



SciForum
MOL2NET

Molecular descriptor from atomic weighted vectors to predict aquatic toxicity.

Yoan Martínez-López^{1, 2}, Stephen J. Barigye³, Oscar Martínez-Santiago², Yovani Marrero-Ponce⁴, James Green⁵ and Juan A. Castillo-Garit^{2,6,*}

¹ Department of Computer Sciences, Faculty of Informatics, Camaguey University, Camaguey City, 74650, Camaguey Cuba.

² Unit of Computer-Aided Molecular “Biosilico” Discovery and Bioinformatic Research (CAMD-BIR Unit), Faculty of Chemistry-Pharmacy. Universidad Central “Martha Abreu” de Las Villas, Santa Clara, 54830, Villa Clara, Cuba.

³ Departamento de Química, Universidade Federal de Lavras, CP 3037, 37200-000, Lavras, MG, Brazil.

⁴ Environmental and Computational Chemistry Group, Facultad de Química Farmacéutica, Universidad de Cartagena, Cartagena de Indias, Bolívar, Colombia.

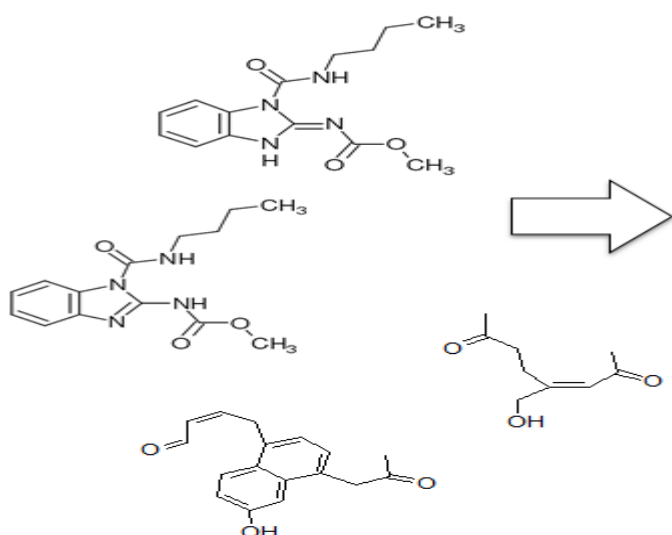
⁵ Department of Systems and Computer Engineering, Carleton University, Ottawa, Ontario, Canada.

⁶ Unidad de Toxicología Experimental, Universidad de Ciencias Médicas “Dr. Serafin Ruiz de Zárate Ruiz”, Santa Clara 50200, Villa Clara, Cuba

* Author to whom correspondence should be addressed; E-Mail: jacgarit@yahoo.es; Tel.: +53-42-273236; Fax: +53-42-271480.

Abstract: Molecular descriptors from atom weighted vectors (MD-AWV) are used in the prediction of aquatic toxicity of a large group of organic compounds of 392 benzene derivatives to the protozoan ciliate *Tetrahymena pyriformis* ($\log(\text{IGC}_{50})^{-1}$). These descriptors are calculated using the MD-LOVIs software and various Aggregation Operators are examined with the aim comparing their performances in predicting aquatic toxicity. Variability analysis is used to quantify the information content of these molecular descriptors by means of an information theory-based algorithm. Principal Component Analysis (PCA) is used to analyze the orthogonality of these descriptors and it is observed that MD-AWV provide linearly independent information from that of descriptors generated using the popular DRAGON package (0-2D). Multiple Linear Regression with Genetic Algorithms is used to obtain models of the structure–toxicity relationships; the best model shows values of $Q^2=0.830$ and $R^2=0.837$ using six variables. Our models compare favorably with other previously published models that use the same data set. The obtained results suggest that MD-AWV provide an effective alternative for determining aquatic toxicity of benzene derivatives.

Keywords: Aggregation Operator, Atom Weighted Vectors, Aquatic Toxicity

Graphical Abstract:

$$X = [X_1, X_2, \dots, X_N]$$

$$DE = \sqrt{\frac{(\sum X_a - M)^2}{n-1}}$$

Introduction:

Benzene derivatives are widely used in industry as pesticides, insecticides, herbicides, lubricants, detergents, polymers, solvents, and in the manufacturing of plastics, resins and nylon [1-2]. Likewise, in the pharmaceutical industry, they are often used as drugs. Many of these derivatives can cause damage to the environment, humans, animals, and plants due to their toxicity. Environmental contamination, in addition to the possible accumulation of substituted benzenes in water and soil, make them potentially damaging chemicals [1, 3]. There is an increasing interest in the toxicity of environmental chemicals. Quantitative structure-activity/toxicity relationships (QSAR/QSTR) studies provides an invaluable tool in the prediction of aquatic toxicity directly from the molecular structure of compounds [3]. The inhibition of growth database [4] of the ciliated protozoan *Tetrahymena pyriformis* is considered to be a high quality dataset [5-6]. It has been developed in a single laboratory over more than two decades. Several studies have attempted to predict the toxicity of benzene derivatives toward *T. pyriformis* using the known toxicity database of 392 benzene derivatives [3, 7-8]. Several other applications of QSAR in toxicology are presented in comprehensive reviews [9-12]. The aim of this paper is to compare the performance of a new family of molecular

descriptors (MD), derived from Atomic Weighted Vectors (AWV) [13], in predicting the aquatic toxicity of a large group of substituted benzenes. These MD are derived from AWV using Aggregation Operators (AOs) to convert the AWV into scalar quantities describing aspects of the molecule. These MD-AWVs are then evaluated for their applicability in ecotoxicological research.

Materials and Methods:

A chemical dataset of 392 benzene derivatives tested in *T. pyriformis* was used to create and assess models of aquatic toxicity. The data comprise diverse structural substituted benzene molecules containing nitrobenzenes, phenols, aminobenzenes, and benzenitriles. For predictive model development, we make use of the same training set and test set previously published [7].

An AWV is a representation, in n space, of the weights of all atoms within a structure. Here, represents the real number space and n represents the number of atoms within the molecular structure represented by the AWV.

$$X = [X_1, X_2, \dots, X_N] \in \mathbb{R}^n \quad (1)$$

On the other hand, an AO is a function which gives a real number y to an n -dimensional vector of real numbers $[X_1, X_2, \dots, X_N]$ [13]:

$$Y = \text{AO}([X_1, X_2, \dots, X_N]) \quad (2)$$

The information codified in a weighted MS may be estimated in an \mathbb{R}^n space, where \mathbb{R} a set of

real numbers and N is the number of atoms in a molecule structure. If the molecular vector W , in the \mathbb{R}^n space is considered to start from the origin, then the vector components represent given atom, atom-type or group properties [13]. This MD were implemented in a software package called MD-LOVIs (Molecular Descriptors from Local Vertex Invariants and Related Maps), which is available in <http://www.tomocomd.com>.

Results and Discussion:

The PCA (Principal Component Analysis) results suggest that the MD-AWVs codify different information than that provided by the DRAGON (0-2D) descriptors. On the other hand, despite the existence of some overlap between the MD-AWV and the DRAGON (0-2D) descriptors, many of the new MD-AWVs are orthogonal to the descriptors implemented in the DRAGON software. This permits us to say that our models encode new chemical information from the molecular structure.

Conclusions:

The results obtained with the MD-AWV models were compared to several QSAR procedures reported in the literature according to the correlation coefficients achieved with the leave-one-out cross-validation (Q^2_{100}) and square coefficient relation (R^2) methods. Generally, performance was observed to be highly competitive with the state of the art.

Conflicts of Interest: The authors declare no conflict of interest

References:

- [1] Huff, J. Benzene-induced cancers: abridged history and occupational health impact. *Int. J. Occup. Environ. Health*, **2007**, *13*, 213–221.
- [2] Tranfo, G., *Benzene and its Derivatives: New Uses and Impacts on Environment and Human Health*. Nova Science Pub Incorporated: Rome, **2011**.
- [3] Salahinejad, M.; Ghasemi, J.B. 3D-QSAR studies on the toxicity of substituted benzenes to *Tetrahymena pyriformis*: CoMFA, CoMSIA and VolSurf approaches. *Ecotoxicology and Environmental Safety* **2014**, *105*, 128–134.
- [4] Schultz, T.W. TERATOX: *Tetrahymena pyriformis* population grow impairment endpoint- A surrogate for fish lethality. *Toxicol. Methods*, **1997**, *7*, 289-309.
- [5] Brito-Sánchez, Y.; Castillo-Garit, J.A.; Le-Thi-Thu, H.; González-Madariaga, Y.; Torrens, F.; Marrero-Ponce, Y.; Rodríguez-Borges, J.E. Comparative study to predict toxic modes of action of phenols from molecular structures. *SAR QSAR Environ Res*, **2013**, *24* (3), 235-251.
- [6] Bradbury, S.P.; Russom, C.L.; Ankley, G.T.; Schultz, T.W.; Walker, J.D. Overview of data and conceptual approaches for derivation of quantitative structure-activity relationships for ecotoxicological effects of organic chemicals. *Environ. Toxicol. Chem.*, **2003**, *22* (8), 1789-1798.
- [7] Castillo-Garit, J.A.; Marrero-Ponce, Y.; Escobar, J.; Torrens, F.; Rotondo, R. A novel approach to predict aquatic toxicity from molecular structure. *Chemosphere*, **2008**, *73*, 415-427.

The obtained MD-AWV-based models were demonstrated to be effective in terms of the R^2 and Q^2_{100} values. The model developed with MD-AWV for six variables is showed below

$$\begin{aligned} \text{Log}(1/\text{IGC50}) = & -1.917(\pm 0.096) \\ & +0.179(\pm 0.009)M^Y_{\text{RA}} \\ & +0.867(\pm 0.109)R^P_{\text{RA}} \\ & +0.451(\pm 0.019)N1^L_{\text{HT}} \\ & -0.010(\pm 7.4 \times 10^4)N3^{\text{IN}11}_{\text{HT}} \\ & +1 \times 10^4(\pm 1 \times 10^4)AC1(\text{PN})^{\text{IN}14}_{\text{HT}} \end{aligned}$$

$$\begin{aligned} R^2 = 89.83 & & F = 64.74 & & s = 0.8009 \\ Q^2_{\text{Loo}} = 87.45 & & Q^2_{\text{Boot}} = 86.02 & & s_{\text{cv}} = 0.7464 \\ a[R^2] = 0.087 & & a[Q^2] = -0.252 & & Q^2_{\text{ext}} = 66.11 \end{aligned}$$

These results suggest that the correlations obtained with MD-AWV-based models for the prediction of aquatic toxicity can be considered to be statistically significant despite their simplicity. Our results were compared with other approaches that use the same dataset and our model with less attributes achieved similar to better statistical parameters than other models.

- [8] Bordbar, M.; Ghasemi, J.; Faal, a.Y.; Fazaeli, R. Chemometric Modeling to Predict Aquatic Toxicity of Benzene Derivatives Using Stepwise-Multi Linear Regression and Partial Least Square. *Asian J. Chem.*, **2013**, *25* (1), 331-342.
- [9] Schultz, T.W.; Tucker, V.A. Structure-toxicity relationships for the effects of N- and N,N'-alkyl thioureas to *Tetrahymena pyriformis*. *Bull Environ Contam Toxicol*, **2003**, *70* (6), 1251-8.
- [10] Castillo-Garit, J.A.; Abad, C.; Rodriguez-Borges, J.E.; Marrero-Ponce, Y.; Torrens, F. A Review of QSAR studies to Discover New Drug-like Compounds Actives Against Leishmaniasis and Trypanosomiasis. *Curr. Top. Med. Chem.*, **2012**, *12* (8), 852-865.
- [11] Gleeson, M.P.e.a. The challenges involved in modeling toxicity data insilico: a review. *Curr. Pharm. Des.*, **2012**, *18*, 1266–1291.
- [12] Castillo-Garit, J.A.; Abad, C.; Casañola-Martin, G.M.; Jones Barigye, S.; Torrens, F.; Torreblanca, A. Prediction of Aquatic Toxicity of Benzene Derivatives to *Tetrahymena pyriformis* According to OECD Principles. *Curr. Pharm. Des.* , **2016**, *22* (33), 5085-5094.
- [13] Martínez-López, Y.; Marrero-Ponce, Y.; Echeverri Jaramillo, G.; Barigye, S.J. The Summation of Atomic Contributions is an Overly Simplified Characterization of the Holistic Molecular Behavior. *Lett. Drug Des. Discov.*, **2016**, *13*, 12.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions defined by MDPI AG, the publisher of the Sciforum.net platform. Sciforum papers authors the copyright to their scholarly works. Hence, by submitting a paper to this conference, you retain the copyright, but you grant MDPI AG the non-exclusive and unrevocable license right to publish this paper online on the Sciforum.net platform. This means you can easily submit your paper to any scientific journal at a later stage and transfer the copyright to its publisher (if required by that publisher). (<http://sciforum.net/about>).