MOL2NET, International Conference Series on Multidisciplinary Sciences

New tool useful for drug discovery validated through benchmark datasets

Yoan Martínez-López[a,b,c], Yaile Caballero[b,c], Stephen J. Barigye[a], Yovani Marrero-Ponce[a,d], Reisel Millán Cabrera[a], Julio Madera[b], Efrain Edgar Chaluisa Quishpe [b,c], Francisco Torrens[e] and Juan Alberto Castillo-Garit[*,a,f]*

[a]*Unit of Computer-Aided Molecular "Biosilico" Discovery and Bioinformatics Research (CAMD-BIR Unit), Faculty of Chemistry-Pharmacy. Universidad Central "Marta Abreu" de Las Villas, Santa Clara, 54830, Villa Clara, Cuba*

[b]*Department of Computer Sciences, Faculty of Computer Sciences, Camagüey University, Camagüey city, 74650, Camagüey, Cuba*

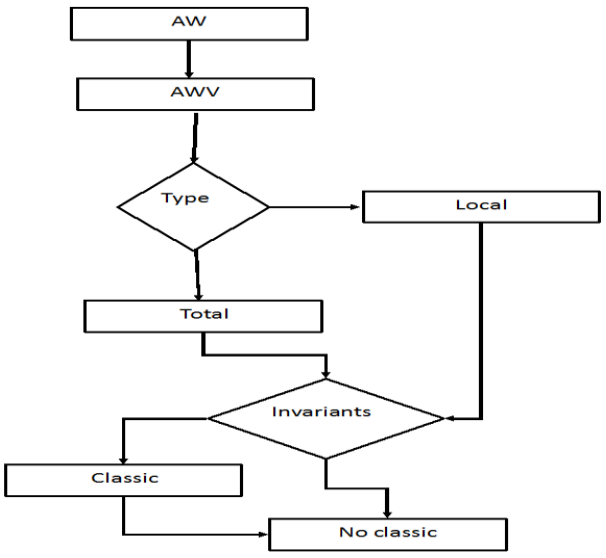
[c]*Artificial Intelligence Researches Group (AIRES), Faculty of Computer Sciences, Camagüey University, Camagüey, Cuba;*

[d]*Grupo de medicina molecular y traslacional (MEM&T), Universidad San Francisco de Quito (USFQ), Quito, Ecuador*

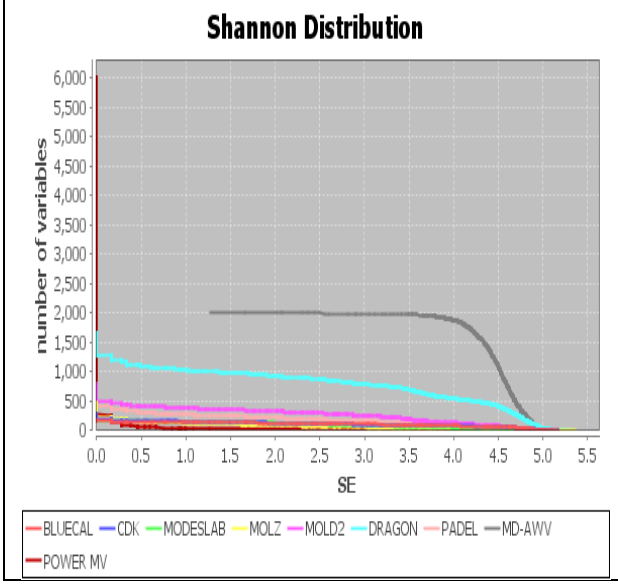
[e]*Institut Universitari de Ciència Molecular, Universitat de València, Edifici d'Instituts de Paterna, P.O. Box 22085, E-46071, València, Spain;*

[f]*Unidad de Toxicología Experimental, Universidad de Ciencias Médicas de Villa Clara, Carretera a acueducto y circunvalación, Santa Clara, Villa Clara, Cuba. CP: 50200, Cuba.*

Graphical Abstract



Abstract. Atomic Weighted Vectors (AWVs) are vectors that contain the codified information of molecular structures, which can apply to a set of Aggregation Operators (AOs) to calculate total and local molecular descriptors (MDs). This article presents an exploratory study of a new tool useful for drug discovery using different datasets, such as DRAGON and Sutherland's datasets, as well as their comparison with other well-known approaches. In order to evaluate the performance of the tool, several statistics and QSAR/QSPR experiments were performed. Variability analyses are used to quantify the information content of the AWVs obtained from the tool, by the way of an information theory-based algorithm. Principal Components Analysis (PCA) is used to analyze the



orthogonality of these descriptors, for which the new MDs from AWVs provide different information from those codified by DRAGON descriptors (0-2D). The QSAR models are obtained for every Sutherland's dataset, according to the original division into training/test sets, by means of the Multiple Linear Regression with Genetic Algorithm (MLR-GA). These models have been validated and compare favorably to several approaches previously published, using the same benchmark datasets. The obtained results show that this tool should be a useful strategy for the QSAR/QSPR studies, despite its simplicity.

References

[1] Malo, M.; Persson, R.; Svensson, P.; Luthman, K.; Brive, L. Development of 7TM receptor-ligand complex models using ligand-biased, semi-empirical helix-bundle repacking in torsion space: application to the agonist interaction of the human dopamine D2receptor. *J. Comput. Aided Mol. Des*, **2013,** *27*, 277–291.

[2] Castillo-Garit, J.A.; Del Toro-Cortés, O.; Kouznetsov, V.V.; Puentes, C.O.; Romero Bohórquez, A.R.; Vega, M.C.; Rolón, M.; Escario, J.A.; Gómez-Barrio, A.; Marrero-Ponce, Y.; Torrens, F.; Abad, C. Identification In Silico and In Vitro of Novel Trypanosomicidal Drug-Like Compounds. *Chem. Biol. Drug Des.*, **2012,** *80* (1), 38-45.

[3] Gonzalez-Diaz, H.; Riera-Fernández, P. New Markov-autocorrelation indices for re-evaluation of links in chemical and biological complex networks used in metabolomics, parasitology, neurosciences, and epidemiology. *J. Chem. Inf. Mod.*, **2012,** *52* (12), 3331-3340.

[4] Välitalo, P.A.J.; Griffioen, K.; Rizk, M.L.; Visser, S.A.G.; Danhof, M.; Rao, G.; van der Graaf, P.H.; van Hasselt, J.G.C. Structure-Based Prediction of Anti-infective Drug Concentrations in the Human Lung Epithelial Lining Fluid. *Pharm. Res.*, **2016,** *33* (4), 856-867.

[5] Todeschini, R.; Consoni, V. New Local Vertex Invariants and Molecular Descriptors Based on Functions of the Vertex Degrees. *MATCH Commun. Math. Comput. Chem.*, **2010,** *64*, 359-372.

[6] Papadatos, G.; Gaulton, A.; Hersey, A.; Overington, J.P. Activity, assay and target data curation and quality in the ChEMBL database. *J. Comput. Aided Mol. Des.*, **2015,** *29*, 885–896.

[7] Gugisch R, Kerber A , Laue R, Meringer M, Weidinger J *MOLGEN*, 4.0; University of Bayreuth, D-95440 Bayreuth, Germany, 1981.

[8] Stefik, M. Planning with constraints (MOLGEN: Part 1). *Artif. Intell.*, **1981,** *16* (2), 111-139.

[9] Benecke, C.; Grüner, T.; Kerber, A.; Laue, R.; Wieland, T. MOLecular structure GENeration with MOLGEN, new features and future developments. *Fresen. J. Anal. Chem.*, **1997,** *359* (1), 23-32.

[10] Ivanciuc, O. CODESSA Version 2.13 for Windows. *J. Chem. Inf. Comp. Sci.*, **1997,** *37* (2), 405-406.

[11] Castillo-Garit, J.A.; Marrero-Ponce, Y.; Torrens, F.; García-Domenech, R.; Rodríguez-Borges, J.E. Applications of Bond-Based 3D-Chiral Quadratic Indices in QSAR Studies Related to Central Chirality Codification. *QSAR & Comb. Sci.*, **2009,** *28*, 1465-1477.

[12] Martínez-López, Y.; Marrero-Ponce, Y.; Barigye, S. *MD-LOVIs for windows: Software for molecular de scriptors calculator from LOVIs. Version 1.0*, 2010.

[13] González-Dıaz, H.; Molina-Ruiz, R.; Hernandez, I. MARCH-INSIDE version 3.0. *MARkov CHains INvariants for SImulation and DEsign*, **2007**.

[14] Castillo-Garit, J.A.; del Toro-Cortés, O.; Vega, M.C.; Rolón, M.; Rojas de Arias, A.; Casañola-Martin, G.M.; Escario, J.A.; Gómez-Barrio, A.; Marrero-Ponce, Y.; Torrens, F.; Abad, C. Bond-based bilinear indices for computational discovery of novel trypanosomicidal drug-like compounds through virtual screening. *Eur. J. Med. Chem.*, **2015,** *96* (0), 238-244.