Bulgarian Chemical Communications, Volume 51, Issue 3 (pp. 445 - 451) 2019 DOI: 10.34049/bcc.51.3.4849

Toxicity prediction of organophosphorus compounds by QSAR

A. Hatipoglu*, S. Aydogdu, H. Dedeoglu, D. Civan

Yıldız Technical University, Chemistry Department, 34220, İstanbul, Turkey

Received April 6, 2019; Accepted Feruary 12, 2018

Organophosphorus compounds are often used as pesticides and they are known as major environmentally hazardous chemicals. Conformational analysis and geometry optimizations of all organophosphorus compounds were carried out to determine the most stable structures. Modeling of the molecules was performed with DFT at B3LYP/6-31G* level. The solvation effects were computed using CPCM as the solvation model. A quantitative structure-activity relationship (QSAR) study was performed to correlate the toxicity of the organophosphorus compounds with calculated molecular descriptors by multilinear regression. The QSAR equations were validated internally and externally. The correlation coefficient (R^2) and cross-validation correlation coefficient (R^2_{CV}) were 0.9422 and 0.9192, respectively. These results show that the QSAR equations have both favourable estimation stability and good prediction power.

Keywords: DFT, toxicity, molecular descriptors, solvent effect.

INTRODUCTION

Organophosphorus compounds (OPs) are often used as pesticides, as insecticides and as warfare agents [1,2]. Organophosphorus compounds are known as major environmentally hazardous chemicals. It is estimated that only 5% of the consumed pesticides reach the target pest, the rest is dispersed into the environment. The increasing use over time is significantly enhancing the risks of environmental contamination of groundwater, food, plants, water resources and human beings [1,3-5]. Organophosphorus compounds are highly toxic and show their acute toxic effects by inhibiting the enzyme acethylcholinesterase (AChE). This inhibition leads to paralysis and even death. Different studies have shown that pesticide exposure is associated to adverse health effects such as depression, memory disorders, respiratory problems, dermal damage, neurological deficit, miscarriages, birth deformities and cancer [4,6,7].

There is a need for evaluating the potential hazard of these pesticides but traditional experimental investigations are often extremely time-consuming and expensive. One practical alternative would be to predict these toxic effects by using quantitative structure-activity relationships (QSARs) or quantitative structure-toxicity relationships (QSTRs) [8,9]. QSAR studies have been developed to assess the amount of toxicity for different chemicals from their molecular structure point of view. In development of QSAR, the descriptors fall into three classes: (i) physical or physicochemical properties (ii) quantum chemical descriptors and (iii) topological descriptors. These descriptors are found to be useful because they help to characterize the electronic environment of a molecule [10,11].

Quantum-chemical descriptors have become quite popular recently and are widely used due to the reliability and accuracy, as well as capability to characterize the electronic properties of the molecule. Among the variety of computational approaches, density functional theory (DFT) is one of the most commonly used. The conceptual density functional theory has been exploited in various occasions to understand the chemical reactivity and site selectivity [1,12]. Several studies have been performed for the toxicity of organophosphates and organothiophosphates against different aquatic organisms. Many models have been derived for OPs with topological indices [1,2,5,13] and some of them derived with the logarithm of the octanol-water partition coefficient or some other physical properties [9, 13-18]. The models mentioned can be used to explain the toxicity of organophosphate and organothiophosphate pesticides.

However, the toxicity studies of OPs need to be improved because these molecules are hazardous for the environment. The aim of this work was to develop QSAR models to find suitable molecular descriptors to predict the toxicity of OPs. For this purpose various molecular descriptors such as hardness (η), energy of the highest occupied molecular orbital (E_{HOMO}), dipole moment (D) and charge of phosphorus atom (P_{ch}) were calculated. The relationships between the experimental rat oral lethal dose (LD_{50}) of the molecules and the calculated descriptors were examined through multilinear regression in order to determine the best

* To whom all correspondence should be sent:
E-mail: hatiparzu@yahoo.com

descriptor of the toxicity of the molecule. In order to determine both the stability and predictive power of QSAR models, leave-one-out cross-validation and external validation were performed on the developed models.

The general chemical structure of OPs is shown in Fig. 1 where R are methyl or ethyl groups and X - leaving groups such as alkyl, heterocyclic, etc. The optimized geometries of the OP molecules in this study are given in Figs. 2 and 3.

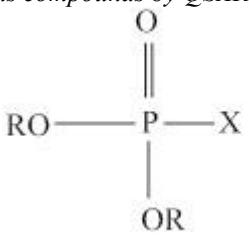

Figure 1. General structure of OPs.

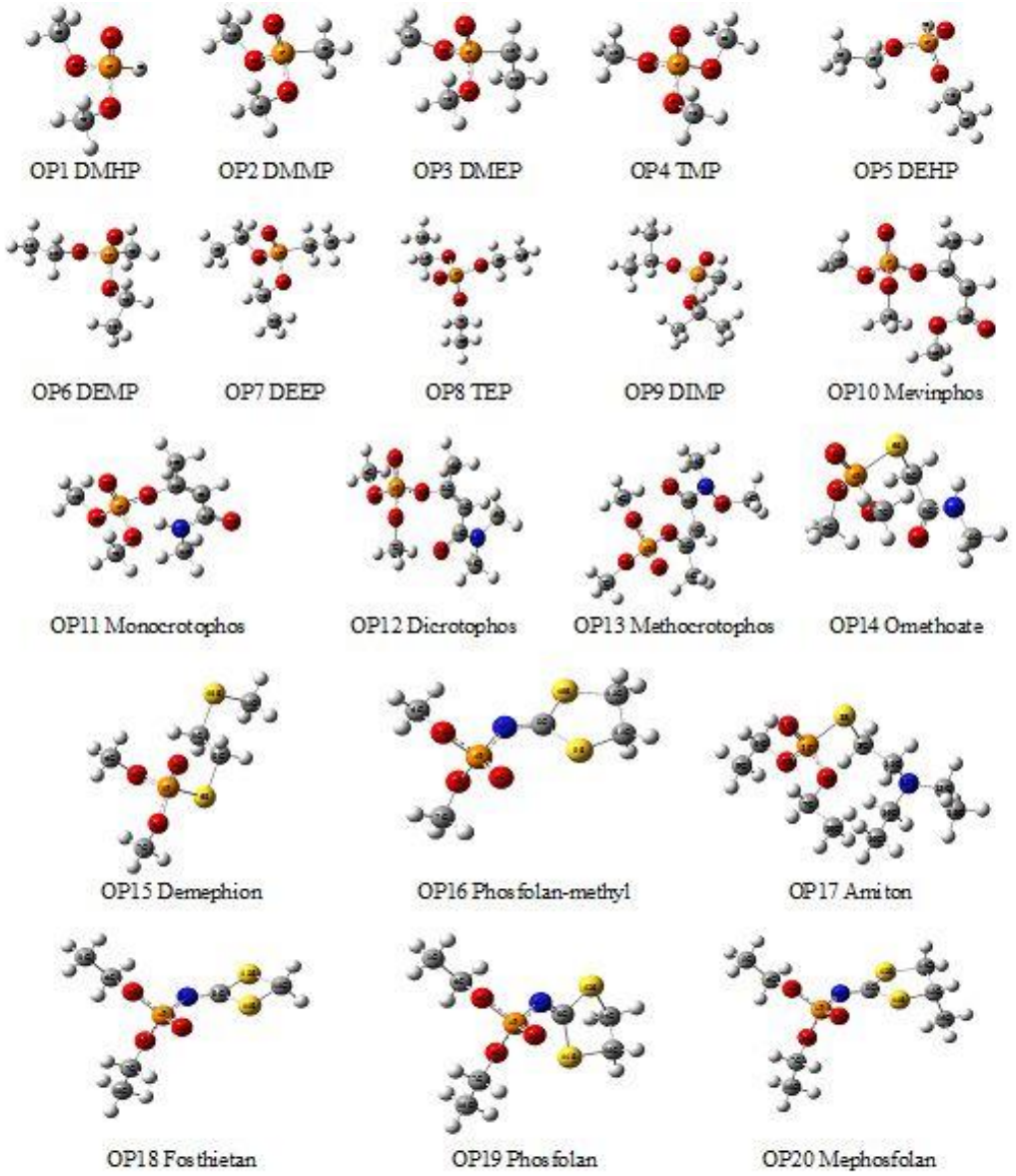

Figure 2. Optimized structure of training set (gray, carbon; red, oxygen; blue, nitrogen; orange, phosphorus; yellow, sulphur; blue, nitrogen; green, chlorine; white, hydrogen)

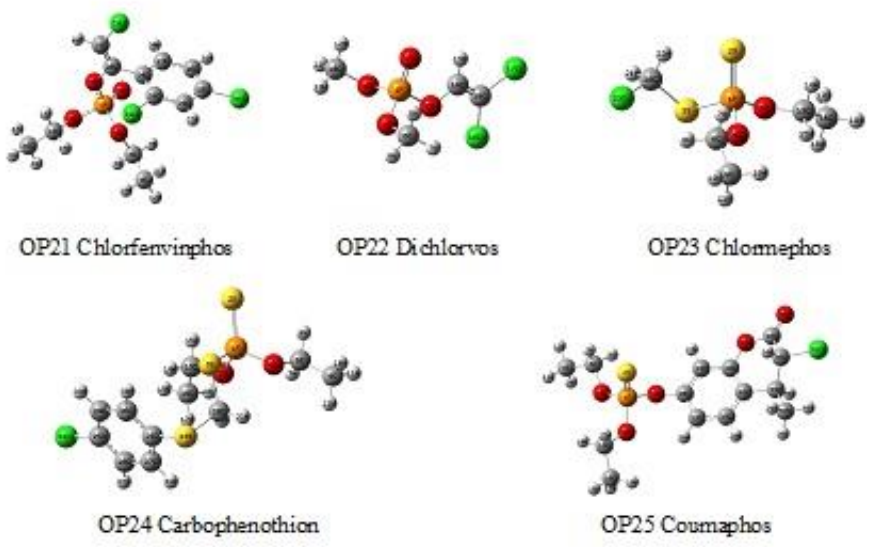

Figure 3. Optimized structure of the test set.

Table 1. Calculated molecular descriptors, η (au), E_{HOMO} (au), D (Debye) , P_{charge} and experimental LD_{50} (mg/kg) for OP compounds.

	η	**E_{HOMO}**	**D**	**P_{charge}**	**LD_{50}**
OP 1	0.1827	-0.3010	2.7848	1.0072	3283
OP 2	0.1801	-0.2963	2.6548	1.1150	8210
OP 3	0.1752	-0.2885	2.5056	1.1482	1900
OP 4	0.1753	-0.3031	0.9631	1.1951	1095
OP 5	0.1807	-0.2987	5.2988	1.0235	4350
OP 6	0.1809	-0.2936	2.8593	1.1134	800
OP 7	0.1789	-0.2899	2.7453	1.1263	2330
OP 8	0.1757	-0.2984	4.7914	1.2348	1600
OP 9	0.1735	-0.2852	2.8630	1.1304	984
OP 10	0.1082	-0.2619	3.7891	1.2534	6.3
OP 11	0.1088	-0.2512	5.9621	1.2434	18.3
OP 12	0.1063	-0.2389	6.2476	1.2504	18.8
OP 13	0.1018	-0.2449	4.6495	1.2618	2
OP 14	0.1229	-0.2539	5.3691	1.0833	22
OP 15	0.1142	-0.2346	4.9309	1.0799	20
OP 16	0.1073	-0.2582	4.8449	1.1762	51
OP 17	0.1022	-0.1999	5.4711	1.0635	3.3
OP 18	0.1047	-0.2647	2.9489	1.1851	4.7
OP 19	0.1059	-0.2585	8.2244	1.1866	8.9
OP 20	0.1074	-0.2568	4.7903	1.1761	8.9
OP21	0.1055	-0.2494	3.7317	1.2874	10
OP22	0.1209	-0.2495	3.4911	1.2067	17
OP23	0.1123	-0.2603	2.9430	0.7863	7
OP24	0.1043	-0.2464	5.9263	0.7924	6.8
OP25	0.1052	-0.2467	7.8190	1.0026	13

Table 2. Statistical parameters for model equations

No	Equations	R	R^2	R^2_{adj}	SD	F
1	$\log LD_{50}$ = - 2.70314 + 0.20416 P_{ch} + 32.53254 η	0.9696	0.9402	0.9327	0.3106	125.7278
2	$\log LD_{50}$ = - 3.13745 – 3.1370 E_{HOMO} + 31.08450 η	0.9688	0.9386	0.9314	0.3237	129.9486
3	$\log LD_{50}$ = -3.13491 + 0.06483 D + 35.14140 η	0.9707	0.9422	0.9354	0.3141	138.4956

MATERIALS AND METHODS

Computational method

Computational studies were carried out using the Gaussian 03 software suite [19]. The molecular structures of 25 OPs were optimized using Density Functional Theory (DFT) with the three-parameter Becke–Lee–Yang–Parr (B3LYP) exchange-correlation functional at 6-31G (d) level. The force constants and vibrational frequencies of the molecules were determined after optimizing their geometries to ensure that they are minimal on the potential energy surface. The molecular descriptors were calculated from the orbital energies of the optimized geometries. The conductor-like polarizable continuum model (CPCM) was used to model solvent effects. The solvent was water at 25°C, with dielectric constant ε = 78.39 [20].

Density functional theory was used extensively to calculate molecular properties of environmental organic pollutants [21]. In this study we calculated the global descriptor, namely hardness (η), for the toxicity predicting of OPs. According to the DFT, hardness, η, is defined as [22]:

$$\eta = \frac{1}{2}\left(\frac{\partial^2 E}{\partial N^2}\right)_{V(r)} \quad (1)$$

where E is the total energy of the system, N is the number of electrons in the system, and ν(r) is the external potential.

Hardness is calculated by using Koopmans' theorem [22]:

$$\eta = \frac{E_{LUMO} - E_{HOMO}}{2} \quad (2)$$

where E_{LUMO} is the energy of the lowest unoccupied molecular orbital and E_{HOMO} is the energy of the highest occupied molecular orbital.

All calculated descriptors such as hardness, energy of the highest occupied molecular orbital, dipol moment and charge of phosphorus atom, are listed in Table 1 for aqueous medium.

Data sources

The QSAR models presented in this paper were developed on 25 OP compounds. These compounds were divided into two groups: 20 OPs for the training set (OP1-OP20) given in Fig. 2 and 5 OPs for the test set (OP21-OP25) given in Fig. 3.

The toxicity of each compound was qualified in terms of LD_{50} (mg/kg). LD stands for "Lethal Dose". The value of LD_{50} for a substance is the dose requir ed to kill half the members of a tested population after a specified test duration. LD_{50} is frequently used as a general indicator for acute toxicity. The experimental rat oral LD_{50} values obtained from the literature are presented in Table 1 [23].

Statistical analysis

A QSAR model is a mathematical relationship between the chemical's quantitative molecular descriptors and its toxicological, biological, and physicochemical activities [24]. Multilinear regression is a common method used in QSAR studies. The multilinear regression is a statistical method used to find a relation between one dependent variable and several independent variables. This relation can be expressed as:

$$Y = A\,x_j + B y_j + C z_j + \ldots + D$$

where A,B,C are regression coefficients and D is the intercept obtained through regression analysis, where x_j, y_j and z_j are the quantum chemical descriptors for the molecule J as independent variables. Y, as dependent variable, represents the expected values by the regression model.

The statistical qualities of the regression equations were judged by parameters such as R^2 (correlation coefficient), R^2_{adj} (adjusted correlation coefficient), F value (Fischer statistics) and SD value (standard deviation). Testing the stability, predictive power and generalization ability of the models is a very important step in QSAR study. As for the validation of predictive power of a QSAR model, two basic principles (internal validation and external validation) are available. The cross-validation is one of the most popular methods for internal validation. In this paper, the stability and prediction ability of models were examined by using leave-one-out (LOO) cross-validation. Cross-validation provides the values of PRESS (Predictive Residual Error Sum of Squares), SSY (Sum of squares of deviation of the experimental values from their mean) and R^2_{CV} coefficient (Cross-validation correlation coefficient) which can test the predictive power of the proposed model [25].

RESULTS AND DISCUSSION

QSAR analysis

This study was carried out for some 25 OP compounds in order to determine a quantitative structure–activity relationship between the molecular descriptors and the rat oral LD_{50} toxicity values. Regression analyses were performed using the experimental toxicity log LD_{50} as the dependent variable and the DFT-based descriptors, namely hardness (η), highest occupied molecular orbital energy (E_{HOMO}), dipole moment (D) and charge of phosphorus atom (P_{ch}) as the independent variables.

Twenty training compounds (OP1-OP20) were used for all model equations.

Table 3. Cross-validation parameters for the model equations.

	PRESS	$\frac{PRESS}{SSY}$	R^2_{CV}
Equation 1	2.0466	0.0835	0.9165
Equation 2	2.4796	0.0855	0.9145
Equation 3	2.3441	0.0808	0.9192

With the intention of finding certain molecular descriptors in order to determine the toxicity of OPs, we derived model equations by forward stepwise regression analysis. Among the several models generated, the three best two-parameter models were selected and they are listed in Table 2. The selection was based on the following mentioned statistically parameters. Generally, the higher R^2 and the higher F value indicate that the model is reliable. It is commonly assumed that a robust and reliable correlation is indicated by $R^2 \geq 0.75$ and SD ≤ 0.5 [26]. As seen in Table 2, the derived model equations are statistically reliable, the correlation coeffients were found to be good (0.9386 – 0.9422) and the standard deviations were below 0.33. According to the model equations, dipole moment, charge of phosphorus atom and hardness positively affected toxicity; in contrast, the energy of the highest occupied molecular orbital had a negative correlation. In the derived model equations, the greatest value of regression coefficient was for hardness, it played a dominant role in the toxicity of OPs.

Validation of the QSAR models

In order to confirm that the models with excellent statistics have excellent predictive power too, we evaluated cross-validation parameters for model equations. Calculated cross-validation results are presented in Table 3 and results indicate that all models proposed were significant. The cross-validation was performed using the leave-one-out method (LOO) in which one compound is removed from the training set and the toxicity is correlated using the rest of the training set. Cross-validation provides the values of PRESS, SSY, PRESS/SSY and R^2_{CV} which we can use to test the prediction power of the model equation. PRESS is a good estimate of the real prediction error of the model equations. If PRESS is smaller than the SSY the model is considered to be statistically significant. In a reasonable QSAR model PRESS/SSY should be smaller than 0.4 and R^2_{CV} should be bigger than 0.5 [25]. In our results, good cross-validation R^2_{CV} was obtained for the models. As seen in Table 3, cross-validation correlation coefficient R^2_{CV} values range from 0.9145 to 0.9192. The ratio PRESS/SSY ranges between 0.0808 – 0.0855 indicating that all proposed models are reliable.

Finally, in order to confirm our findings, the toxicity of the OPs predicted log LD_{50} by model equations was compared with corresponding observed log LD_{50} values. These comparisons are shown in Table 4. As is seen in the table, the predicted toxicity values agree with the experimental ones. The residual is the difference between observed and predicted log LD_{50}. The two numeric values are close enough to each other. The plot between the predicted and the observed toxicity values of OP compounds is shown in Fig. 4. The predictive ability of the QSAR models was also evaluated by external validation. The external validation results are given in Table 4. As seen in the table, the two numeric values (observed and predicted) are close to each other. This shows that the equations have excellent determining capability of the OP compounds toxicity.

Toxicity interpretation by QSAR analysis

Based on the QSAR equations, the main descriptors that could impact the toxicity of OPs were η, E_{HOMO}, D and P_{ch}. Hardness is a measure of the stability of the molecule. According to the maximum hardness principle (MHP) molecules arrange themselves so as to be as hard as possible [22]. Therefore, stable molecules are likely to be harder than less stable molecules and thus they have low reactivities. Hardness is the most important descriptor for the toxicity of OPs, as mentioned above. The positive coefficient of hardness demonstrated that the toxicity increases with hardness. E_{HOMO} was negatively correlated with the toxicity which could be seen in equation 2. E_{HOMO} is the electronic energy identical to the corresponding negative value of the ionization potential which can be used to measure the donating electron ability [22]. E_{HOMO} models the nucleophilic nature of the OPs which is important for their reaction with the active site of AChE [14]. Atomic charges are often used as an important concept and describe electronic aspects both of the whole molecule and of particular regions or fragments. They are often used in QSAR studies as descriptors [1,5]. In this study, the charge on the phosphorus atom is particularly important, because it is related with the toxicity of the OPs. The Mulliken charge analysis of the OP compounds shows that the increase in the positive charge on the phosphorus atom in a molecule leads to an increase in toxicity in most of the OP compounds. Fig. 5 shows the molecular electrostatic potential (MEP)

plot for OP12, OP14 and OP19. In the MEP plot, blue and green colors show positive electrostatic potential whereas red color shows negative electrostatic potential. As can be seen in Fig. 5, positive electrostatic charge regions belong to phosphorus and hydrogen atoms. Negative electrostatic charge regions represent oxygen atoms. In the electrostatic potential map, the phosphorus atom is in the centre of negative and positive charge separation. So inhibition reaction occurs at this center and X is the leaving group shown in Fig. 1. The inhibition is a nucleophilic substitution which replaces the X group with the hydroxyl group of serine in the active site of AChE [14,27]. Overall, the toxic effect of OPs is mostly related with electronic descriptors as η, E_{HOMO}, D and P_{ch} which show that electron exchange occurs between OPs and biological molecules.

Table 4. Predicted and observed log LD_{50} of OPs from model equations

	log LD_{50exp}	Eq.1	Residual	Eq.2	Residual	Eq.3	Residual
OP 1	3.5163	3.4462	0.0701	3.4859	0.0304	3.4660	0.0503
OP 2	3.9143	3.3836	0.5307	3.3904	0.5239	3.3662	0.5481
OP 3	3.2788	3.2310	0.0478	3.2136	0.0652	3,1843	0.0945
OP 4	3.0394	3.2438	-0.2044	3.2625	-0.2231	3.0878	-0.0484
OP 5	3.6385	3.3844	0.2541	3.4165	0.2220	3.5587	0.0798
OP 6	2.9031	3.4093	-0.5062	3.4068	-0.5037	3.4075	-0.5044
OP 7	3.3674	3,3469	0.0205	3.3330	0.0344	3.3299	0.0375
OP 8	3.2041	3.2649	-0.0608	3.2602	-0.0561	3.3501	-0.1460
OP 9	2.9930	3.1720	-0.1790	3.1504	-0.1574	3,1477	-0.1547
OP 10	0.7993	1.0728	-0.2735	1.0475	-0.2482	0.9130	-0.1137
OP 11	1.2625	1.0903	0.1722	1.0326	0.2299	1.0750	0.1875
OP 12	1.2742	1.0104	0.2638	0.9163	0.3579	1.0057	0.2685
OP 13	0.3010	0.8663	-0.5653	0.7952	-0.4942	0.7439	-0.4429
OP 14	1.3424	1.5163	-0.1739	1.4793	-0.1369	1.5320	-0.1896
OP 15	1.3010	1.2325	0.0685	1.1483	0.1527	1.1979	0.1031
OP 16	1.7076	1.0277	0.6799	1.0079	0.6997	0.9499	0.7577
OP 17	0.5185	0.8388	-0.3203	0.6665	-0.1480	0.8112	-0.2927
OP 18	0.6721	0,9450	-0.2729	0.9475	-0.2754	0.7356	-0.0635
OP 19	0.9494	0.9843	-0.0349	0.9653	-0.0159	1.1198	-0.1704
OP 20	0.9494	1.0310	-0,0816	1.0066	-0,0572	0.9498	-0.0004
OP21	1.0000	0.9919	0.0081	0.9243	0,0757	0.8144	0.1856
OP22	1.2304	1.4764	-0.2460	1.4033	-0.1729	1.3400	-0.1096
OP23	0.8451	1.1108	-0.2657	1.1699	-0.3248	1.0023	-0.1572
OP24	0.8325	0.8518	-0.0193	0.8776	-0.0451	0.9145	-0.0820
OP25	1.1139	0.9240	0.1900	0.9065	0.2074	1.0689	0.0451

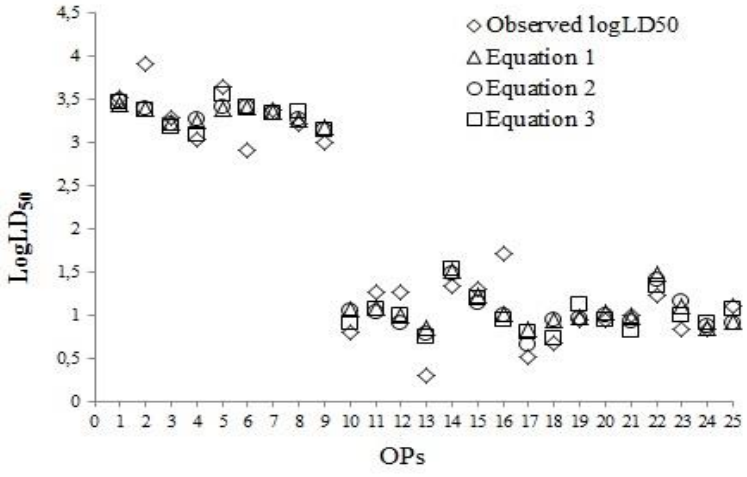

Figure 4. Comparison between the predicted and observed log LD_{50} values of OPs

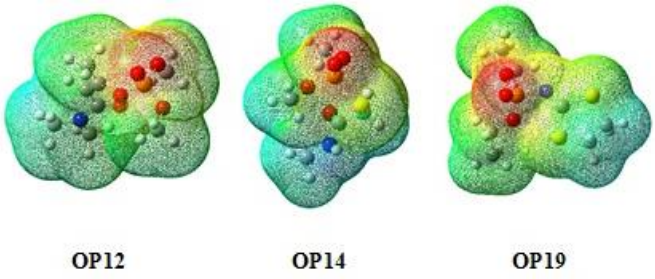

Figure 5. Molecular electrostatical potential representation of selected OPs

CONCLUSIONS

QSAR analysis was performed in this work for selected OPs on their oral toxicity to rats using quantum chemical descriptors. The usefulness of descriptors in the development of QSAR analysis was clarified by statistical analysis. The derived model equations were statistically significant and can be used for prediction purposes; they may be helpful for a better understanding of the toxicity of this class of compounds. It could be concluded that hardness, dipole moment, highest occupied molecular orbital energy and charge of phosphorus atom can be used as descriptors in the prediction of the toxicity of OPs. Due to the success of the developed regression model, it may be utilized to predict the toxicity of other OPs whose experimental toxicity data are not available.

Acknowledgement: *The authors express their thanks to Yıldız Technical University Research Foundation (Project No: 2012-01-02-KAP02), for financial support.*

REFERENCES

1. Y. Paukku, G. Hill, *Int. J. Quant. Chem*., **112**, 1343 (2012).
2. R. Garcı´A-Domenech, P. Alarco N-Elbal, G. Bolas,R. Bueno-Marı, F. A. Chorda -Olmos, S. A. Delacour, M. C. Mourın Oz, A. Vıdal, J. Ga´ Lvez, *J. SAR QSAR Environ. Res.*, **18**, 745 (2007).
3. G. Davila-Vazquez, R.Tinoco, M.A.Pickard, R.Vazquez-Duhalt, *Enzyme Microb. Technol*., **36**, 223 (2005).
4. C.T. Duarte, R.Roman, R.Tinoco, R.V.Duhalt, *Chemosphere*,**77**, 687 (2009).
5. S.A. Senior, M.D.Madbouly, A.M. El Massry, *Chemosphere*, **85**, 7 (2011).
6. M. B.Colovic, V. M. Vasic, N. S. Avramovic, M. M. Gajic, D. M. Djuric, D. Z. Krstic, *Toxicol. Lett*., **233**, 29 (2015).
7. Y. Arima, H. Shiraishi, A. Saito, K. Yoshimoto, A. Namera, R. Makita, K. Murata, K. Imaizumi, M. Nagao, *J. Toxicol. Sci*., **141**, 617 (2016).
8. S. Tao, X.H. Xi, F.L. Xu, B.G.Li, J.Cao, R. Dawson, *Environ. Pollut*., **116**, 57 (2002).
9. D. Yan , X. Jiang, S. Xu, L. Wang, Y. Bian, G. Yu, *Chemosphere*, **71**, 1809 (2008).
10. A. Mohajeri, M.H. Dinpajooh, *J. Mol. Struct: THEOCHEM*, **855**, 1 (2008).
11. C. Hansch, D. Hoekman, A. Leo, D. Weininger, C.D. Selassie, *Chem.Rev*., **102**, 783 (2002).
12. M. Elango, R. Parthasarathi, V. Subramanian, U. Sarkar, P.K. Chattaraj, *J. Mol. Struct: THEOCHEM*, **723**, 43 (2005).
13. F.Worrall, M.Thomsen, *Chemosphere*, **54**, 585 (2004).
14. E. Zvinavashe,T. Du, T. Griff, H.H.J. van den Berg, A.E.M.F. Soffers, J. Vervoort, A. J. Murk, I.M.C.M. Rietjens, *Chemosphere*, **75**, 1531 (2009).
15. D. Yan, X. Jiang, G. Yu, Z. Zhao,Y. Bian, F. Wang, *Chemosphere*, **63**, 744 (2006).
16. R.Todeschini, M. Vighi, R. Provenzani, A. Finizio, P. Gramatica, *Chemosphere*, **8,** 1527 (1996).
17. J. Devillers, *J. Toxicol. Methods*, **10**, 69 (2000).
18. J. Devillers, *J. SAR QSAR Environ. Res.*, **15**, 501(2004).
19. Gaussian 03, Revision B.04.Gaussian, Inc., Pittsburgh, PA, 2003.
20. J. B. Foresman, A. Frisch, Exploring Chemistry with Electronic Structure Methods, Gaussian Inc., USA, 1996.
21. J.Shi, M. Feng, X. Zhang, Z.Wei, Z. Wang, *Environ. Sci. Pollut. Res*., **20,** 6138 (2013).
22. R.G.Pearson, *J. Chem. Sci*., **117**, 369 (2005).
23. .https://www.chem.nlm.nih.gov/chemidplus/ accessed 2017.
24. K. Roy, R. Narayan Das, P. Ambure, R. B. Aher, *Chemometr. Intell. Lab*., **152**, 18 (2016).
25. H.V.Waterbeemd, Chemometric Method in Molecular Design, Wiley, N.Y., 2008.
26. A. Furuhama, K. Hasunuma, Y. Aoki, *SAR QSAR in Environ. Res.*, **26**, 301 (2015).
27. M. Stoytcheva, Pesticides - The Impacts of Pesticide Exposure, InTech, India, 2011.