Advances in Biomechanics
2017; 1(2): 42-46
http://www.sciencepublishinggroup.com/j/abm
doi: 10.11648/j.abm.20170102.13

Density Functional Theory (DFT) Based Quantitative Structure Toxicity Relationship (QSTR) Modelling of the Acute Toxicity of Phenols

Ibraheem Wasiu Aderemi[1, *], Ameji Philip John[1], Awor George Okorn[1], Racheal U. Akpa[1], Raji Saheed Akinleye[2]

[1]Department of Chemistry, Ahmadu Bello University, Zaria, Kaduna, Nigeria

[2]Chemical Engineering Department, Ladoke Akintola University of Technology, Ogbomosho, Nigeria

Email address:

okaye4real@gmail.com (I. W. Aderemi)

[*]Corresponding author

To cite this article:

Ibraheem Wasiu Aderemi, Ameji Philip John, Awor George Okorn, Racheal U. Akpa, Raji Saheed Akinleye. Density Functional Theory (DFT) Based Quantitative Structure Toxicity Relationship (QSTR) Modelling of the Acute Toxicity of Phenols. *Advances in Biomechanics*. Vol. 1, No. 2, 2017, pp. 42-46. doi: 10.11648/j.abm.20170102.13

Received: April 26, 2017; **Accepted**: May 9, 2017; **Published**: July 6, 2017

Abstract: A toxicity data set of 58 phenols to *Tetrahymena pyriformis* expressed as pEC_{50} (Log to base 10 of effective concentration, EC_{50}) was taken from literature and the molecular structure of each molecule was optimized to obtain their minimum energy geometry. The descriptors of each optimized molecule were computed and subsequently used to build QSTR models. The best QSTR model hinted that the toxicity of phenol was dominantly influenced by the descriptors; molecular complexity (FMF), valence path cluster (VPC) and topological diameter (topo). The results of the statistical analysis of the tri-parametric model include; n = 41, Lack of fit (LOF) score = 0.06566, R^2 = 0.7629, $R^2_{adj.}$= 0.7437, Q^2_{LOO} = 0.7, F-value = 39.69. The generated QSTR model has been proven to possess statistical significance, high predictive power and wide applicability domain. Thus, it is recommended for environmental risk assessment of phenols.

Keywords: QSTR, Phenols, EC_{50}, Model, XlogP, LOF

1. Introduction

Phenols are important group of chemicals that has found wide applications in medicine, industry and agriculture [1]. Because of their extensive usage, they are constantly being released into the environments constituting a serious threat to humans and the ecosystem owing to their high toxicity and persistence in the environment [2-4]. The high toxicity of phenol made it one of the first compounds inscribed into The List of Priority Pollutants by the US Environmental Protection Agency [5]. Acute poison with nitro-phenol is characterized by burning pain in mouth and throat, white necrotic lesions in mouth, esophagus and stomach, vomiting, headache, irregular pulse, decrease of temperature and muscle weakness, depression, convulsions and death [6].

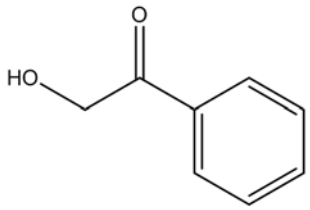
The large quantity of these chemicals been released into the environment stemming from anthropogenic and natural sources, and the high toxicity associated with them has made proper evaluation of their risk a sine qua non (very necessary). QSTR modelling has unquantifiable role to play in this regard as it provides theoretical predictive methods that can fill gaps in the data and identify those compounds that are most promising for empirical assessment since it is not practically and economically feasible to conduct toxicity tests on all phenols released into the environment owing to the unavailability of property/activity data of several classes of this compound [7-10].

Jionghao and Zhang [11] studied the acute toxicity of 31 halogenated phenols using semi empirical PM3 and coding of halogens' substitutional positions. Peng and Liu [12] investigated quantum chemical descriptors based QSAR models on toxicity of 22 halogenated phenols with Hartee-Fock and DFT methods with the basis set 6-31G (d, p) and 6-311G (d, p) respectively. In the above-mentioned models, which have high correlation coefficient, external validation of the developed models was not performed making it hard to know the external predictive power.

More recently, Guangyu *et al*. [13] investigated the link between structures of halogenated phenols and their acute toxicity in their work titled: A QSTR Study of the Acute Toxicity of Halogenated Phenols in which DFT-B3LYP method, with the basis set 6-31G (d, p), was employed to calculate some quantum chemical descriptors of 43 halogenated phenols compounds. The above descriptors along with the octanol-water partition coefficient were used to establish a QSTR of the toxicity of these compounds to *tetrahymena pyriformis* by multiple linear regression (MLR) and support vector machine (SVM). The statistical results indicate that the multiple correlation coefficient (R^2) and cross validation using leave-one-out were 0.922, 0.892 and 0.944, 0.924, respectively. To validate the predictive power of the resulting models, external validation multiple correlation coefficient and cross validation (Q^2 ext) were 0.975, 0.919 and 0.957, 0.934, respectively. The results revealed that there are good correlations among the acute toxicity of halogenated phenols to *Tetrahymena pyriformis* and the octanol-water partition coefficient, highest occupied molecular orbital, dipole moment, the sum of halogenated electric charges. Though their result showed that the QSAR models have both favorable estimation stability and good prediction power but the models could only be applied to halogenated phenols only. Thus, it can be inferred that the QSTR models have very narrow applicability domain.

2. Materials and Methods

A toxicity data set of 58 phenols to *Tetrahymena pyriformis* expressed as pEC_{50} (Log to base 10 of EC_{50}) was taken from literature [14]. The entire data set and their respective pEC_{50} are presented in Table 1. 70% (41 phenols) of the data was used as training set while 30% (17 phenols) was used as test set. Spartan' 14 software was used to optimize the geometries of all the molecules using DFT method at the B3LYP/6-31G level. Padel descriptor tool kit was used to compute all the descriptors of the molecules. QSTR models were generated using Genetic Function Approximation (GFA) – Multilinear Regression techniques in Material Studio software with the pEC_{50} as dependent variable and the computed descriptors as independent variables. Many QSTR models were generated but Model 1 was selected based on the model with the best statistical significance (LOF score = 0.06566, $R^2 = 0.7629$, $R^2_{adj.} = 0.7437$, $Q^2_{LOO} = 0.7$, F-value = 39.69]. The chemical structure of a representative of the data set studied is shown in Figure 1.



2-Hydroxyacetophenone

Figure 1. *Chemical structure of member of the data set.*

Table 1. *Name and Experimental pEC_{50} of the Phenols Studied.*

Cpd.	Name	pEC_{50}	Cpd.	Name	pEC_{50}
1	4-Fluorophenol	0.02	30	2-Bromo-4-methylphenol	0.60
2	2-Cyanophenol	0.03	31	2,4-Difluorophenol	0.61
3	5-F-2-OH-acetophenone	0.04	32	3-Isopropylphenol	0.61
4	2,4-Dimethylphenol	0.07	33	2-Cl-4,5-dimethylphenol	0.69
5	2-Hydroxyacetophenone	0.08	34	4-Butoxyphenol	0.70
6	2,5-Dimethylphenol	0.08	35	4-Chloro-2-methylphenol	0.70
7	3,5-Dimethylphenol	0.11	36	3-tert-Butylphenol	0.73
8	4'-Hydroxypropiophenone	0.12	37	4-Chloro-3-methylphenol	0.80
9	2,3-Dimethylphenol	0.12	38	4-Iodophenol	0.85
10	3,4-Dimethylphenol	0.12	39	2,2'-Biphenol	0.88
11	2-Ethylphenol	0.16	40	4-tert-Butylphenol	0.91
12	2-Chlorophenol	0.18	41	3,4,5-Trimethylphenol	0.93
13	4-OH-2-CH_3-acetophenone	0.19	42	4-sec-Butylphenol	0.98
14	4-Ethylphenol	0.2	43	2,4-Dichlorophenol	1.04
15	3-Ethylphenol	0.23	44	4-Chloro-3-ethylphenol	1.08
16	2,3,6-Trimethylphenol	0.28	45	2-Phenylphenol	1.09
17	2,4,6-Trimethylphenol	0.28	46	3-Chloro-4-fluorophenol	1.13
18	2-OH-5-CH_3-acetophenone	0.31	47	6-tert-Butyl-2,4-dimethylphenol	1.16
19	2-Bromophenol	0.33	48	4-Chloro-3,5-dirnethylphenol	1.20
20	5-Br-2-OH-benzyl alcohol	0.34	49	4-Cyclohexylphenol	1.56
21	2,3,5-Trimethylphenol	0.36	50	3,4-Dinitrophenol	0.27
22	2-Chloro-5-methylphenol	0.39	51	2,6-Dinitrophenol	0.54
23	4-Allyl-2-methoxyphenol	0.42	52	2,6-Dichloro-4-nitrophenol	0.63
24	2-Hydroxybenzaldehyde	0.42	53	2,5-Dinitrophenol	0.95
25	2,6-Difluorophenol	0.47	54	4-Bromo-2-fluoro-6-nitrophenol	1.62
26	4-Cyanophenol	0.52	55	2-Amino-4-nitrophenol	0.47
27	4-Propyloxyphenol	0.52	56	2,6-Diiodo-4-nitrophenol	1.71
28	4-Chlorophenol	0.55	57	3-Fluoro-4-nitrophenol	0.94
29	5-Methyl-2-nitrophenol	0.59	58	4-Hexyloxyphenol	1.64

Literature source for the toxicity data: Jiang *et al*., 2011

3. Results and Discussion

Model *1:

n = 41, LOF score = 0.06566, R^2 = 0.7629, $R^2_{adj.}$= 0.7437, Q^2_{LOO} = 0.7, F-value = 39.69

The definition of the descriptors in the models include;

FMF = complexity of molecule, VPC = valence path cluster, topo = topological diameter, In the equation, *n* is the number of compounds, R^2 is the multiple correlation coefficient, R^2_{adj} is adjusted R^2, F stands for significance of regression and Q^2_{LOO} stands for Leave one out cross-validation coefficient.

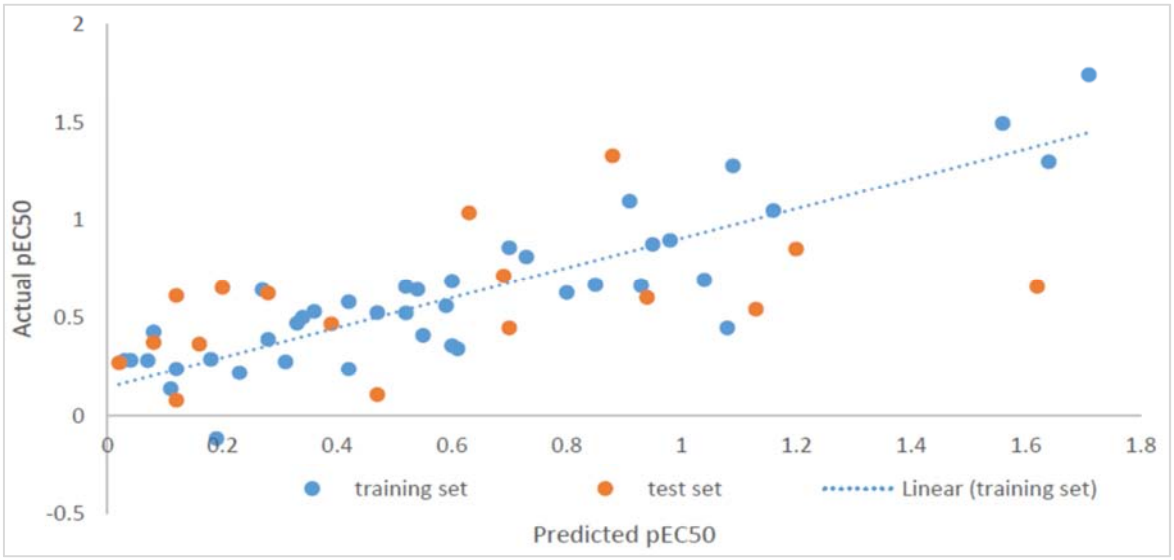


Figure 2. *Comparison Between the Predicted and Experimental Values of pEC_{50}.*

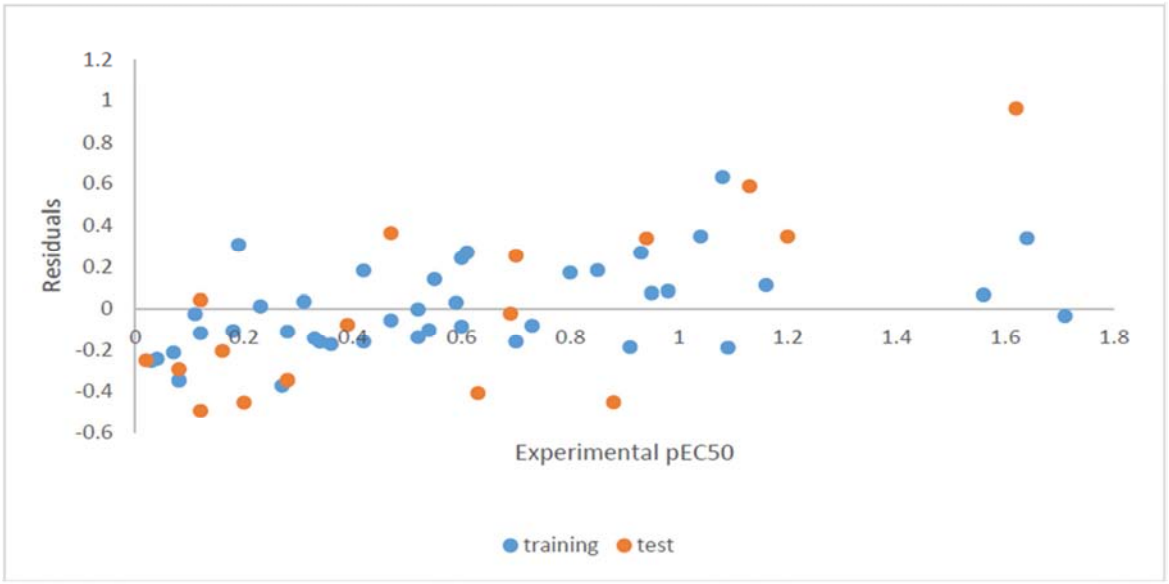


Figure 3. *Plot of the Residuals Versus the Experiment pEC_{50} Values of Model 1.*

The high R^2 of Model 1 is an indication that the model explained a very high percentage of the variation of the response variables (descriptor), high enough for a robust QSTR model. The high R^2_{adj} value and its closeness in value to the value of R^2 implies that the model has excellent explanatory power to the descriptors in it. Also, the high Q^2 value of the model shows that it is not over-fitted. F value judges the overall significance of the regression coefficients. The high F value of the model is an indication that the regression coefficients are significant.

A good predictive ability of the model 1 for the training and test set compounds is depicted by the high linearity of Figure 2 which gives the plot of predicted values of the test and training sets and their experimental values. Also, Figure 3 gives the residual plot of the optimum model. Most of the calculated residuals are distributed on both sides of the zero line, a conclusion may be drawn that there is no systematic error in the development of the present model.

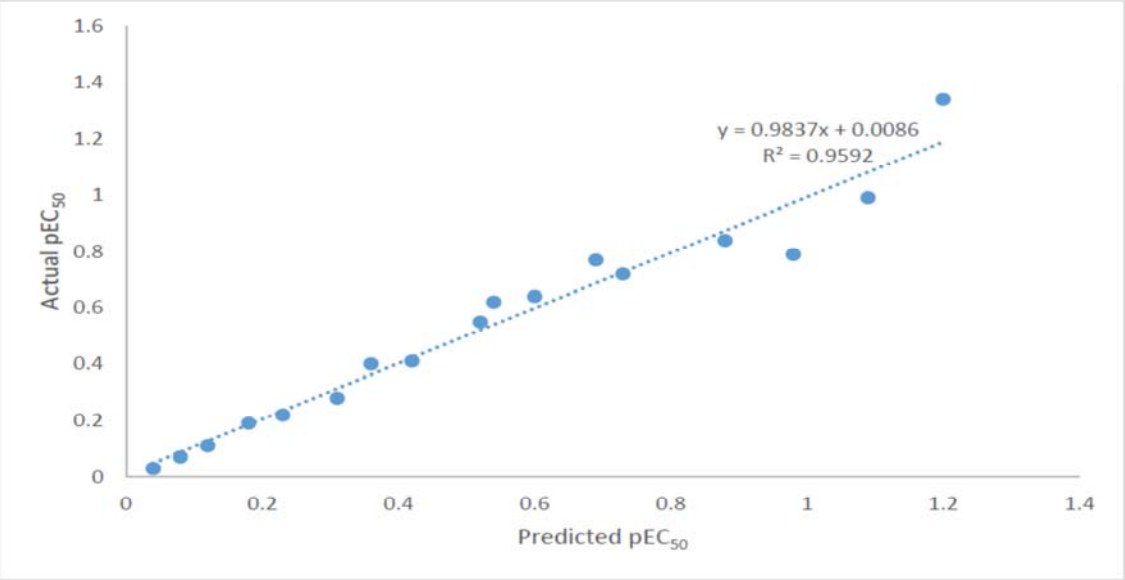


Figure 4. *Plot of Actual Verses Predicted pEC_{50} of Test Set Molecules.*

Table 2. *External Validation Table.*

Test set Compound	Actual pEC_{50}	Predicted pEC_{50}	Residual
3	0.04	0.03	0.01
6	0.08	0.07	0.01
9	0.12	0.11	0.01
12	0.18	0.19	-0.01
15	0.23	0.22	0.01
18	0.31	0.28	0.03
21	0.36	0.4	-0.04
24	0.42	0.41	0.01
27	0.52	0.55	-0.03
30	0.6	0.64	-0.04
33	0.69	0.77	-0.08
36	0.73	0.72	0.01
39	0.88	0.84	0.04
42	0.98	0.79	0.19
45	1.09	0.99	0.1
48	1.2	1.34	-0.14
51	0.54	0.62	-0.08

The comparison of the predicted pEC_{50} of the test set compounds with their experimental values by the optimum QSTR model is presented in Table 2. The low residual values shown in the table confirms the high predictive power of the model. To further augment the claim of high agreement revealed by Table 2, the experimental pEC_{50} of test set molecules were plotted against their predicted pEC_{50} (Figure 4). The high Linearity of these plot indicates reveal an all-encompassing agreement between the experimental and predicted values indicative of the high predictive power of the model.

Based on the information provided by model 1, the main factors that could impact the biological toxicity of phenols include FMF (complexity of molecule), VPC (valence path cluster, describing molecular connectivity), and topo (topological diameter). According to statistic learning theory, comparing the importance of each parameter entails the knowledge of the standardize coefficient of them in the regression equation, the bigger the absolute value of the standardized coefficient, the greater the influence of the parameter. In the equation, the standardized coefficient of FMF, VPC, and topo parameters are 2.41, 0.95, and 0.27, respectively. Thus, it could be inferred that the dominant descriptor of phenol toxicity is FMF descriptor. The three parameters describe the size of the molecules. Their positive coefficient implies that the EC_{50} of the studied molecules increase with the value of these descriptors in the molecule.

4. Conclusion

According to the QSTR study, pEC_{50} of halogenated phenols to *Tetrahymena pyriformis* increases with the descriptors; molecular complexity, valence path cluster and molecular topological diameter, descriptors of molecular size. Validation of the optimum model shows that it has good stability and great predictive power and as such could be of immense help in cheaper and safer quantitative risk assessment of these chemicals in our environment.

Recommendation

In view of the robustness and stability of the generated Genetic Function Approximation (GFA) derived model, it is suggested that environmental regulatory agencies locally and internationally should adopt it in the environmental risk assessment of phenols. This will undoubtedly reduce the level of pollution of our ecosystem by these toxic chemicals.

References

[1] Liu, H., Tan J., Yu, H., Liu, X., Wang, H. X., Wang, Z. Y. (2010). Determination of the Apparent Reaction Rate Constants for Ozone Degradation of Substituted Phenols and QSPR/QSAR Analysis. *Int. J. Environ. Res*., 4 (3): 507-512.

[2] Bukowska, B., Kowalska, S. (2004) Phenol and catechol induce prehemolytic and hemolytic changes in human erythrocytes. *Toxicol. Lett*., 152: 73–84.

[3] Cunningham, W. P., Cunningham, M. A., Saigo, B. (2005) *A Global Concern*, Mc Graw-Hill Education. New York: Environmental Science.

[4] Kušić, H. (2009). Prediction of rate constants for radical degradation of aromatic pollutants in water matrix: A QSAR study. *Chemosphere*, 75; 1128–1134.

[5] Michałowicz, J., Duda, W. (2007). Phenols – Sources and Toxicity. *Polish J. of Environ. Stud*. 16 (3): 347-362.

[6] Gosselin, R. E., Smith, R. P., Hodge, H. C. (1984). *Clinical Toxicology of Commercial Products*, 5th ed.; Williams and Wilkings: Baltimore, pp. III -192-346.

[7] Basak, S. C., Grunwald, G. D., Gute, B. D., Balasubramanian, K., Opit, D. (2000). Use of statistical and neural net approaches in predicting toxicity of chemicals, *J. Chem. Inf. Comput. Sci*. 40: 885–890.

[8] Auer, C. M., Nabholz, J. V., Baetcke, K. P. (1990). 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.,* 87: 183–197.

[9] Walker, J. D. (2003). Applications of QSAR in toxicology: a US Government perspective, *J. Mol. Struct. (Theochem).,* 622: 167–184.

[10] Luís, C., Gonçalo, B., Carrera, V. S. M., Aires-de-Sousa, J., Martin, I. L., Frade, R., Afonso, C. A. (2016). *Physico-Chemical Properties of Task-Specific Ionic Liquids*. Retrieved January 09, 2016, from http://cdn.intechopen.com/pdfs/13913.pdf.

[11] Jionghao HX, Zhang YT. A QSAR Study of Halogen Phenols Toxicity to the *Tetrahymena Pyriformis*, *Computers and Applied Chemistry*, 2007; 24: 87-90.

[12] Peng YF, Liu TB. (2009). QSAR Study of Halogen Phenols Toxicity to *Tetrahymena Pyriformis*, Chinese Journal of Structure Chemistry, 28: 218-222.

[13] He G, Feng L, Chen H. A QSAR Study of the Acute Toxicity of Halogenated Phenols. *Procedia Engineering* 2012; 43, 204 – 20.

[14] Jiang D. X., Li Y., Li J., Wang G. X. (2011). Prediction of the Aquatic Toxicity of Phenols to Tetrahymena Pyriformis from Molecular Descriptors. *International Journal of Environmental Research*, 5 (4), 923-938.