

Quantitative Structure–Toxicity Prediction of log(1/EC_{50}) for Some Benzene Derivatives from Their Density Functional Theory Calculated Molecular Descriptors

Mina Haghdadi*[1] and Mohammad H. Fatemi[2]

[1]Department of Chemistry, Islamic Azad University, Babol Branch, Iran

[2]Department of Chemistry, Mazandaran University, Babolsar, Iran

Received November 24, 2009; E-mail: mhaghdadi2@yahoo.co.uk

In this work the 50% effective inhibition concentration after 48 h (log(1/EC_{50})) of 39 substituted benzenes to the algae *Scenedesmus obliquus* was predicted by quantitative structure–toxicity relationship (QSTR) approaches. Descriptors which appeared in the best model are; highest occupied molecular orbital, molecular weight, lowest unoccupied molecular orbital, chemical potential, zero point energy, and hardness parameters. These descriptors were obtained from density functional theory (DFT) calculation at B3LYP level together with 6-31G* basis set. Results of this modeling provide the statistics of $R = 0.91$, $F = 19$, and $SE = 0.30$ for training set and $R = 0.91$, $F = 18$, and $SE = 0.34$ for test set. Also the average error and average absolute error in the prediction of log(1/EC_{50}) are -0.00041 and 0.198 for training set and for prediction set are -0.056 and 0.217, respectively. Moreover, the credibility of the model was tested by cross-validation and Y-scrambling, the results of these tests indicated the reliability of the model. The results of this investigation show that it was possible to predict the molecular toxicity of organic compounds from their DFT-calculated molecular descriptors.

One of the current interests in medicinal chemistry, environmental science, and toxicology is the ranking of chemical substances with respect to their potential hazardous effects on humans, wildlife, aquatic flora, and fauna.[1,2] Quantitative structure–activity relationships (QSARs) and quantitative structure–toxicity relationships (QSTRs) have provided a valuable approach in research into the toxicity of organic chemicals in such studies, especially in the case that there are not any experimental data about the chemicals of interest.[3] As well known, QSAR/QSTR studies could also help to understand the molecular mechanism of biological activities of environmentally important molecules including medicines and pesticides. It is essential for this purpose to obtain reliable models with high-quality descriptors, because the success of a QSAR/QSTR model is highly dependent upon the choice of descriptors. So exploring the use of reliable descriptors, especially descriptors based on reasonably high level electronic structure calculations, could lead to significantly better QSAR/QSTR results.[4] The earliest QSAR/QSTR studies were limited to the use of physicochemical descriptors or electronic structural descriptors based on semiempirical molecular orbital (MO) calculations which have been used to obtain electronic descriptors for many years.[5,6] The latest development of computer technology, software and electronic structural theory allows the calculation of quantum chemical descriptors at first-principles levels, such as density functional theory (DFT) calculations, with higher accuracy including some effective consideration of electron correlation effects.[7,8] There have been seen dramatic changes in computational chemistry over the last two decades. Therefore electronic structure theory has become an effective and powerful tool for use in predicting the properties of a wide range of molecular characteristics including geometries, energies, reactivates, and spectroscopic properties.[7,9] The combination of relatively low computational cost with reasonably accurate methods has led to the successful application of the DFT calculated molecular descriptors in the prediction of a broad range of properties of molecules.[4] So, DFT has emerged as a practical and versatile tool to obtain accurate information on molecular systems of chemical interest.[10–19] This approach, which includes the dynamic correlation effects, represents a valid alternative to the Hartree–Fock (HF) theory, or to more sophistical post-HF methods, such as those of the Møller–Plesset theory, coupled-cluster theory, or configuration interaction approach, that are highly demanding in terms of CPU time. One of the most popularly used hybrid density functionals is Becke's three-parameter hybrid exchange functional and the Lee–Yang–Parr correlation functions (B3LYP).[13,16] DFT has become a powerful computational approach to investigate the precise electronic characteristics of molecular structure, which is the key factor to determine the interactions between receptors and the ligands.[20] There are some reports about QSAR/QSTR modeling by using quantum chemical derived molecular descriptors. For example Lu et al. reported a QSTR study on the toxicity of substituted benzenes to S*cenedesmus obliquus*.[21] They used the energy of the lowest unoccupied molecular orbital and the hydrophobicity parameter of $\log K_{ow}$ as independent variables in their QSTR model. Their best model had the statistics of $R^2 = 0.793$, $SE = 0.316$, and $F = 71$ for training set, but they did not evaluate the prediction power of their model. In another work,

Published on the web March 26, 2010; doi:10.1246/bcsj.20090309

Table 1. Data Set and Corresponding Observed and Calculated Values of $\log(1/EC_{50})$[a)]

No.	Name	$\log(1/EC_{50})_{Exp}$	$\log(1/EC_{50})_{Cal}$	Residuals
1	4-Nitrotoluene	3.74	3.57	−0.17
2	1,2-Dinitrobenzene	5.04	4.45	−0.59
3	1,3-Dinitrobenzene	4.85	4.71	−0.14
4	1,4-Dinitrobenzene	4.96	4.94	−0.02
5	2,4-Dinitrotoluene	4.52	4.56	0.04
6	2,6-Dinitrotoluene	4.06	4.40	0.34
7	2-Chloronitrobenzene	3.94	3.98	0.04
8	3-Chloronitrobenzene	3.95	4.14 p	0.19
9	4-Chloronitrobenzene	4.01	4.14	0.13
10	Nitrobenzene	3.26	3.62	0.36
11	3,4-Dichloronitrobenzene	4.52	4.56	0.04
12	2,5-Dichloronitrobenzene	4.31	4.30 p	−0.01
13	2-Nitroanisole	3.44	3.49	0.05
14	3-Nitroanisole	3.71	3.65	−0.06
15	4-Nitroanisole	3.65	3.57	−0.08
16	3-Bromonitrobenzene	4.32	4.41	0.09
17	4-Bromonitrobenzene	3.88	4.42	0.54
18	2-Nitroaniline	3.33	3.42	0.09
19	3-Nitroaniline	3.48	3.45	−0.03
20	4-Nitroaniline	3.40	3.35 p	−0.05
21	2,4-Dinitroaniline	4.68	4.08	−0.60
22	Aniline	2.56	2.73	0.17
23	2-Methylaniline	2.34	2.30	−0.04
24	4-Methylaniline	3.19	3.19	0.00
25	2-Chloroaniline	2.89	3.47	0.58
26	4-Chloroaniline	2.79	3.29	0.50
27	2,3-Dichloroaniline	3.98	3.32	−0.66
28	2,4-Dichloroaniline	3.74	3.68	−0.06
29	2,5-Dichloroaniline	3.82	3.28	−0.54
30	3-Bromoaniline	2.80	3.58 p	0.78
31	2,4,6-Tribromoaniline	4.37	4.33	−0.04
32	2-Nitrophenol	3.51	3.51	0.00
33	3-Nitrophenol	3.75	3.67 p	−0.08
34	4-Nitrophenol	3.57	3.77	0.20
35	2,4-Dinitrophenol	4.16	4.22	0.06
36	Phenol	2.46	2.44	−0.02
37	2,4-Dichlorophenol	3.62	3.50 p	−0.12
38	2,4,6-Trichlorophenol	3.81	3.74	−0.07
39	Pentachlorophenol	4.63	4.34 p	−0.29

a) p is referring to prediction set.

Lameira et al. reported the application of quantum chemical calculated descriptors (at the DFT/B3LYP theory level, with the 6-31G* basis set) in QSAR prediction of anti-HIV activities for 26 flavonoid compounds.[22] The correlation between biological activity and structural properties was obtained by using multiple linear regression. They demonstrated that the anti-HIV activity of compounds can be related to the molecular hydrophobicity ($C\log P$), the electronegativity, and the charges on some key atoms. Furthermore Pasha and co-workers used molecular weight, hardness, chemical potential, total energy, and electrophilicity index as independent variables in QSAR study of the toxicity of phenols.[23] They reported that the DFT calculated molecular descriptors provide models with higher predictive power than AM1, PM3, and PM5 methods, in terms of correlation coefficient and cross validation coefficient. The main aim of the present work was to investigate the prediction of toxicity of some benzene derivatives using DFT calculated molecular descriptors and QSTR approaches.

Methodology

Data Set. The data set was taken from Ref. 21 and is shown in Table 1. In this work Lu and co-workers reported the 50% effective inhibition concentration after 48 h of 39 substituted benzenes to the algae *Scenedesmus obliquus*. These values were determined according to the algae inhibition test and reported as $\log(1/EC_{50})$.[24] As can be seen in Table 1 the values of $\log(1/EC_{50})$ ranged from 1.03 to 5.12 for aniline and pentachlorophenol, respectively.

Quantum Chemical Descriptors. The structures of 39 substituted benzene compounds have been optimized using the Gaussian 2003W computational package.[25] The Becke's three parameter exact exchange functional (B3)[16] combined with

Table 2. The DFT-Calculated Values of Molecular Descriptors

No.	HOMO /eV	η /eV	LUMO /eV	μ /eV	MW /g mol^{-1}	*ZPE* /Hartrees
1	−7.646	2.435	−2.775	−5.210	137	0.130
2	−8.299	2.420	−3.456	−5.877	156	0.105
3	−8.734	2.571	−3.592	−6.163	156	0.105
4	−8.653	2.350	−3.945	−6.299	156	0.105
5	−8.408	2.503	−3.401	−5.904	182	0.133
6	−8.218	2.476	−3.265	−5.742	182	0.133
7	−7.646	2.381	−2.884	−5.265	158	0.093
8	−7.701	2.273	−3.156	−5.428	158	0.093
9	−7.809	2.354	−3.102	−5.456	158	0.093
10	−7.891	2.489	−2.912	−5.402	123	0.103
11	−7.809	2.259	−3.290	−5.549	192	0.084
12	−7.565	2.232	−3.102	−5.314	180	0.089
13	−6.938	2.217	−2.503	−4.720	153	0.135
14	−6.962	2.043	−2.876	−4.919	153	0.135
15	−7.088	2.214	−2.660	−4.874	153	0.136
16	−7.558	2.202	−3.153	−5.355	202	0.094
17	−7.661	2.269	−3.123	−5.392	202	0.094
18	−6.431	1.884	−2.662	−4.546	138	0.119
19	−6.462	1.853	−2.756	−4.609	138	0.120
20	−6.558	2.054	−2.449	−4.504	138	0.119
21	−7.275	1.993	−3.290	−5.283	169	0.122
22	−5.513	1.694	−2.125	−3.819	93	0.116
23	−5.415	2.653	−0.109	−2.762	107	0.144
24	−5.497	2.612	−0.272	−2.884	107	0.144
25	−5.797	2.625	−0.547	−3.172	128	0.107
26	−6.020	2.691	−0.638	−3.329	128	0.107
27	−6.014	2.626	−0.762	−3.388	127	0.097
28	−5.900	2.501	−0.898	−3.309	127	0.097
29	−6.068	2.598	−0.871	−3.469	127	0.097
30	−6.019	2.846	−0.326	−3.172	171	0.108
31	−6.122	2.381	−1.360	−3.741	330	0.089
32	−7.156	2.272	−2.612	−4.884	139	0.107
33	−7.156	2.095	−2.966	−5.061	139	0.107
34	−7.782	2.451	−2.880	−5.331	139	0.106
35	−7.197	1.952	−3.292	−5.244	184	0.109
36	−6.313	2.912	−0.489	−3.401	94	0.104
37	−6.585	2.721	−1.143	−3.864	163	0.085
38	−7.184	2.884	−1.415	−4.299	198	0.075
39	−7.020	2.639	−1.741	−4.380	267	0.057

gradient corrected correlation functional of Lee–Yang–Parr (LYP)[12] DFT methods by implementing the 6-31G* basis sets have been employed. The nature of optimized geometries at the B3LYP levels has been checked with frequency calculations. Total energy and corrected zero point (*ZPE*) have been calculated for all the optimized molecules. Quantum chemical descriptors (listed in Table 2) taken from DFT calculations were used to analyze variations in the biological activity of these compounds. The minimum energy conformations were used to calculate electronic descriptors such as the energy of the highest occupied molecular orbital (HOMO), the lowest unoccupied molecular orbital (LUMO), hardness (η), softness (S), electronegativity (χ), electrophilicity (ω), and dipole moment (dp). It was found that the stability of molecules is related to hardness and the electronegativity defined as the power of an atom in a molecule to attract electrons to it. The hardness and electronegativity are defined as:

$$\eta = 1/2(\partial^2 E/(\partial N^2)_{v(r)} \quad (1)$$

$$\chi = -\mu = -[\partial E/\partial N]_{v(r)} \quad (2)$$

where E and $v(r)$ are the electronic energy and external potential of an N-electron system, respectively and μ is chemical potential. Also, the operational definitions of absolute hardness and electronegativity are given as:

$$\eta = 1/2(IP - EA) \quad (3)$$

$$\chi = -\mu = -1/2(IP + EA) \quad (4)$$

IP and *EA* are the ionization potential and electron affinity, respectively. According of Koopman's theorem, the *IP* is simply the eigenvalue of the HOMO with change of sign and *EA* is the eigenvalue of the LUMO with change of sign;[26] hence eqs 3 and 4 may be written as:

$$\eta = 1/2(\varepsilon\text{LUMO} - \varepsilon\text{HOMO}) \quad (5)$$

$$\chi = -\mu = 1/2(\varepsilon\text{HOMO} + \varepsilon\text{LUMO}) \quad (6)$$

Softness is a property of molecules that measures the extent of chemical reactivity. It is the reciprocal of hardness:

$$S = 1/\eta \quad (7)$$

Parr et al.[26,27] have proposed electrophilicity index as a measure of energy lowering due to maximal electron flow between donor and acceptor. They defined electrophilicity index (ω) as follows:

$$\omega = \mu^2/2\eta \quad (8)$$

Finally, a more general, but important property of a molecular system, the molecular weight, has also been tested as a descriptor.

Statistical Analysis. In order to develop a reliable QSTR model, more relevant features (independent variables, descriptors) were selected by stepwise-selection multiple linear regression (SMR) which combined forward-addition and backward-elimination strategies.[28] In this manner independent variables were individually added to or deleted from the model at each step of regression depending on the correlation coefficient (R), Fisher statistics value (F), and standard error (SE). Variables were selected to enter or to remove until the best model was obtained. The best equation was tested for their predictive power by comparing the calculated $\log(1/EC_{50})$ with experimental data by external and internal validation tests. The SPSS 13.0 package was used for statistical analysis. In the first step the data set sorted according to their $\log(1/EC_{50})$ values and the test set was chosen from this set with desired distances from each other and remaining molecules were chosen as training set. The diversity of molecules in training and test sets was examined by diversity analysis. Then the model was constructed by using the training set, which consists of 32 molecules, and the obtained model was used to predict the $\log(1/EC_{50})$ for molecules existing in the test set. Also leave-many-out cross validation test was used as internal validation procedure to qualify the predictive ability of the model. Finally in order to investigate any chance correlation between dependent and independent variables Y-scrambling was applied on the data set.

Results and Discussion

Model Development. We established different predictive relationships capable to link the DFT calculated molecular descriptors of benzene derivatives with their toxicities against the algae (*Scenedesmus obliquus*) as $\log(1/EC_{50})$. Based on the analyzing of different statistical parameters of these models (including R, F, and SE parameters), we conclude that the best QSTR model is one that has five molecular descriptors, which are HOMO, molecular weight (MW), LUMO, chemical potential (μ), zero point energy (ZPE), and hardness parameters (η). The appearances of these descriptors in the model can represent different aspects of molecule which was affected on molecular activity. For example ZPE can represent the stability of molecule and correlate with its size and bulkiness of molecule which can affects on the ADEME (absorption, distribution, metabolism, and excretion) properties of molecules. Other descriptors are also affected on ADEME properties of molecule by steric, lipophilic and dispersion interactions. The specifications of this model are shown in following equation:

$$\begin{aligned}\log(1/EC_{50}) &= -12.226(\pm 3.744) - 2.062(\pm 0.427)\mathrm{HOMO} \\ &- 217.766(\pm 48.293)(\mathrm{MW})^{-1} + 0.215(\pm 0.069)(\mathrm{LUMO})^{-1} \\ &- 36.565(\pm 8.756)(\mu)^{-1} + 0.048(\pm 0.019)(ZPE)^{-1} \\ &- 2.336(\pm 0.598)\eta \qquad (9)\\ n = 32, \quad R = 0.906, \quad SE = 0.30, \quad F = 19\end{aligned}$$

The DFT calculated values of these descriptors are shown in Table 2. The calculated values of $\log(1/EC_{50})$ using this model for training and test sets are shown in Table 1. Results of this model provide the statistics of $R = 0.91$, $F = 19$, and $SE = 0.30$ for training set and $R = 0.91$, $F = 18$, and $SE = 0.34$ for test set. Also the average error and average absolute error in prediction of $\log(1/EC_{50})$ are -0.00041 and 0.198 for training set and are -0.056 and 0.217 for prediction set, respectively. The calculated values of $\log(1/EC_{50})$ and their residuals were plotted against their experimental values in Figures 1 and 2, respectively. The random propagation of residuals on both sides of zero line indicated that there are not any systematic errors in the developed models.

Diversity Analysis. One of the most critical aspects of constructing the training set is to warrant enough molecular diversity for it. In this study, diversity analysis was performed on the data set to make sure the structures of the training or test sets can represent those of the whole sets.[29] A MATLAB program was written to combine a maximum dissimilarity search algorithm and general multidimensional measurements of chemical similarity based on different molecular descriptors which were calculated by DFT methods. We considered a database of n compounds generated from m highly correlated chemical descriptors $\{X_j\}_{j=1}^{m}$. Each compound X_i is represented as the following vector:

$$X_i = (x_{i1}, x_{i2}, x_{i3}, \ldots, x_{im}) \quad \text{for} \quad i = 1, 2, \ldots, n \qquad (10)$$

where x_{ij} denotes the value of descriptor j of compound X_i. The collective database $X = \{X_i\}_{i=1}^{n}$ is represented as an $n \times m$ matrix of X:

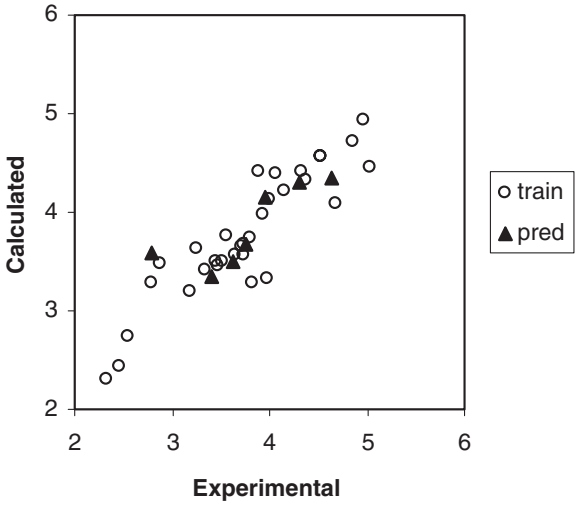

Figure 1. Plot of predicted $\log(1/EC_{50})$ versus their experimental values.

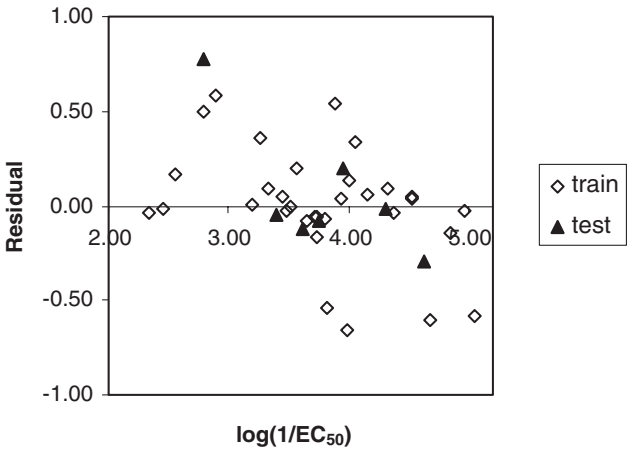

Figure 2. The residuals of predicted $\log(1/EC_{50})$ versus their experimental values.

$$X = (X_1, X_2, \ldots, X_n)^T = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \qquad (11)$$

Here the superscript T denotes the vector/matrix transpose. A distance score for two different compounds X_i and X_j can be measured by the Euclidean distance norm based on the compound descriptors:

$$d_{ij} = \|X_i - X_j\| = \sqrt{\sum_{k=1}^{m}(x_{ik} - x_{jk})^2} \qquad (12)$$

The mean distances of one sample to the remaining ones were computed as follow:

$$\overline{d_i} = \frac{\sum_{j=1}^{n} d_{ij}}{n-1} \quad i = 1, 2, \ldots, n \qquad (13)$$

Then the mean distances were normalized within the interval of zero to one. The mean distances of samples were plotted versus $\log(1/EC_{50})$ in Figure 3. As can be seen from this figure, the

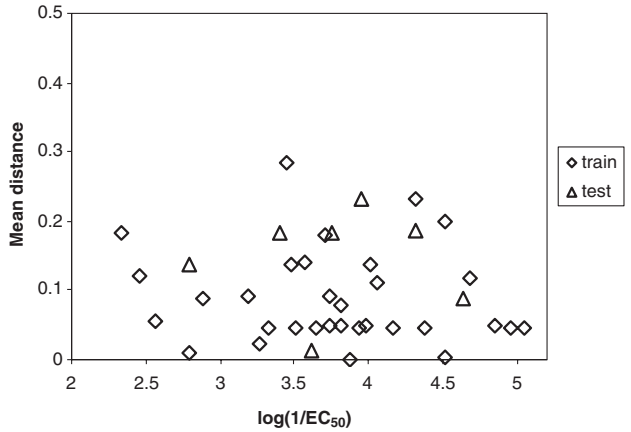

Figure 3. Diversity plot of date set.

dispersion of compounds is diverse in both training and test sets. The training set with a broad representation of the chemistry space was adequate to ensure model stability and the diversity of a test set can prove the predictive capability of the model.

Validation of the Models. In order to ensure that the results were not conditioned by the data distribution in descriptor space, the built model was also validated by evaluating the model's external predictive power on the selected prediction set by leave-many-out (LMO) scheme.[30] In this method a model was built with *N*–*m* compounds that *m* represents a group of randomly selected data points. Each *m* group was left out of the model derivation and predicted in turn. The outcomes from this procedure were a cross-validated correlation coefficient (Q^2) and standardized predicted error sum of squares (*SPRESS*), which were calculated according to the following formulas:

$$Q^2 = 1 - \frac{\sum (y_O - y_i)^2}{\sum (y_i - y_m)^2} \tag{14}$$

$$SPRESS = \sqrt{\frac{\sum (y_O - y_i)^2}{n - k - 1}} \tag{15}$$

In the above expressions, y_m is the mean of dependent variable (experimental values), n is the number of observations, and k is the number of independent variables in the regression equation. Q^2 is the proportion of variability in a data set that is accounted by a statistical model and *SPRESS* is criteria of deviation from observed data. The obtained statistical results of leave-five-out cross validation test on the obtained model were: $Q^2 = 0.781$ and $SPRESS = 0.352$, which revealed the reliability of the obtained model.

Another widely used approach to establish the model robustness is the so called Y-randomization test (randomization of response, i.e., in our case, retention).[31] It consists of repeating the calculation procedure with randomized retention vector and subsequent probability assessment of the resultant statistics. It is expected that models obtained for the data set with randomized retention should have low values of R^2. However, sometimes models based on the randomized data have high R^2 values due to chance correlation or structural redundancy.[32] The results of 30 repetitions in randomization of Y vectors (retention factors) on MLR model are shown in Table 3. As can be seen from this table the random models have significantly lower R^2 values ($\overline{R^2} = 0.129$) than the original model, which indicates good results in our original model are not due to the chance or structural dependency in the training set. As mentioned earlier Lu and co-workers proposed a QSTR model on the same data but they did not report any results about the prediction power of their model. In order to compare the results of these two models we performed leave-one-out cross-validation test to these two models, that provided the statistics of $Q^2 = 0.7154$ and $SPRESS = 0.87$ for Lu model and $Q^2 = 0.741$ and $SPRESS = 0.34$ for our model, which revealed the superiority of our model. In addition they used an experimentally determined parameter ($\log K_{ow}$) in their proposed model as independent parameter while it was preferred to use theoretically derived molecular descriptors in QSAR/QSTR models.[21] In contrast to their work the present model uses only theoretical derived molecular descriptors and the validation of model was evaluated by internal and external validation procedures as well as by Y-scrambling methods.

Table 3. R^2 Values after Several Y-Randomization Tests

No.	R^2	No.	R^2
1	0.200	16	0.124
2	0.197	17	0.081
3	0.133	18	0.088
4	0.081	19	0.076
5	0.063	20	0.051
6	0.032	21	0.074
7	0.030	22	0.009
8	0.093	23	0.115
9	0.089	24	0.028
10	0.070	25	0.024
11	0.049	26	0.030
12	0.038	27	0.086
13	0.146	28	0.060
14	0.085	29	0.025
15	0.044	30	0.047

Conclusion

The results of this investigation revealed that the DFT-calculated molecular descriptors can property relate to the toxicity of benzene derivatives. The obtained model involves various quantum chemical descriptors including MW, HOMO, LUMO, hardness, dipole moment, and zero-point energy. The reliability of the obtained model has been tested by internal and external validation methods as well as by Y-scrambling.

References

1 T. W. Schultz, L. B. Kier, L. H. Hall, *Bull. Environ. Contam. Toxicol.* **1982**, *28*, 373.

2 B. D. Gute, S. C. Basak, *SAR QSAR Environ. Res.* **1997**, *7*, 117.

3 M. H. Fatemi, S. Gharaghani, *Bioorg. Med. Chem.* **2007**, *15*, 7746.

4 M. H. Fatemi, M. Haghdadi, *J. Mol. Struct.* **2008**, *886*, 43.

5 M. H. Fatemi, E. Baher, *SAR QSAR Environ. Res.* **2009**, *20*, 77.

6 P. P. Singh, F. A. Pasha, H. K. Srivastava, *QSAR Comb. Sci.*

2003, *22*, 843.
7 H. Chermette, *J. Comput. Chem.* **1999**, *20*, 129.
8 L. Zhang, J. Wan, G. Yang, *Bioorg. Med. Chem.* **2004**, *12*, 6183.
9 P. Geerlings, F. De Proft, W. Langenaeker, *Chem. Rev.* **2003**, *103*, 1793.
10 R. G. Parr, W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, Oxford, **1989**.
11 A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098.
12 A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 1372.
13 A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.
14 J. B. Foresman, A. Frisch, *Exploring Chemistry with Electronic Structure Methods*, Gaussian Inc., Pittsburgh, PA, **1996**.
15 F. Bernardi, A. Bottoni, M. Garavelli, *Quant. Struct.-Act. Relat.* **2002**, *21*, 128.
16 C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.
17 M. Sulpizi, G. Folkers, U. Rothlisberger, P. Carloni, L. Scapozza, *Quant. Struct.-Act. Relat.* **2002**, *21*, 173.
18 J. P. Perdew, Y. Wang, *Phys. Rev. B* **1992**, *45*, 13244.
19 S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*, 1200.
20 S. Arulmozhiraja, M. Morita, *Chem. Res. Toxicol.* **2004**, *17*, 348.
21 G.-H. Lu, X. Yuan, Y.-H. Zhao, *Chemosphere* **2001**, *44*, 437.
22 J. Lameira, C. N. Alves, V. Moliner, E. Silla, *Eur. J. Med. Chem.* **2006**, *41*, 616.
23 F. A. Pasha, H. K. Srivastava, P. P. Singh, *Bioorg. Med. Chem.* **2005**, *13*, 6823.
24 OECD Test Guideline 201, Paris Decision of the Council C, **1981**, Vol. 81, pp. 320–323.
25 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Startmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rubuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Oritz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Comperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, J. A. Pople, *Gaussian 98, Revision A.6*, Gaussian. Inc., Pittsburgh, PA, **2003**.
26 R. G. Parr, R. A. Donnelly, M. Levy, W. E. Palke, *J. Chem. Phys.* **1978**, *68*, 3801.
27 R. G. Parr, L. V. Szentpály, S. Liu, *J. Am. Chem. Soc.* **1999**, *121*, 1922.
28 D. L. Massart, *Handbook of Chemometrics and Qualimetrics, Part A*, Elsevier Science, Netherland, **1997**.
29 A. G. Maldonado, J. P. Doucet, M. Petitjean, B.-T. Fan, *Mol. Diversity* **2006**, *10*, 39.
30 B. Efron, *J. Am. Stat. Assoc.* **1983**, *78*, 316.
31 S. Wold, L. Eriksson, *Statistical Validation of QSAR Results*, in *Chemometric Methods in Molecular Design*, ed. by H. van de. Waterbeemd, VCH, Weinheim, **1995**.
32 B. Walczak, D. L. Massart, *Chemom. Intell. Lab. Syst.* **1995**, *27*, 41.