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

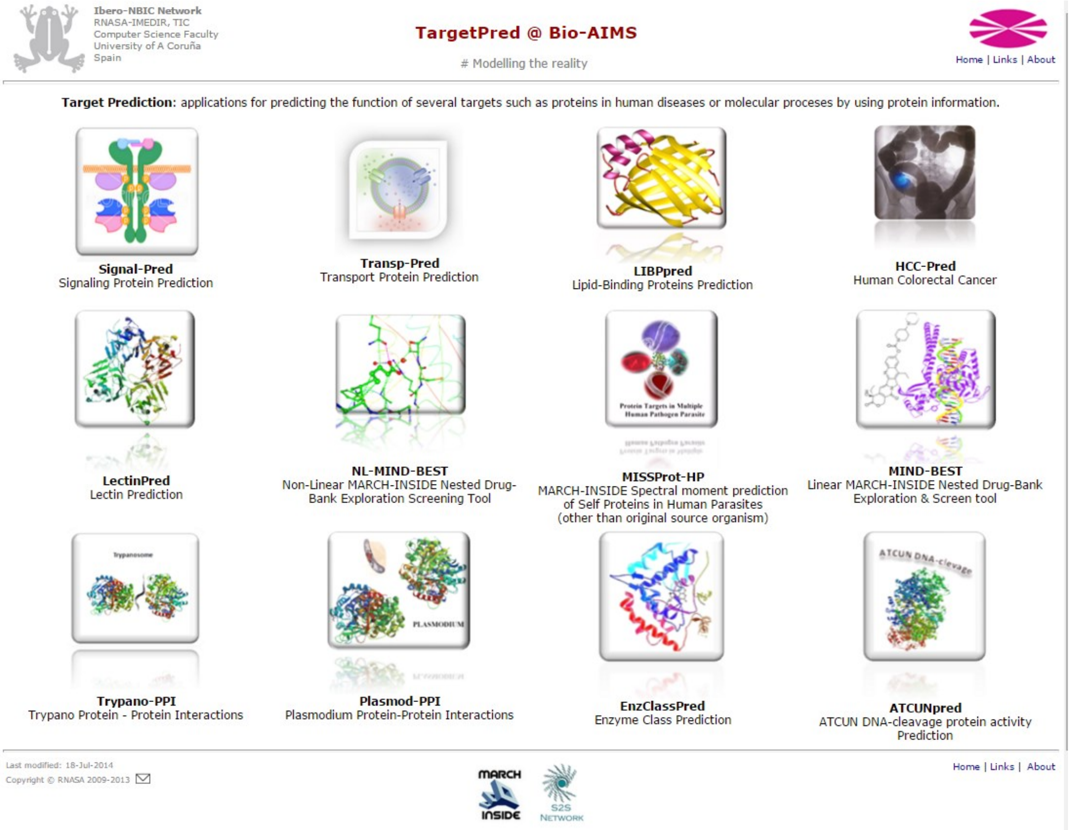
The *in silico* screening methods are very important in Drug Development or proteomics. The theoretical screening is a fast and low cost option to filter the large number of molecules or macromolecules for a specific biological action or chemical property.

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

Target Prediction: applications for predicting the function of several targets such as proteins in human diseases or molecular proceses by using protein information.

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

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