Comparative QSTR Study Using Semi-Empirical and First Principle Methods Based Descriptors for Acute Toxicity of Diverse Organic Compounds to the Fathead Minnow

Erol Eroglu [1,*], Selami Palaz [1], Oral Oltulu [1], Hasan Turkmen [2] and Cihat Ozaydın [1]

1 Department of Physics, Harran University, Osmanbey Campus, 63300, Sanliurfa, Turkey
E-mail: eeroglu@harran.edu.tr, spalaz@harran.edu.tr, oltulu@harran.edu.tr, cozaydin@harran.edu.tr.

2 Department of Chemistry, Harran University, Osmanbey Campus, 63300, Sanliurfa, Turkey
E-mail: hturkmen@harran.edu.tr

* Author to whom correspondence should be addressed; E-mail: eeroglu@harran.edu.tr, Tel:+90 414 3440089, Fax:+90 414 3440051. Postal address: Department of Physics, Harran University, Osmanbey Campus, 63300, Sanliurfa, Turkey.

Abstract: Several quantum-mechanics-based descriptors were derived for a diverse set of 48 organic compounds using AM1, PM3, HF/6-31+G, and DFT-B3LYP/6-31+G (d) level of the theory. LC_{50} values of acute toxicity of the compounds were correlated to the fathead minnow and predicted using calculated descriptors by employing Comprehensive Descriptors for Structural and Statistical Analysis (CODESSA) program. The heuristic method, implemented in the CODESSA program for selecting the 'best' regression model, was applied to a pre-selection of the most-representative descriptors by sequentially eliminating descriptors that did not satisfy a certain level of statistical criterion. First model, statistically, the most significant one has been drawn up with the help of DFT calculations in which the squared correlation coefficient R^2 is 0.85, and the squared cross-validation correlation coefficient R^2_{CV} is 0.79. Second model, which has been drawn up with the help of HF calculations, has its statistical quality very close to the DFT-based one and in this model value of R^2 is 0.84 and that of R^2_{CV} is 0.78. Third and fourth models have been drawn up with the help of AM1 and PM3 calculations, respectively. The values of R^2 and R^2_{CV} in the third case are correspondingly 0.79 and 0.66, whereas in the fourth case they are 0.78 and 0.65 respectively. Results of this study clearly demonstrate that for the calculations of

descriptors in modeling of acute toxicity of organic compounds to the fathead minnow, first principal methods are much more useful than semi-empirical methods.

Keywords: Comparative QSTR; fathead minnow; acute toxicity; DFT; HF; AM1; PM3.

1. Introduction

Many of QSAR studies are based on the assumption that molecules from the same chemical domain will behave in a similar manner, so that QSAR models drawn up with the analogical molecules are hypothesized to exhibit better performance than that derived from miscellaneous data set. The traditional approach to QSARs for acute toxicity of organic compounds to the fathead minnow is the modeling of the activity of homologous or congeneric series of chemicals such as nitroaromatics [1], alkylamines [2], halogenated hydrocarbons and phenols [3], and chlorobenzenes and chloroalinines [4]. This congeneric series approach is conservative. Often, such chemicals have a single functional group or toxicophore and an alkyl moiety of variable size. Some other studies [5-11] by using diverse molecule sets have usually relied on dividing a molecule set into subgroups (chemical classes) by clustering the molecules based on their mode of action. Then, local QSTRs built up for each subgroup are applicable only to certain mode of action. It is worthy mention here that there has been a successful effort to draw up a global QSTR model by using a single descriptor, namely, logarithm of 1-octanol/water partition coefficient LogP. This model is applicable to quite miscellaneous data set, but still counts quite a big number of molecules as outliers [12]. Hydrophobicity of a molecule is characterized by LogP which is directly related to bio-uptake of chemicals by fish or many other organisms. It has been successfully used for the modeling of acute toxicity of chemicals with different modes and mechanism of toxic action to *Pimephales promelas*, combined with additional parameters such as energy of lowest unoccupied molecular orbital (E_{LUMO}) [13] and maximum superdelocalizability (S_{max}), which is a molecular orbital parameter that quantifies the electro (nucleo) philicity of a molecule [14]. Developing a better QSTR for the modeling of acute toxicity of diverse chemicals is a subject of interest due to its demand by the organizations such as OECD (Organization for Economic Co-operation and Development) or EC (European Communities) to use the QSTR model for regulatory purpose.

The aim of the present study is two folds. The first one is to build QSTR multiple regression model using quantum-mechanics-based molecular descriptors that correlate and predict the LogLC$_{50}$ value of acute toxicity of 48 compounds to the fathead minnow. LC$_{50}$ (mg/l), aquatic toxicity on *Pimephales promelas* expressed as the chemical concentration at which 50% lethality is observed in a test batch of fish within a 96 h exposure period. Molecules used in this study are quite a diverse set and were taken from a study [12]. However, they were not strictly selected to ensure that they are sufficiently diverse. The second aim of this study is to compare the accuracy of semi-empirical and first principle methods for calculation of molecular descriptors. AM1 [15] and PM3 [16, 17] are fast in computation, well suited to organic compounds, and belong to semi-empirical method family. These methods have been traditionally used to calculate the optimized 3D geometry and quantum mechanics descriptors of

molecules in most of QSAR studies. Some previous comparative QSAR works [1,18-22] have shown that using descriptors calculated by HF [23-25] or DFT [26] together with B3LYP [27] hybrid function instead of semi-empirical AM1 or PM3 methods improve the accuracy of the results that lead more reliable QSARs. On the other hand, there is an interesting comparative QSTR study relevant in this area [28]. In that study, a huge molecule set (568 molecules) has been used to establish QSTR models. These QSTR models have been built up from descriptors which were calculated using two different theory levels namely AM1 and DFT/B3LYP (6-31G**). Their study has shown that the choice of the precise but time-consuming DFT/B3LYP method does not have an advantage over AM1 method for the quality of the derived QSTRs.

Table 1. 48 compounds used in this study and their LogP and toxicity values to fathead minnow (*Pimephales promelas*).

Comp. No	CAS No	Chemical Name	[a]LogP	[a]LogLC$_{50}$ (mol/l)
1	57-55-6	1,2-Propanediol	-0.78	-0.838
2	68-12-2	Formamide, N,N-dimethyl-	-0.93	-0.839
3	71-36-3	1-Butanol	0.84	-1.601
4	78-87-5	Propane, 1,2-dichloro-	2.25	-2.907
5	78-92-2	2-Butanol	0.77	-1.305
6	79-00-5	Ethane, 1,1,2-trichloro-	2.01	-3.214
7	79-34-5	Ethane, 1,1,2,2-tetrachloro-	2.19	-3.917
8	80-05-7	Phenol, 4,4'-(1-methylethylidene)bis-	3.64	-4.696
9	80-62-6	2-Propenoic acid, 2-methyl-, methyl ester	1.28	-2.552
10	95-50-1	Benzene, 1,2-dichloro-	3.28	-3.411
11	96-18-4	Propane, 1,2,3-trichloro-	2.5	-3.346
12	96-29-7	2-Butanone, oxime	1.69	-2.014
13	100-37-8	Ethanol, 2-(diethylamino)-	0.05	-1.818
14	106-46-7	Benzene, 1,4-dichloro-	3.28	-4.015
15	107-06-2	Ethane, 1,2-dichloro-	1.83	-2.931
16	107-41-5	2,4-Pentanediol, 2-methyl-	0.58	-1.089
17	107-98-2	2-Propanol, 1-methoxy-	-0.49	-0.637
18	108-88-3	Benzene, methyl-	2.54	-3.549
19	120-83-2	Phenol, 2,4-dichloro-	2.8	-4.277
20	122-99-6	Ethanol, 2-phenoxy-	1.1	-2.604
21	123-54-6	2,4-Pentanedione	0.05	-2.860
22	123-86-4	Acetic acid, butyl ester	1.85	-3.810
23	124-04-9	Hexanedioic-acid-	0.23	-3.178
24	141-78-6	Acetic-acid-ethyl-ester-	0.86	-2.583
25	760-23-6	1-Butene, 3,4-dichloro-	2.6	-4.184
26	770-35-4	2-Propanol, 1-phenoxy-	1.52	-2.735
27	868-77-9	2-Propenoic acid, 2-methyl-, 2-hydroxyethyl ester	0.3	-2.758

28	1634-04-4	Propane, 2-methoxy-2-methyl-	1.43	-2.118
29	4169-04-4	1-Propanol, 2-phenoxy-	1.52	-2.735
30	101-84-8	Diphenyl ether	4.21	-4.62
31	693-65-2	Dipentyl ether	4.04	-4.69
32	108-20-3	Diisopropyl ether	1.52	-3.04
33	109-99-9	Tetrahydrofuran	0.46	-1.52
34	142-96-1	Dibutyl ether	3.21	-3.60
35	110-00-9	Furan	1.34	-3.04
36	64-17-5	Ethanol	-0.31	0.51
37	5673-07-4	2,6-dimethoxytoluene	2.64	-3.87
38	115-20-8	2,2,2-trichloroethanol	1.42	-2.69
39	120-82-1	1,2,4-trichlorobenzene	4.05	-4.79
40	541-73-1	1,3-dichlorobenzene	3.52	-4.27
41	150-78-7	1,4-dimethoxybenzene	2.15	-3.07
42	4412-91-3	3-furanmethanol	0.30	-2.28
43	95-75-0	3,4-dichlorotoluene	4.06	-4.74
44	67-64-1	Acetone	-0.24	-0.85
45	98-86-2	Acetophenone	1.58	-2.87
46	67-56-1	Methanol	-0.77	-0.06
47	108-94-1	Cyclohexanone	0.81	-2.27
48	79-01-6	Trichloroethene	2.42	-3.47

[a] LogP and toxicity data (LogLC$_{50}$) taken from the literature [12].

2. Procedures and Calculations Methods

2.1. Computational details

For all molecules studied here, 3-D modeling and calculations were performed using the Gaussian 03 quantum chemistry package [29]. To save in computational time, initial geometry optimizations were carried out with the molecular mechanics (MM) method using Amber force field. The lowest energy conformations of the molecules obtained by the MM method were further optimized by the DFT method by employing Becke's three-parameter hybrid functional (B3LYP) and the 6-31+G (d) basis set; their fundamental vibrations were also calculated using the same method to check if there were true minima. All the computations were carried out for the ground states of these molecules as singlet state. The lowest energy conformations of the compounds obtained using DFT were used as an input geometry for the calculations for HF/6-31+G, AM1 and PM3 methods. (CODESSA PRO) Comprehensive Descriptors for Structural and Statistical Analysis, Version 2.7.2 [30], was used for extracting descriptors of quantum mechanics and 3D geometry of the compounds from Gaussian 03 output files. CODESSA PRO enables the generation of hundreds of molecular descriptors (constitutional, topological, and quantum mechanical) from a loaded 3D geometry, and uses diverse statistical structure property/activity correlation techniques for the analysis of experimental data in combination with calculated molecular descriptors.

Table 2. DFT/B3LYP-based descriptors and predicted toxicity of the compounds by Eq 7.

Comp. No	S_{tr}	I_A	ω_H	ω_L	O-LogLC$_{50}$	P-LogLC$_{50}$	[b]Residual
1	38.902	0.263	3774.1	114.565	-0.838	-0.788	0.049
2	38.782	0.296	3171.2	106.871	-0.839	-1.184	-0.345
3	38.823	0.623	3755.3	110.929	-1.601	-1.423	0.177
4	40.055	0.224	3186.1	111.732	-2.907	-3.215	-0.308
5	38.823	0.256	3742.2	102.240	-1.305	-1.394	-0.089
6	40.544	0.114	3189.9	107.486	-3.214	-3.386	-0.172
7	41.227	0.057	3167.6	73.008	-3.917	-3.826	0.090
8	42.176	0.029	3753.8	36.841	-4.696	-4.399	0.296
9	39.720	0.116	3242.6	51.096	-2.552	-2.496	0.055
10	40.845	0.062	3224.5	135.387	-3.411	-4.086	-0.675
11	40.845	0.071	3188.4	80.855	-3.346	-3.727	-0.381
12	39.305	0.125	3762.3	65.570	-2.014	-1.984	0.029
13	40.189	0.067	3759.7	25.457	-1.818	-1.764	0.053
14	40.845	0.188	3229.6	101.569	-4.015	-4.039	-0.024
15	39.657	0.968	3196.7	118.164	-2.931	-2.813	0.117
16	40.214	0.103	3739.4	46.314	-1.089	-2.042	-0.953
17	39.406	0.244	3754.1	84.761	-0.637	-1.175	-0.538
18	39.472	0.184	3204.8	11.093	-3.549	-2.855	0.693
19	41.155	0.070	3674.9	144.475	-4.277	-3.677	0.599
20	40.680	0.150	3772.8	49.425	-2.604	-2.497	0.107
21	39.720	0.137	3163.5	44.404	-2.860	-2.051	0.809
22	40.163	0.179	3175.7	35.795	-3.810	-3.023	0.786
23[a]	40.847	0.159	3680.2	-46.051[a]	-3.178	-2.194	0.983
24	39.339	0.279	3175.9	38.760	-2.583	-2.148	0.434
25	40.359	0.084	3242.2	80.663	-4.184	-3.443	0.740
26	40.968	0.107	3755.0	47.032	-2.735	-2.847	-0.112
27	40.502	0.071	3782.8	44.232	-2.758	-2.050	0.707
28	39.340	0.145	3111.9	61.752	-2.118	-2.470	-0.352
29	40.968	0.101	3772.7	38.393	-2.735	-2.820	-0.085
30[a]	41.301	0.080	3217.2	18.694	-4.620	-4.587	0.032
31	41.085	0.269	3099.9	28.820	-4.690	-4.511	0.178
32	38.570	0.314	3306.7	608.797	-3.040	-2.520	0.519
33	38.741	0.236	3126.8	47.891	-1.520	-1.697	-0.177
34	40.503	0.304	3102.1	40.606	-3.600	-3.850	-0.250
35	39.780	0.115	3136.5	85.249	-3.040	-2.767	0.272
36	37.406	1.144	3744.7	264.649	0.510	-0.345	-0.855
37	40.968	0.073	3232.8	63.657	-3.870	-3.794	0.075
38	40.885	0.062	3767.0	-61.022[a]	-2.690	-2.609	0.080

39	41.469	0.060	3238.4	95.552	-4.790	-4.697	0.092
40	40.845	0.093	3238.5	167.828	-4.270	-4.214	0.055
41	40.680	0.150	3222.7	60.164	-3.070	-3.432	-0.362
42	39.659	0.236	3739.2	76.473	-2.280	-1.648	0.631
43	41.119	0.060	3217.4	5.630	-4.740	-4.405	0.334
44	38.096	0.336	3161.1	-57.828[a]	-0.850	-0.875	-0.025
45	40.263	0.122	3222.8	61.969	-2.870	-2.961	-0.091
46	36.324	4.240	3763.7	323.726	-0.060	0.403	0.463
47	39.721	0.142	3738.8	160.354	-2.270	-2.174	0.095
48	40.498	0.127	3251.8	172.065	-3.470	-3.556	-0.086

O-LogLC$_{50}$, observed toxicity (mol/l) taken from Ref. 12.

P-LogLC$_{50}$, predicted toxicity (mol/l) by Eq. 7.

S_{tr}, translational entropy (at 300 K).

I_A, principal moment of inertia A .

ω_H, is the highest vibrational wavenumber and ω_L, is the lowest vibrational wavenumber.

[a]Data points not included in the deriving equation 7.

[b]Residual is the differences between O-LogLC$_{50}$ and P-LogLC$_{50}$ values.

A QSAR/QSTR model can be developed for a given set of molecules by using a various types of descriptors. Sometimes, a model might have very good statistical parameters, but still not suffice to explore the mechanism of interaction between the ligand and receptor mechanistically. Building a model with physically interpretable descriptors is an important task for value of a QSAR/QSTR work. In this study, we aimed to draw up a QSTR model by using quantum mechanically calculated thermodynamical descriptors by virtue of which obtained models are usually mechanistically interpretable. About 50 thermodynamical descriptors depending on the number of atoms in a molecule were calculated using CODESSA PRO and Gaussian 03 packages. The heuristic method [29] implemented in CODESSA PRO was used to build up a multi-able regression model. By this method, a pre-selection of descriptors is accomplished. All descriptors are checked to ensure the following: (a) the values of each descriptor are available for each structure, and (b) there is a variation in these values. The descriptors for which values are not available for every structure in the data in question are discarded. Descriptors having a constant value for all structures in the data set are also discarded. A printout showing the values of descriptors discarded in this manner is provided. Thereafter, the one-parameter correlation equations for each descriptor are calculated. To further reduce the number in the "starting set" of descriptors, the following criteria are applied and a descriptor is eliminated if any of the following conditions are met with: (a) the F-test's value for the one-parameter correlation with the descriptor is below 1.0, (b) the squared correlation coefficient of the one-parameter equation is less than R^2_{min} by default 0.01, (c) the parameter's t-value is less than t_1 (where R^2_{min} 0.1 by default and t_1 1.5 by default are user defined values), (d) the descriptor is highly inter-correlated (above r_{full}, where r_{full} is a user specified value by default 0.99), with another descriptor. All the remaining descriptors are then listed in decreasing order according to the correlation coefficient of the corresponding one-parameter correlation equation. All two-parameter regression models with remaining descriptors are

developed and ranked by the regression correlation coefficient R^2. A stepwise addition of the further descriptors' scales is performed to find the best multi-parameter regression models with the optimum values of statistical criteria (highest values of R^2, the cross-validated, R^2_{CV} and the F value). R^2_{CV}, the 'leave one out' (LOO) squared cross-validated coefficient, is a practical and reliable method for testing the predictive performance and stability of a regression model. LOO approach consists in developing a number of models with one sample omitted at a time. After developing each model, the omitted data are predicted and the differences between the experimental and predicted activity values are calculated. Then the R^2_{CV} is calculated according to the following formula [31]:

$$R^2_{CV} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2} \tag{1}$$

where y_i is the actual experimental activity, $\bar{y}$ is the average actual experimental activity and $\hat{y}_i$ is the predicted activity of compound i computed by the new regression equation obtained each time after leaving out one datum point (No. i).

Table 3. HF-based descriptors and predicted toxicity of the compounds by Eq 8.

Comp. No	S_{tr}	I_A	ω_H	ω_L	O-LogLC$_{50}$	P-LogLC$_{50}$	[b]Residual
1	38.862	0.229	4036.5	119.378	-0.838	-0.930	-0.092
2	38.782	0.298	3335.0	116.641	-0.839	-1.828	-0.989
3	38.823	0.625	4030.4	110.442	-1.601	-1.021	0.579
4	40.055	0.223	3408.6	108.918	-2.907	-3.128	-0.221
5	38.823	0.258	4019.2	109.284	-1.305	-0.898	0.406
6	40.544	0.110	3412.8	102.260	-3.214	-3.622	-0.408
7	41.227	0.055	3392.0	76.661	-3.917	-4.350	-0.433
8	42.176	0.029	4039.8	39.720	-4.696	-4.441	0.254
9	39.720	0.171	3442.8	73.673	-2.552	-2.602	-0.050
10	40.845	0.062	3409.3	152.827	-3.411	-4.063	-0.652
11	40.845	0.068	3407.3	78.833	-3.346	-3.904	-0.558
12	39.305	0.128	4134.1	79.521	-2.014	-1.172	0.841
13	40.189	0.067	4031.9	30.933	-1.818	-2.188	-0.370
14	40.845	0.191	3415.0	112.760	-4.015	-4.014	0.000
15	39.657	0.965	3417.5	114.598	-2.931	-2.951	-0.020
16	40.214	0.104	4013.3	49.077	-1.089	-2.296	-1.207
17	39.406	0.247	4029.9	86.093	-0.637	-1.490	-0.853
18[a]	39.472	0.186	3385.2	-39.526[a]	-3.549	-2.154	1.394
19	41.155	0.071	4039.2	137.736	-4.277	-3.515	0.762
20	40.680	0.154	4044.1	46.404	-2.604	-2.796	-0.192
21	39.720	0.140	3317.3	40.210	-2.860	-2.690	0.169

22	40.163	0.181	3331.4	49.892	-3.810	-3.211	0.598
23[a]	40.847	0.160	3980.7	-60.874[a]	-3.178	-2.837	0.340
24	39.339	0.280	3331.4	64.207	-2.583	-2.343	0.239
25	40.359	0.084	3418.0	85.160	-4.184	-3.357	0.826
26	40.968	0.110	4110.3	41.233	-2.735	-3.004	-0.269
27	40.502	0.114	3975.3	53.356	-2.758	-2.689	0.068
28	39.340	0.146	3262.9	52.717	-2.118	-2.364	-0.246
29	40.968	0.103	4045.4	32.456	-2.735	-3.071	-0.336
30[a]	41.301	0.081	3400.4	-3.077[a]	-4.620	-4.254	0.365
31	41.085	0.259	3248.3	24.087	-4.690	-4.345	0.344
32	38.570	0.315	3514.0	651.847	-3.040	-2.536	0.504
33	38.741	0.234	3293.5	103.901	-1.520	-1.787	-0.267
34	40.503	0.284	3251.9	38.393	-3.600	-3.720	-0.120
35	39.780	0.117	3290.4	77.601	-3.040	-2.870	0.169
36	37.406	1.147	4021.2	266.514	0.510	0.036	-0.473
37	40.968	0.073	3407.1	59.104	-3.870	-4.002	0.132
38	40.885	0.061	4038.1	78.853	-2.690	-3.075	-0.385
39	41.469	0.061	3416.5	106.258	-4.790	-4.659	0.131
40	40.845	0.093	3422.3	187.518	-4.270	-4.134	0.135
41	40.680	0.141	3402.2	39.039	-3.070	-3.662	-0.592
42	39.659	0.236	4013.5	80.456	-2.280	-1.783	0.496
43	41.119	0.059	3401.6	36.388	-4.740	-4.125	0.614
44	38.096	0.342	3313.3	58.043	-0.850	-0.964	-0.114
45	40.263	0.124	3406.3	53.685	-2.870	-3.208	-0.338
46	36.324	4.417	4035.2	311.330	-0.060	-0.023	0.036
47	39.721	0.143	4015.3	168.667	-2.270	-2.013	0.256
48	40.498	0.123	3465.5	181.294	-3.470	-3.677	-0.207

O-LogLC$_{50}$, observed toxicity (mol/l) taken from Ref. 12.

P-LogLC$_{50}$, predicted toxicity (mol/l) by Eq. 8.

S_{tr}, translational entropy (at 300 K).

I_A, principal moment of inertia A .

ω_H, is the highest vibrational wavenumber and ω_L, is the lowest vibrational wavenumber .

[a]Data points not included in the deriving equation 8.

[b]Residual is the differences between O-LogLC$_{50}$ and P-LogLC$_{50}$ values.

2.2. Theory

Among the thermodynamical descriptors, translational entropy (at 300 K), principal moment of inertia A, highest normal mode of vibrational frequency, and lowest normal mode of vibrational frequency were involved in the models that are presented in this study. Thermodynamical properties of a molecule arise from the energetics of vibrational frequencies. This connection is based upon

partitioning the total energy of a macroscopic system among the constituent molecules. Translational entropy (at 300 K) is defined as [32];

$$S_{tr} = \ln\left(\frac{2\pi mkT}{h^2}\right)^{1/2} \frac{Ve^{5/2}}{N}$$

(2)

where V is the volume of the system, N is the Avogadro's number, h is the Planck constant, m is the molecular mass, and kT is the Boltzman temperature. Highest normal mode of vibrational frequency and lowest normal mode of vibrational frequency are actually not frequencies; they are wavenumbers (in cm^{-1} unit). It is customary to call normal modes of vibration of molecule as frequency in infrared and Raman spectroscopy. Definition of normal mode of vibration arises from quantum mechanical harmonic oscillator model of a diatomic molecule. In this model, energy of the vibrational states is given as [33],

$$E_v = hv\left(v + \frac{1}{2}\right)$$

(3)

where h is the Planck constant, v is the vibrational quantum number (0, 1, 2,), and v is the classical vibrational frequency given by,

$$v = \frac{1}{2\pi}\left(\frac{k}{\mu}\right)^{1/2}$$

(4)

where k is the force constant of the chemical bond and μ is the reduced mass for nuclei of two atoms. More commonly, equation 5 is used as vibrational wavenumber (ω) form rather than frequency form, where

$$E_v = hc\omega\left(v + \frac{1}{2}\right)$$

(5)

where ω is the vibrational wavenumber, c is the velocity of light. Final descriptor involved in our model is principal moment of inertia A (I_A) that is obtained from the 3D-cooordinate of the atoms in the given molecule. Its definition is given as [34],

$$I_A = \sum_i m_i r_{ix}^2$$

(6)

where m_i are the atomic masses and r_{ix} denotes the distance of the i-*th* atomic nucleus from the main rotational axes, x. I_A characterizes the mass distribution in the molecule.

3. Results and Discussion

3.1. Results

The assessment of toxcity of a hypothetical compound is a subject of interest. The QSAR/QSTR method saves time and cost in determining the toxicity of a series of newly synthesized compounds with the help of toxicity of previously known compounds. Forty-eight compounds have been taken in this study, and their toxicity (LogLC$_{50}$) and calculated logarithm of 1-octanol/water partition coefficient (LogP) values to fathead minnow have been taken from the literature [12] and are given in Table 1. Among the quantum mechanically calculated descriptors, translational entropy (at 300 K), principal moment of inertia A, and highest normal mode of vibrational frequency and lowest normal

mode of vibrational frequency have been identified which are capable of modeling the toxicity and the structure of a molecule. The data matrix of these descriptors obtained from first principal (HF and DFT-B3LYP) and semi-empirical (AM1 and PM3) methods calculations are shown in Table 2 to 5. By using DFT-based descriptors, several equations were generated by using all the variables and the statistically best model that we have obtained is four-parameters equation, which is as follows:

$$\text{DFT-LogLC}_{50} = 36.37\ -1.11 S_{tr}\ +1.65 \text{x} 10^{-3}\ \omega_H -2.29 \text{x} 10^{-3}\ \omega_L -0.34 I_A$$
$$N=45,\ R^2=0.85,\ R^2_{CV}=0.79,\ F=60.34,\ \text{and}\ s^2=0.27 \tag{7}$$

where N is the number of compounds included in the model, R^2 is the squared correlation coefficient, R^2_{CV} is the squared cross-validation correlation coefficient, F is the Fisher test for significance of the equation and s^2 is the standard deviation of the regression. The statistical quality of the above equation is good as evident from its correlation coefficient R^2 value = 0.85 and a cross-validation coefficient R^2_{CV} value = 0.79. The predicted toxicity of the compounds is given in Table 2 by using Equation 7. In this DFT-based model, compounds **23**, **38,** and **44** were selected as outliners due to the fact that during the calculation of their geometry and vibrational frequencies, all our attempts had failed to get all the frequencies as positive. **23, 38** and **44** have given one negative value of vibrational frequency. This means that obtained structures of these molecules do not correspond to the global minima of potential energy surface. Second, model based on other first principle method used in present study, namely, HF method is found as follows:

$$\text{HF-LogLC}_{50} = 38.00\ -1.13 S_{tr}\ +1.38 \text{x} 10^{-3}\ \omega_H -2.22 \text{x} 10^{-3}\ \omega_L -0.36 I_A$$
$$N=45,\ R^2=0.84,\ R^2_{CV}=0.78,\ F=58.73,\ \text{and}\ s^2=0.28 \tag{8}$$

Statistical quality of this model is very close to the DFT-based one. The predicted toxicity of the compounds by using Equation 8 is given in Table 3. In this HF-based model, compounds **18**, **23,** and **30** were selected as outliers due to the same reason as DFT-based model. By using AM1-based descriptors, the statistically best model that we have obtained is as follows:

$$\text{AM1-LogLC}_{50} = 43.94\ -1.14 S_{tr}\ -1.14 \text{x} 10^{-4}\ \omega_H -4.21 \text{x} 10^{-3}\ \omega_L -0.188 I_A$$
$$N=48,\ R^2=0.76,\ R^2_{CV}=0.72,\ F=34.47,\ \text{and}\ s^2=0.43 \tag{9}$$

Second model, namely, PM3 based on other semi-empirical method used in present study was found as follow:

$$\text{PM3-LogLC}_{50} = 45.04\ -1.18 S_{tr}\ -3.32 \text{x} 10^{-5}\ \omega_H -3.59 \text{x} 10^{-3}\ \omega_L -0.253 I_A$$
$$N=48,\ R^2=0.75,\ R^2_{CV}=0.66,\ F=32.58,\ \text{and}\ s^2=0.44 \tag{10}$$

Statistical fit of equation 10 is similar to equation 9. R^2_{CV} value of the equation 10 is relatively lower than that of equation 9. This result indicates that AM1 and PM3 based models are similar statistical fit, but PM3 based model has a lower predictive power as is evident from its lower value of squared cross-validated coefficient (R^2_{CV} =0.66). Finally, in order to elucidate the relationship between the hydrophobicity of compounds and their toxicity to fathead minnow, we have added LogP value of compounds as an additional descriptor to the DFT-based model (equation 7). Influence of the LogP to statistical fit of the equation 7 is as follows:

$$\text{DFT and LogP-LogLC}_{50} = 22.08 - 0.68 S_{tr} + 9.45 \text{x} 10^{-4} \omega_H - 1.22 \text{x} 10^{-3} \omega_L - 0.147 I_A - 0.373 \text{CLogP} \tag{11}$$

$$N=45, R^2=0.89, R^2_{CV}=0.80, F=70.82, \text{ and, } s^2=0.19$$

As can be seen in equation 11, statistical quality of DFT-based model was increased dramatically by adding LogP to the model.

Table 4. AM1-based descriptors and predicted toxicity of the compounds by Eq 9.

Comp. No	S_{tr}	I_A	ω_H	ω_L	O-LogLC$_{50}$	P-LogLC$_{50}$	[b]Residual
1	38.916	0.263	3502.6	46.572	-0,838	-1,353	-0.515
2	38.796	0.294	3103.6	133.306	-0,839	-1,541	-0.702
3	38.838	0.612	3157.6	64.729	-1,601	-1,366	0.234
4	40.094	0.227	3154.8	73.827	-2,907	-2,773	0.133
5	38.838	0.265	3163.5	66.472	-1,305	-1,309	-0.004
6	40.589	0.103	3079.7	224.825	-3,214	-3,945	-0.731
7	41.273	0.060	3005.6	51.915	-3,917	-3,985	-0.068
8	42.190	0.028	3461.1	26.172	-4,696	-4,974	-0.278
9	39.734	0.129	3232.5	61.837	-2,552	-2,299	0.252
10	40.878	0.064	3197.0	121.451	-3,411	-3,847	-0.436
11	40.887	0.074	3090.3	56.361	-3,346	-3,573	-0.227
12	39.319	0.144	3156.8	83.654	-2,014	-1,910	0.103
13	40.203	0.074	3158.6	46.823	-1,818	-2,755	-0.937
14	40.878	0.187	3192.5	95.185	-4,015	-3,759	0.255
15	39.699	0.983	3101.5	75.687	-2,931	-2,464	0.466
16	40.228	0.104	3161.8	17.174	-1,089	-2,665	-1.576
17	39.420	0.245	3159.1	39.893	-0,637	-1,861	-1.224
18	39.486	0.182	3202.7	196.531	-3,549	-2,590	0.958
19	41.186	0.070	3192.5	107.448	-4,277	-4,142	0.134
20	40.693	0.148	3206.1	19.546	-2,604	-3,223	-0.619
21	39.734	0.172	3163.8	27.915	-2,860	-2,157	0.703
22	40.177	0.177	3157.5	47.599	-3,810	-2,748	1.061
23	40.861	0.162	3425.5	35.172	-3,178	-3,508	-0.330
24	39.353	0.276	3162.1	47.922	-2,583	-1,824	0.759
25	40.395	0.090	3211.5	50.659	-4,184	-3,001	1.182
26	40.982	0.107	3206.1	19.068	-2,735	-3,544	-0.809
27	40.515	0.081	3199.7	41.240	-2,758	-3,096	-0.338
28	39.354	0.147	3163.9	13.189	-2,118	-1,655	0.463
29	40.982	0.091	3206.0	11.418	-2,735	-3,508	-0.773
30	41.315	0.079	3204.6	25.514	-4,620	-3,948	0.671
31	41.098	0.258	3157.4	16.950	-4,690	-3,692	0.997
32	38.584	0.306	3304.1	513.629	-3,040	-2,927	0.113
33	38.756	0.235	3122.5	42.932	-1,520	-1,105	0.414
34	40.517	0.293	3157.5	26.151	-3,600	-3,070	0.529

35	39.794	0.116	3159.9	89.923	-3,040	-2,477	0.563
36	37.421	1.124	3161.4	146.790	0,510	-0,183	-0.693
37	40.982	0.073	3205.6	43.190	-3,870	-3,639	0.230
38	40.926	0.063	3073.9	105.914	-2,690	-3,823	-1.133
39	41.505	0.061	3187.5	87.350	-4,790	-4,422	0.367
40	40.878	0.094	3195.2	162.544	-4,270	-4,026	0.243
41	40.693	0.146	3203.1	37.611	-3,070	-3,298	-0.228
42	39.673	0.238	3299.0	43.605	-2,280	-2,181	0.098
43	41.150	0.061	3190.5	97.797	-4,740	-4,058	0.681
44	38.111	0.330	3157.3	83.337	-0,850	-0,558	0.291
45	40.277	0.122	3199.0	17.486	-2,870	-2,731	0.138
46	36.339	4.051	3149.1	295.043	-0,060	-0,118	-0.058
47	39.735	0.143	3107.1	134.076	-2,270	-2,593	-0.323
48	40.544	0.131	3152.6	166.703	-3,470	-3,505	-0.035

O-LogLC$_{50}$, observed toxicity (mol/l) taken from Ref. 12.

P-LogLC$_{50}$, predicted toxicity (mol/l) by Eq. 9.

S_{tr}, translational entropy (at 300 K).

I_A, principal moment of inertia A.

ω_H, is the highest vibrational wavenumber and ω_L, is the lowest vibrational wavenumber.

[b]Residual is the differences between O-LogLC$_{50}$ and P-LogLC$_{50}$ values.

The above linear regression models obtained using the descriptors from DFT and HF calculations are much better than those obtained from semi-empirical AM1 and PM3 methods. Introducing LogP to DFT-based model, there is a rapid rising in statistical quality of the regression equation. Figure 1 gives the plots of observed LogLC$_{50}$ versus the predicted LogLC$_{50}$ by equations (7)-(11) for compounds.

3.2. Discussion

QSAR/QSTR model quality depends on the reliability of the dataset (i.e. uncertainty in toxicological and physicochemical and/or structural data). The authors usually have to rely on experimental toxicological data set that is taken from literature. The data are assumed to provide a uniform measure of toxicity for all of the compounds studied. When the uncertainties in physicochemical and/or structural data are considered, the accuracy of the descriptors is an important element for the QSAR/QSTRs. The computational level of theory is a major task for the accuracy of descriptor calculation. Above presented results clearly demonstrate the effect of used level of theory for calculations. Semi-empirical methods such as AM1 and PM3 use the empirical or experimental parameters to deal with the Schrödinger equation and omit some molecular integral calculations, so they are much faster than the first principle HF and DFT-B3LYP schemes. Therefore, they are utilized more widely in the calculations of molecular properties. But accuracy of their results is inferior to ab initio or DFT methods. The best statistical quality of the equations obtained in this study was DFT – based one with the correlation coefficient R^2 value = 0.85 and a cross-validation coefficient R^2_{CV} value

= 0.79. The second best equation was HF-based one, and in this model R^2 is 0.84 and R^2_{CV} is 0.78. This result is as expected. HF does not include the effects of an instant electronic correlation, whereas DFT-B3LYP does, so that HF is inferior to DFT in theory. The statistical quality of equations obtained from the semi-empirical methods was lower than that of equations obtained from the DFT and HF methods as expected. AM1 and PM3 based models have given similar statistical fits, but PM3 based model has a lower predictive power as evident from its lower value of R^2_{CV} =0.66 in contrast to that of AM1 value of R^2_{CV} =0.72.

For all of the above mentioned models, the most representative descriptor with the highest coefficient is the translational entropy (at 300 K) S_{tr}, which negatively correlates to LogLC$_{50}$. It should be noted that all of the thermochemical properties of a molecule such as S_{tr} arise from the energetics of vibrational frequencies. This connection is based upon partitioning of the total energy of a macroscopic system among the constituent molecules. Other two descriptors involved in the models are lowest normal mode of vibrational frequency, ω_L and highest normal mode of vibrational frequency, ω_H . ω_L correlates negatively to LogLC$_{50}$ in all the models, whereas ω_H correlates positively to LogLC$_{50}$ in DFT, HF, and DFT with LogP models. In several QSAR studies, fundamental vibrational frequencies of molecules have been used as descriptors [35-39]. They suggested that the eigen value ('EVA') descriptors are derived from fundamental IR and Raman range of molecular vibrational frequencies. Vibrational frequency of a molecule is sensitive to 3D structure. The idea behind the use of such data as descriptors was that a significant amount of information pertaining to molecular properties might be contained within the molecular vibration wave function [35]. Another descriptor involved in the models in this study is principal moment of inertia A, (I_A). It is a geometrical descriptor and originates from the rigid rotator approximation model of molecules. I_A is sensitive to 3D structure and characterizes the mass distribution in the molecule. It correlates negatively to LogLC$_{50}$ in all models.

Results of present study demonstrate that quantum mechanically calculated thermo chemical descriptors, S_{tr}, ω_H, and ω_L jointly with geometrical descriptor, I_A are capable of modeling the acute toxicity of the compounds to the fathead minnow. First principle DFT and HF methods led to statistically better models than that of semi-empirical AM1 and PM3 methods. This result is normal because DFT and HF can calculate molecular properties such as optimized geometry and spectroscopic properties more accurately than semi-empirical methods. Our results are in disagreement with the conclusions of the other comparative study [28] which has concluded that the use of DFT/B3LYP does not have an advantage over AM1 for the quality of the derived QSTRs. The disagreement between two studies may result from several reasons. It may be due to the difference of the size of molecule sets between two studies. Another possible reason is that the models were restricted to build up using only quantum chemical and thermodynamical descriptors in our study whereas Natzeva *et al.* [28] has used considerably large amount of descriptors to derive their models.

Finally, LogP has been inserted as an additional descriptor into the statistically best model (DFT-based one). This resulted in an increase of statistical quality of the model for the parameters (R^2 from 0.85 to 0.89, F from 60.34 to 70.82, and s^2 from 0.27 to 0.19). As mentioned in introduction section, LogP itself has been used as single descriptor for many QSTR models. Most of the compounds used in this study act in narcotic mode of action. Narcosis is a general term that describes noncovalent interaction between xenobiotics and cellular membranes. Whereas it is generally accepted that narcosis

is the result of the accumulation of the compounds in cell membranes that disturbs their function, the exact mechanism is not known yet [40]. LogP characterizes the hydrophobicity of a molecule that is directly related to bio-uptake of chemicals by fish or many other organisms. Results presented in this study demonstrates that quantum mechanically calculated thermo chemical descriptors in combination with LogP are capable of modeling the acute toxicity of a quite diverse set of 48 compounds to the fathead minnow.

Table 5. PM3-based descriptors and predicted toxicity of the compounds by Eq 10.

Comp. No	S_{tr}	I_A	ω_H	ω_L	O-LogLC$_{50}$	P-LogLC$_{50}$	[b]Residual
1	38.916	0.261	3182.6	67.778	-0.838	-1.372	-0.534
2	38.796	0.281	3131.1	198.042	-0.839	-1.702	-0.863
3	38.838	0.620	3182.7	82.322	-1.601	-1.423	0.177
4	40.094	0.233	3176.2	69.425	-2.907	-2.763	0.143
5	38.838	0.269	3183.6	121.401	-1.305	-1.474	-0.169
6	40.589	0.106	3051.0	209.619	-3.214	-3.815	-0.601
7	41.273	0.057	2945.1	29.974	-3.917	-3.963	-0.046
8	42.190	0.026	3162.5	41.232	-4.696	-5.086	-0.390
9	39.734	0.111	3165.8	50.229	-2.552	-2.236	0.315
10	40.878	0.066	3078.0	122.576	-3.411	-3.835	-0.424
11	40.887	0.077	3059.8	44.269	-3.346	-3.566	-0.220
12	39.319	0.136	3183.9	100.781	-2.014	-1.935	0.078
13	40.203	0.062	3179.1	49.434	-1.818	-2.776	-0.958
14	40.878	0.188	3073.9	93.850	-4.015	-3.763	0.252
15	39.699	0.963	3065.2	63.709	-2.931	-2.457	0.473
16	40.228	0.103	3182.6	56.271	-1.089	-2.841	-1.752
17	39.420	0.244	3183.2	54.483	-0.637	-1.916	-1.279
18	39.486	0.184	3171.8	191.436	-3.549	-2.470	1.078
19	41.186	0.072	3067.9	107.791	-4.277	-4.147	0.129
20	40.693	0.150	3081.5	22.099	-2.604	-3.277	-0.673
21	39.734	0.175	3183.0	39.014	-2.860	-2.213	0.646
22	40.177	0.182	3182.4	40.669	-3.810	-2.744	1.065
23	40.861	0.164	3851.1	41.276	-3.178	-3.573	-0.395
24	39.353	0.277	3185.8	39.706	-2.583	-1.792	0.791
25	40.359	0.093	3144.0	50.603	-4.184	-2.993	1.190
26	40.968	0.107	3902.2	19.928	-2.735	-3.569	-0.834
27	40.515	0.073	3164.7	31.179	-2.758	-3.082	-0.324
28	39.354	0.145	3182.3	115.054	-2.118	-2.030	0.087
29	40.982	0.091	3172.0	35.494	-2.735	-3.654	-0.919
30	41.315	0.079	3081.0	15.268	-4.620	-3.969	0.650
31	41.098	0.260	3182.5	19.088	-4.690	-3.776	0.913

32	38.584	0.311	3176.5	507.271	-3.040	-2.571	0.469
33	38.756	0.238	3075.1	54.747	-1.520	-1.126	0.393
34	40.517	0.294	3182.3	29.091	-3.600	-3.134	0.465
35	39.794	0.111	3175.0	27.889	-3.040	-2.228	0.811
36	37.421	1.162	3187.0	169.851	0.510	-0.200	-0.710
37	40.982	0.062	3166.9	52.972	-3.870	-3.709	0.160
38	40.926	0.065	2969.0	70.581	-2.690	-3.701	-1.011
39	41.505	0.063	3070.5	87.839	-4.790	-4.451	0.338
40	40.878	0.095	3076.5	157.507	-4.270	-3.968	0.301
41	40.693	0.142	3145.7	20.095	-3.070	-3.270	-0.200
42	39.673	0.240	3164.6	30.095	-2.280	-2.125	0.154
43	41.150	0.063	3171.4	97.945	-4.740	-4.070	0.669
44	38.111	0.335	3181.9	60.492	-0.850	-0.413	0.436
45	40.277	0.122	3168.8	53.232	-2.870	-2.893	-0.023
46	36.339	4.245	3141.5	283.579	-0.060	-0.110	-0.050
47	39.735	0.142	3046.7	143.340	-2.270	-2.577	-0.307
48	40.544	0.139	3065.4	147.246	-3.470	-3.503	-0.033

O-LogLC$_{50}$, observed toxicity (mol/l) taken from Ref. 12.

P-LogLC$_{50}$, predicted toxicity (mol/l) by Eq. 10.

S_{tr}, translational entropy (at 300 K).

I_A, principal moment of inertia A.

ω_H, is the highest vibrational wavenumber and ω_L, is the lowest vibrational wavenumber.

[b]Residual is the differences between O-LogLC$_{50}$ and P-LogLC$_{50}$ values.

3.3. Conclusion

This QSTR study has been based on quantum mechanically calculated descriptors (such as entropy (at 300 K), principal moment of inertia A, highest normal mode of vibrational frequency, and lowest normal mode of vibrational frequency) and on the acute toxicity of the 48 organic compounds to the fathead minnow. All of these descriptors are sensitive to 3D structure of the molecules. The reliability of this study has been tested by four different methods, namely, AM1, MP3, HF, and DFT/B3LYP. A comparison of all the methods indicates that the DFT/B3LYP method is more reliable than others and has a high predictive power. Introduction of LogP as an additional descriptor into DFT-based model has resulted in an increase of statistical parameters of the model.

(a) DFT/B3LYP

(b) HF

(c) AM1

(d) PM3
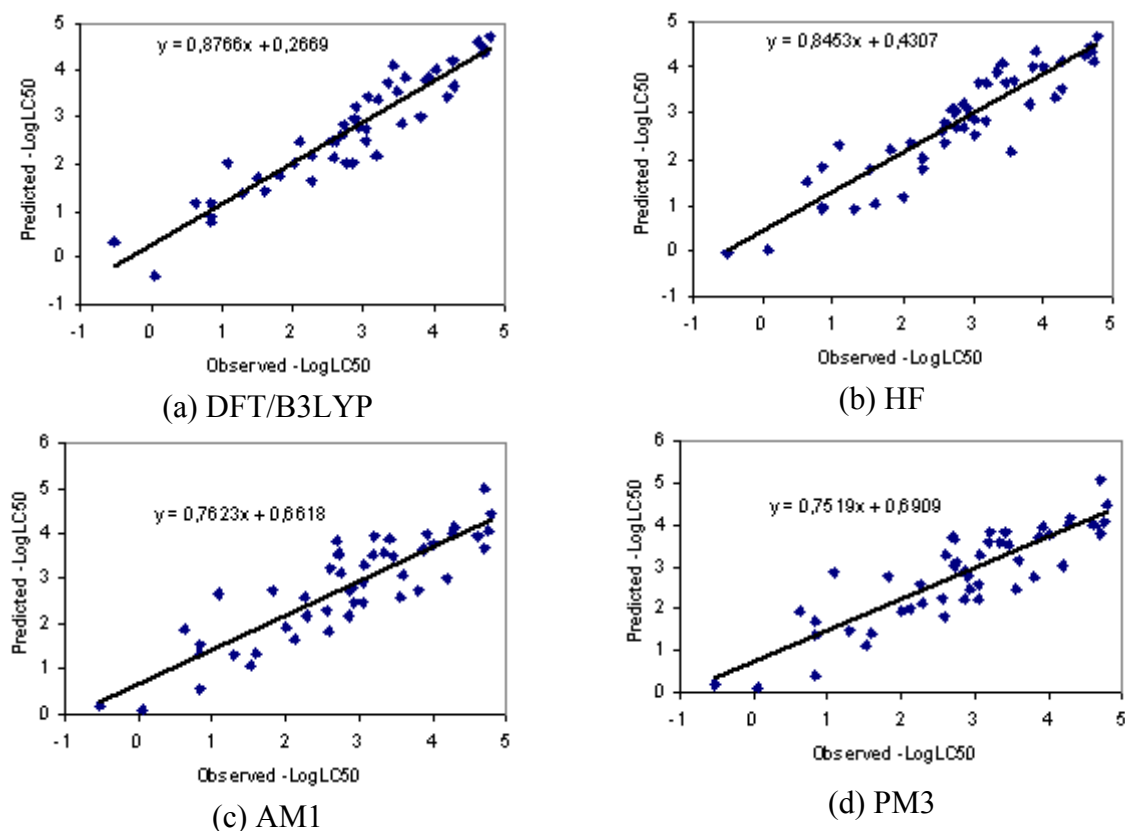
Figure 1. . The plot of observed versus predicted LogLC$_{50}$ values by using
(a) eq.7 (DFT/B3LYP), (b) eq. 8 (HF), (c) eq. 9 (AM1) and (d) eq. 10 (PM3).

References and Notes

1. Yan, X.F.; Xiao, H.M.; Gong, X.D.; Ju, X.H. A comparison of semiempirical and first principle methods for establishing toxicological QSARs of nitroaromatics. *J. Mol. Struct. (Theochem)* **2006**, *764*, 141-148.
2. Newsome, L.D.; Johnson, D.E.; Lipnick, R.L.; Broderius, S.J.; Russom, C.L. A QSAR study of the toxicity of amines to the fathead minnow. *Sci. Tot. Env.* **1991**, *109-110*, 537-551.
3. Protic, M.; Sabljic, A.; Quantitative structure-activity relationships of acute toxicity of commercial chemicals on fathead minnows: effect of molecular size. *Aquatic Toxicology* **1989,** *14,* 47-64.
4. Van Leeuwen, C.J.; Adema, D.M.M.; Hermens, J. Quantitative structure-activity relationships for fish early life stage toxicity. *Aquatic Toxicology.* **1990**, *16*, 321-334.
5. Yuan, H.; Wang, Y.Y.; Cheng, Y.Y. Local and Global Quantitative Structure-Activity Relationship Modeling and Prediction for the Baseline Toxicity. *J. Chem. Inf. Model.* **2007** *47*, 159-169.

6. Devillers, J. A new strategy for using supervised artificial neural networks in QSAR. *Sar And Qsar In Environmental Research* **2005**, *16*, 433-442.

7. He, L. ; Jurs, P.C. Assessing the reliability of a QSAR model's predictions. *J. Mol. Graph. Model.* **2005**, *23*, 503-523.

8. Papa, E**.;** Villa, F.; Gramatica, P. Statistically Validated QSARs, Based on Theoretical Descriptors, for Modeling Aquatic Toxicity of Organic Chemicals in *Pimephales promelas* (Fathead Minnow). *J. Chem. Inf. Model*. **2005**, *45*, 1256-1266.

9. McKinney, J.D.; Richard, A.; Waller, C.; Newman, M.C.; Gerberick, F. The Practice of Structure Activity Relationships (SAR) in Toxicology. *Toxicol. Sci.* **2000**, *56*, 8-17.

10. Cronin, M.T.D.; Dearden, J.C. QSAR in Toxicology. 1. Prediction of Aquatic Toxicity. *Quant. Struct.–Act. Relat*. **1995**, *14*, 1-7.

11. Hermens, J. Prediction of environmental toxicity based on structure-activity relationships using mechanistic information. *Sci. Total Environ*. **1995**, *171*, 235-242.

12. Pavan, M.; Worth, A.; Netzeva, T. Preliminary analysis of an aquatic toxicity dataset and assessment of QSAR models for narcosis. *JRC report EUR 21479 EN*. **2005,** European Commission, Joint Research Centre, Ispra, Italy.
    (http://ecb.jrc.it/Documents/QSAR/Report_QSAR_model_for_narcosis.pdf)

13. Veith, G.D. ; Mekenyan, O.G. A QSAR Approach for Estimating the Aquatic Toxicity of Soft Electrophiles QSAR for Soft Electrophiles. *Quant. Struct.-Act. Relat*. **1993**, *12*, 349-356.

14. Karabunarliev, S.; Mekenyan, O.G.; Karcher, W.; Russom, C.L.; Bradbury, S.P. Quantum-chemical Descriptors for Estimating the Acute Toxicity of Substituted Benzenes to the Guppy (*Poecilia reticulata*) and Fathead Minnow (*Pimephales promelas*). *Quant. Struct.-Act. Relat*. **1996**, *15*, 311-320.

15. Dewar, M.; Zoebisch E.G.; Healy, E. F.; Stewart, J.J.B. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1993**, *115*, 5348-5356.

16. Stewart, J.J.B. Optimization of parameters for semiempirical methods I. Method. *J. Comp. Chem.* **1989**, *10*, 209-220.

17. Stewart, J.J.B. Optimization of parameters for semiempirical methods II. Applications. *J. Comp. Chem.* **1989**, *10*, 221-264.

18. Reis, M.; Lobato, B.; Lameira, J.; Santos, A.S.; Alves, C.N. A theoretical study of phenolic compounds with antioxidant properties. *Eur. J. Med. Chem*. **2006**, *41*, 1-7.

19. Pasha, F.A.; Srivastava, H.K.; Singh, P.P. Comparative QSAR study of phenol derivatives with the help of density functional theory. *Bioorg. Med. Chem*. **2005**, *13*, 6823-6829.

20. Zhang, L.; Wan, J.; Yang, G. A DFT-based QSARs study of protoporphyrinogen oxidase inhibitors: phenyl triazolinones. *Bioorg. & Med. Chem.* **2004**, *12*, 6183-6191.

21. Trohalaki, S.; Giffort, E.; Pachter, R. Improved QSARs for predictive toxicology of halogenated hydrocarbons**.** *Computers and Chemistry* **2000**, *24*, 421-427.

22. Eroglu, E.;Turkmen, H. A DFT-based quantum theoretic QSAR study of aromatic and heterocyclic sulfonamides as carbonic anhydrase inhibitors against isozyme, CA-II. *J. Mol. Graph. Model*. **2007**, doi:10.1016/j.jmgm.2007.03.015

23. Roothan, C.C.J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys*. **1951**, *23*, 69-89.

24. Pople, J.A.; Nesbet, R.K. Self-Consistent Orbitals for Radicals. *J. Chem. Phys*. **1954**, *22*, 571-572.

25. McWeeny, R.; Dierksen, G. Self-Consistent Perturbation Theory. II. Extension to Open Shells. *J. Chem. Phys*. **1968**, *49,* 4852-4856.

26. Parr, R.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford University Pres: New York, 1989.

27. Becke, A.D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys*. **1993**, *98*, 5648-5652.

29. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Montgomery, Jr., J.A.; Vreven, T.; Kudin, K.N.; Burant, J.C.; Millam, J.M.; Iyengar, S.S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G.A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J.E.; Hratchian, H.P.; Cross, J.B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R.E.; Yazyev, O.; Austin, A.J.; Cammi, R.; Pomelli, C.; Ochterski, J.W.; Ayala, P.Y.; Morokuma, K.; Voth, G.A.; Salvador, P.; Dannenberg, J.J.; Zakrzewski, V.G.; Dapprich, S.; Daniels, A.D.; Strain, M.C.; Farkas, O.; Malick, D.K.; Rabuck, A.D.; Raghavachari, K.; Foresman, J.B.; Ortiz, J.V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B.B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R.L.; Fox, D.J.; Keith, T.; Al-Laham, M.A.; Peng, C.Y.; Nanayakkara, A.; Challacombe, M.; Gill, P.M.W.; Johnson, B.; Chen, W.; Wong, M.W.; Gonzalez, C.; and Pople, J.A.; Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford CT, 2004.

30. CODESSA (Comprehensive Descriptors for Structural and Statistical Analysis), Semichem, 7204, Mullen, Shawnee, KS 66216 USA, Copyright© Semichem and the University of Florida, 1995-2004.

31. Mayers, R.H. *Classical and Modern Regression With Applications*; PWS-KENT Publ. Co.: Boston, 1990.

32. CODESSA™, *References Manual, V. 2.13 (PC).* Semichem, 7204, Mullen, Shawnee, KS 66216 USA, Copyright© Semichem and the University of Florida, 2002.

33. Hollas, J. M. *Modern Spectroscopy*, 2nd Ed.; John Wiley & Sons Ltd.: London, UK, 1992.

34. David, R.L. *Handbook of the Chemistry and Physics*, 85nd ed., CRC Press: Cleveland OH, 2004.

35. Ferguson, A.M.; Heritage, T.W.; Jonatson, P.; Pack, S.E.; Phillips, L.; Rogan, J.; Snaith, P.J. EVA: A new theoretically based molecular descriptor for use in QSAR/QSPR analysis. *J. Comput.-Aided Mol. Des*. **1997**, *11*, 143-152.

36. Turner, D.B.; Willett, P.; Ferguson, A.M.; Heritage, T. Evaluation of a novel infrared range vibration-based descriptor (EVA) for QSAR studies. 1. General application. *J. Comput.-Aided Mol. Des*. **1997**, *11*, 409-422.

37. Turner, D.B.; Willett, P.; Ferguson, A.M.; Heritage, T. Evaluation of a novel molecular vibration-based descriptor (EVA) for QSAR studies: 2. Model validation using a benchmark steroid dataset. *J. Comput.-Aided Mol. Des*. **1999**, *13*, 271-296.

38. Turner, D.B.; Willett, P. Evaluation of the EVA descriptor for QSAR studies: 3. The use of a genetic algorithm to search for models with enhanced predictive properties (EVA_GA). *J. Comput.-Aided Mol. Des*. **2000**, *14*, 1-21.

39. Ginn, C.M.R.; Turner, D.B.; Willett, P.; Ferguson, A.M.; Heritage, T.W. Similarity Searching in Files of Three-Dimensional Chemical Structures: Evaluation of the EVA Descriptor and Combination of Rankings Using Data Fusion. *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 23-37.

40. Schultz, T.W.; Cronin, M.T.D.; Walker, J.D.; Aptula, A.O. Quantitative structure–activity relationships (QSARs) in toxicology: a historical perspective. *Journal of Molecular Structure (Theochem)* **2003**, *622*, 1-22.