MOL2NET, **2020**, 6, **ISSN**: 2624-5078,
http://sciforum.net/conference/mol2net-06/chembiomol-01

CHEMBIOMOL-01: Chem. Biol. & Med. Chem. Workshop, Rostock, Germany-Galveston, Texas, USA, 2020

MOL2NET, International Conference Series on Multidisciplinary Sciences

In silico toxicity prediction of phenol derivatives with ISIDA descriptors using multiple linear regression and machine learning approach

Karel Diéguez-Santana [a, b], *Amilkar Puris* [c], *Oscar M. Rivera-Borroto* [d, e, f], and *Gerardo M. Casañola-Martin* [g]

[a] *Departamento Ciencias de la Vida, Universidad Estatal Amazónica, km 2 1/2 Vía Tena, Puyo, Pastaza, Ecuador.*

[b] *Department of Organic Chemistry II, University of the Basque Country UPV/EHU, 48940, Leioa, Spain.*

[c] *Facultad de Ciencias de La Ingeniería, Universidad Técnica Estatal de Quevedo, Ecuador.*

[d] *Department of Dev. Mathematics, Houston Community College-West Loop Campus, Houston TX, 77081, USA.*

[e] *Department of Mathematics, Lone Star College-CyFair Campus, Houston, TX, 77433, USA*

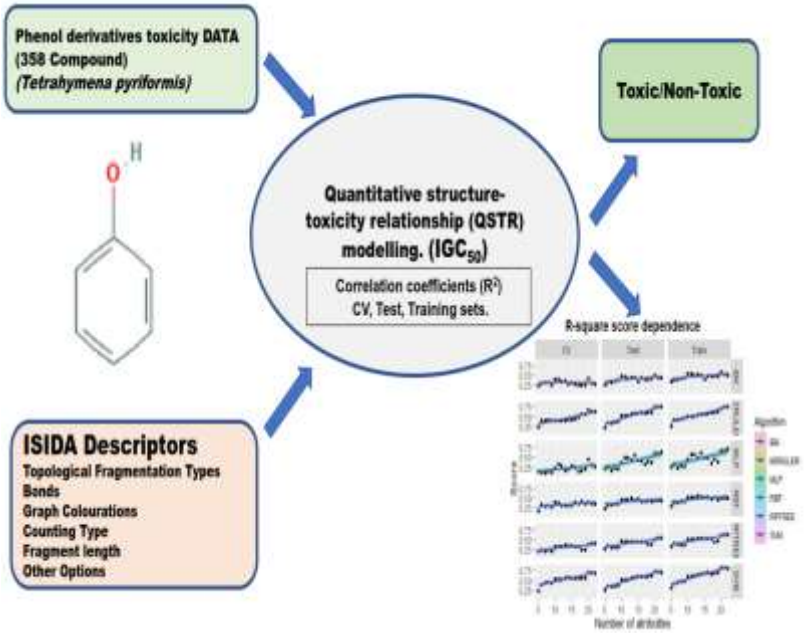
[f] *Departamento de Química Física Aplicada, Facultad de Ciencias, Universidad Autonoma de Madrid, 28049 Madrid, Spain.*

[g] *Department of Coatings and Polymeric Materials, North Dakota State University, Fargo, ND, 58102, USA.*

*** Corresponding authors:** K. Dieguez-Santana (karel.dieguez.santana@gmail.com) and G. M. Casañola-Martin gmaikelc@gmail.com)

..

Graphical Abstract



Abstract.

Phenolic compounds are considered as dangerous pollutants, which produces serious environmental problems by pollution of water streams because of their great water solubility and high toxicity. In this paper we present the modeling of inhibitory grown activity against Tetrahymena pyriformis with structural feature descriptors. Quantitative structure-toxicity relationship model for acute toxicity of phenol derivatives was performed using Multiple Linear Regression (MLR), Reduced Error Pruning Tree (REPTree), M5 Model Rules (M5R), Multilayer Perceptron (MLP), Instance-Based Learning algorithms using K nearest neighbor (IBk-ANN), Support vector machine (SVM), and Radial basis function network (RBF). The correlation coefficients (R^2) of training sets and test sets were 0.88 and 0.86 for the best MLR model, 0.82 and 0.72 for the best machine learning model (SVM), respectively. Following to the obtained results, our proposed model may be useful to predict of toxicity and risk assessment of phenol derivatives compound.

Keywords: Phenolic compounds; acute toxicity; *Tetrahymena pyriformis*; structural feature descriptors; Support vector machine

References

.

1. Chen, X.H.; Shan, Z.J.; Zhai, H.L. QSAR models for predicting the toxicity of halogenated phenols to Tetrahymena. *Toxicological and Environmental Chemistry* **2017**, *99*, 273-284, doi:10.1080/02772248.2016.1165818.
2. Dieguez-Santana, K.; Pham-The, H.; Villegas-Aguilar, P.J.; Le-Thi-Thu, H.; Castillo-Garit, J.A.; Casañola-Martin, G.M. Prediction of acute toxicity of phenol derivatives using multiple linear

regression approach for Tetrahymena pyriformis contaminant identification in a median-size database. *Chemosphere* **2016**, *165*, 434-441, doi:10.1016/j.chemosphere.2016.09.041.
3. Enoch, S.J.; Cronin, M.T.D.; Schultz, T.W.; Madden, J.C. An evaluation of global QSAR models for the prediction of the toxicity of phenols to Tetrahymena pyriformis. *Chemosphere* **2008**, *71*, 1225-1232, doi:10.1016/j.chemosphere.2007.12.011.
4. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter* **2009**, *11*, 10-18.
5. Basant, N.; Gupta, S.; Singh, K.P. Modeling the toxicity of chemical pesticides in multiple test species using local and global QSTR approaches. *Toxicology Research* **2015**, *5*, 340-353, doi:10.1039/c5tx00321k.
6. Abbasitabar, F.; Zare-Shahabadi, V. In silico prediction of toxicity of phenols to Tetrahymena pyriformis by using genetic algorithm and decision tree-based modeling approach. *Chemosphere* **2017**, *172*, 249-259, doi:10.1016/j.chemosphere.2016.12.095.