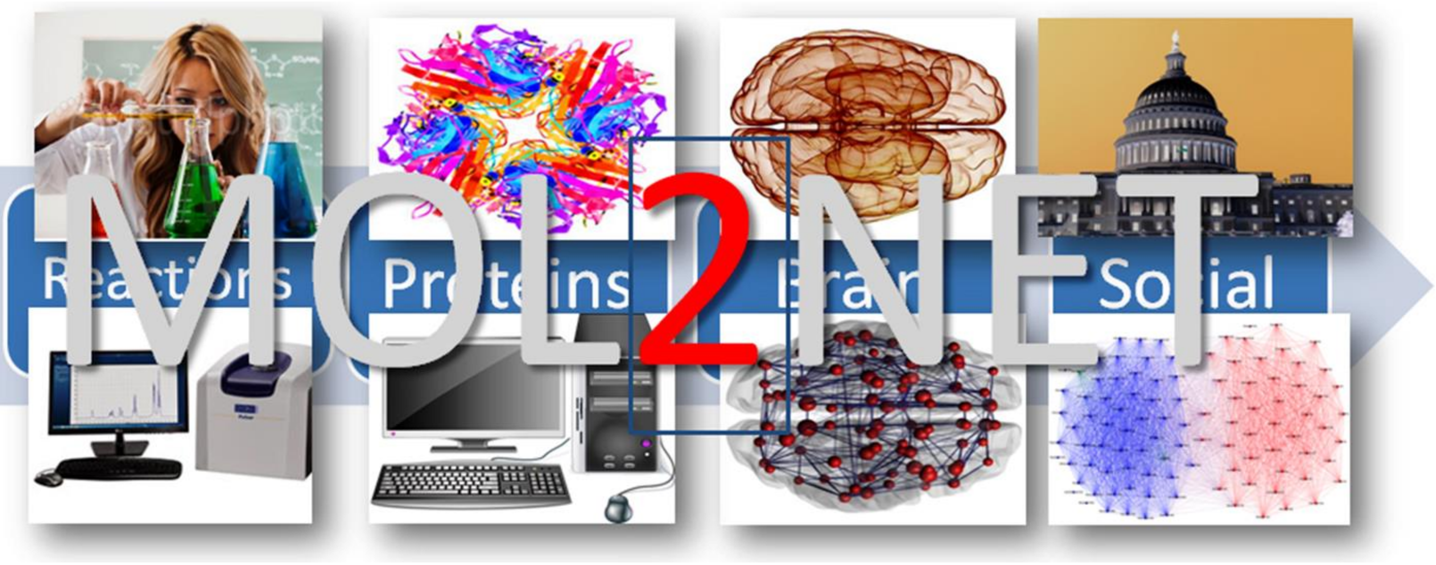
Acknowledgments

This work received financial support from the European Union (FEDER funds through COMPETE) and National Funds (FCT, Fundação para a Ciência e a Tecnologia) through projects Pest-C/EQB/LA0006/2013 (REQUIMTE) and Pest-C/QUI/UI0081/2013 (CIQ). The work also received financial support from the European Union (FEDER funds) under the framework of QREN through Project NORTE-07-0124-FEDER-000067-NANOCHEMISTRY. RC acknowledges also FCT and the European Social Fund for financial support (Grant SFRH/BPD/80605/2011). To all financing sources the authors are greatly indebted.

References and Notes

1. Concu, R.; Perez, M.; Cordeiro, M.N.; Azenha, M. Molecular dynamics simulations of complex mixtures aimed at the preparation of naproxen-imprinted xerogels. *Journal of chemical information and modeling* **2014**, *54*, 3330-3343.
2. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J. Gromacs: Fast, flexible, and free. *Journal of computational chemistry* **2005**, *26*, 1701-1718.
3. Jorgensen, W.L.; Maxwell, D.S.; Tirado-Rives, J. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society* **1996**, *118*, 11225-11236.
4. Martinez, L.; Andrade, R.; Birgin, E.G.; Martinez, J.M. Packmol: A package for building initial configurations for molecular dynamics simulations. *Journal of computational chemistry* **2009**, *30*, 2157-2164.

IKERBASQUE UPVEHU



sciforum

Basel, Switzerland
ISBN: 978-3-03842-817-6
ISSN: 2624-5078
Year: **2018**