

## Inhibition of *Tetrahymena pyriformis* growth by Aliphatic Alcohols and Amines: a QSAR Study

Nadia Ziani, Khadidja Amirat & Djelloul Messadi

Laboratoire de Sécurité Environnementale et Alimentaire (LASEA),  
Université Badji Mokhtar – Annaba, 23 000, Annaba, Algérie

Soumis le : 07.06.2013

Révisé le : 30.04.2014

Accepté le : 04.06.2014

### ملخص

أنجزت دراسة العلاقة بين الكمية البنيوية والنشاط (QSAR) لتحقيق التسمم النسبي لمزيج مكون من 21 كحول (ذات سلاسل خطية و متفرعة) و 9 أمينات اليقاتية عادية للتعبير على تركيز المعيقات بنسبة لنمو 50% (IGC<sub>50</sub>) (*Tetrahymena pyriformis*). طريقة الانحدار الخطي البسيط اعتمدت على الموصفات الجزئية النظرية (هندسية), ثلاثية الأبعاد 3D, المتحصل عليها اعتمادا على برنامج DRAGON, والواصفات المختلفة log P المحسوبة. الصلابة و القدرة التنبؤية للنموذج تم التحقق منها اعتمادا على التصديق الداخلي (تصديق متقاطع) (Loo, LMO, bootstrap) و كذلك التصديق الخارجي Clog P ظهر بأنه أحسن واصف لنمذجة التسممية, و الذي يمكن تعويضه بالواصف الهندسي ADDD, بدون تغير ملحوظ للمعايير الإحصائية.

الكلمات المفتاحية: كحول وأمينات-تسمم مائي QSAR - الوصفات الهندسية و خصائص مائية - الانحدار الخطي البسيط.

### Résumé

Une étude Relation Quantitative Structure- Activité (QSAR) a été réalisée pour évaluer la toxicité relative d'un mélange composé de 21 alcools (à chaînes linéaires et ramifiées) et 9 amines aliphatiques normales, en terme de concentration d'inhibition 50% de la croissance (IGC<sub>50</sub>) de *Tetrahymena pyriformis*. L'approche par régression linéaire simple est basée sur des descripteurs moléculaires théoriques (géométriques) 3D obtenus à l'aide du logiciel DRAGON et différents descripteurs logP calculés. La robustesse et la capacité prédictive des modèles ont été vérifiées à l'aide de statistiques de validations internes (validations croisées LOO et LMO ; bootstrap) et externe. Clog s'est avéré le meilleur descripteur pour la modélisation de la grandeur d'intérêt considérée. Il peut être remplacé par le descripteur géométrique ADDD sans variations appréciables des paramètres statistiques.

**Mots clés:** Alcools et Amines – Toxicité aquatique – QSAR – Descripteurs géométriques et caractère hydrophobe – Régression linéaire simple.

### Abstract

A Quantitative Structure- Activity Relationship (QSAR) study was undertaken to evaluate the relative toxicity of a mixed series of 21 (linear and branched-chain) alcohols and 9 normal aliphatic amines in term of the 50% inhibitory growth concentration (IGC<sub>50</sub>) of *Tetrahymena pyriformis*. The applied simple linear regression approach is based on theoretical 3D (geometrical) molecular descriptors from DRAGON package, and some calculated logP descriptors. The robustness and the predictive performance of the models were verified using both internal (cross-validation by LOO and LMO; bootstrap) and external statistical validations. ClogP turned out to be the best descriptor to model the considered endpoint. It may be interchanged with geometrical descriptor ADDD without relevant variations in the statistical parameters.

**Keywords:** Alcohols and Amines –Aquatic toxicity – QSAR – Geometrical descriptors and hydrophobic character – Simple linear regression.

Auteur correspondant : d\_messadi@yahoo.fr

## 1. INTRODUCTION

The impact of the potential hazard of untested chemicals, a challenge confronting national and international regulatory agencies [1-4] can be measured by experimental investigations, but this approach is both quite expensive and time consuming [5]. An alternative is to rely on QSAR (Quantitative Structure-Activity Relationships) models that describe a mathematical relationship between the structural features of a set of chemicals and the particular activity associated with them [6,7].

Several QSAR models predicting acute chemical toxicity for aquatic environment have been published [8-12]. They are based mainly on the logarithm of the octanol-water coefficient (logP, also referred to as logKow) as this hydrophobicity term reproduces the ability of a substance to enter cells through the lipid membranes and indicates both the toxicant uptake and baseline toxicity.

Albeit the number of compounds with a measured value for the logP was estimated to be 30 000 [13], which seems at a first glance to be high, this is negligible compared to the rapidly increasing number of compounds for which logP values are desired but missing. Furthermore, the experimental determination is tedious, time-consuming and demands a high purity of the solute [14]; none of these preconditions are compatible with high-throughput techniques, there is, therefore, an ongoing interest in methods for the prediction of logP values.

Over recent decades various approaches (fragmental, atom-based, conformation – dependent methods) [15-18] have been developed that are mostly implemented and available as computer programs. However, even in these calculations it is not uncommon to have differences of several order of magnitude [19, 20].

For these reasons logP cannot be considered a univocal descriptor, which brought different authors [19-24] to propose toxicity models based exclusively on other structural theoretical molecular descriptors.

The present paper proposes predictive simple linear regression QSAR models to evaluate the relative toxicity of organic chemicals, in terms (of the logarithm of the inverse) of the 50% inhibitory growth concentration (IGC50) of

*Tetrahymena pyriformis*. Models based on different kinds of logP (calculated values for AlogP, MlogP and ClogP), are compared to the optimal model constructed using a single 3D (geometrical) descriptor calculated from the chemical structure alone.

## 2. METHODS

### 2.1 Experimental Data

Two different toxicants were studied: a set of 21 (linear and branched-chain) alcohols and 9 normal aliphatic amines, selected to reflect diversity in chain length and branching.

These toxicants, which are both nonionic and nonreactive, inhibit the growth concentration of *Tetrahymena pyriformis* the most tested common freshwater hymenostome ciliate, which approximately measures 50  $\mu\text{m}$  in length and 30  $\mu\text{m}$  in width [25].

The ciliates were grown in axenic culture with population density being measured spectrophotometrically as optical density (absorbance) at 540 nm following 48h of incubation.

The set of experimental data was taken from Schultz [26].

### 2.2 Estimation of octanol /water partition coefficient

**ClogP** ( $\equiv$  *calculated logP*) [17]

The software version of the system developed by Hansch and Leo [27], using the Rekker's additive scheme [28], is known as ClogP (or calculated logP). It is based on different fragmental constants and correction terms. Fragment constants were derived from solutes where the fragment occurs in isolation. Furthermore, the bonding environment was taken into account (alkyl, benzyl, vinyl, styryl, and aromatic neighbors) resulting in five values per fragment. If a fragment in combination with the bonding environment was missing but at least two values for the same fragment with different neighbors could be found an interpolation was attempted to derive the missing data. The correction factors have been calculated from the corrections required for the specific interactions being modelled. For instance, the interaction of the two hydroxyl groups in diethylene glycol increases the logP value by 0.85 compared with two hydroxyl

groups that do not interact. This value is then taken as the correction term for a two-neighbored hydroxyl group [29].

The decomposition of the molecular structure into fragments is performed by using a unique and simple set of rules, thus obtaining a unique solution; the fragments are either atoms or polyatomic groups.

**AlogP** ( $\equiv$  Ghose and Crippen model based on atomic increment system) [30]

Several models have been published where the fragments are defined on a purely atomic level. This simplifies both the recognition of fragments and the calculation, as correction substructures are not applied (see Eq.(1)).

The most frequently used atomic increment system, AlogP, was developed by Ghose and Crippen [31]. Atoms are classified by their neighboring environment and carbon atoms additionally by their hybridization.

Estimated logP for any compound is given by:

$$\text{AlogP} = \sum_i n_i a_i \quad (1)$$

where  $n_i$  is the occurrence of the  $i$ th atom type and  $a_i$  is the corresponding hydrophobicity constant.

The AlogP model implemented in DRAGON has been evaluated on a set of 2648 compounds with known experimental logP taken from the NCI open Data Base. The resulted correlation coefficient  $r$  is 0.915.

**MlogP** ( $\equiv$  Moriguchi model based on structural parameters)

This is a model described by a regression equation based on 13 structural parameters [32, 33].

The regression coefficients have been evaluated by a training set of 1230 organic molecules including general aliphatic aromatic and heterocyclic compounds containing the following atoms: C, H, N, O, S, P, F, Cl, Br, I [30]. The statistical parameters of the model are  $r = 0.952$ ;  $SE = 0.422$ ;  $F_0(13;1216) = 900.4$

### 2.3 Geometrical Descriptors Generation

The chemical structure of each compound was sketched on a PC using the HYPERCHEM program [34] and pre optimized using MM<sup>+</sup> molecular mechanics method (Polack-Ribiere

algorithm). The final geometries of the minimum energy conformation were obtained by the semi empirical PM3 method at a restricted Hartree-Fock level with no configuration interaction applying a gradient norm limit of 0.01 kcal. Å<sup>-1</sup>. mol<sup>-1</sup> as a stopping criterion.

The resulted geometries were used as input for the generation of (74) 3D- geometrical descriptors using the DRAGON software (version 5.3) [30].

Geometrical descriptors being defined from the three dimensional structure of the molecule, which involves the knowledge of the relative positions of the atoms in 3D space provide information and discrimination power also for similar molecular structures and molecule conformations.

### 2.4 Chemometric Methods

Models with one variable were performed by the software MOBYDYGs [35] using the Ordinary Least Square regression (OLS) method.

Population of 74 regression models corresponding to each of the 74 Geometry descriptors were ordered according to their decreasing internal predictive performance, verified by  $Q^2$ , optimal non-logP model was then selected and compared to the three logP – based models.

The goodness of fit of the calculated models were assessed by means of the multiple determination coefficients,  $R^2$ , and the standard deviation error in calculation (SDEC).

$$SDEC = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

Cross validation techniques allow the assessment of internal predictivity ( $Q^2_{LMO}$  cross validation; bootstrap) in addition to the robustness of model ( $Q^2_{LOO}$  cross validation).

Cross validation methods consist in leaving out a given number of compounds from the training set and rebuilding the model, which is then used to predict the compounds left out. This procedure is repeated for all compounds of the training set, obtaining a prediction for everyone. If each compound is taken away one at a time the cross validation procedure is called leave-one-out technique (LOO technique), otherwise

leave-more-out technique (LMO technique). An LOO or LMO correlation coefficient, generally indicated with  $Q^2$ , is computed by evaluating the accuracy of these “test” compounds prediction.

$$Q^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_{i/i})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{PRESS}{TSS} \quad (3)$$

The “hat” of the variable  $y$ , as is the usual statistical notation, indicates that it is a predicted value of the studied property, and the sub index “i/i” indicates that the predicted values come from models built without the predicted compound.

TSS is the total sum of squares.

The predictive residual sum of squares (PRESS) measures the dispersion of the predicted values. It is used to define  $Q^2$  and the standard deviation error in prediction (SDEP).

$$SDEP = \sqrt{PRESS/n} \quad (4)$$

A value  $Q^2 > 0.5$  is generally regarded as a good result and  $Q^2 > 0.9$  as excellent [36, 37]. However, studies [38, 39] have indicated that while  $Q^2$  is a necessary condition for high predictive power a model, is not sufficient.

To avoid overestimating the predictive power of the model LMO procedure (repeated 5000 times, with 5 objects left out at each step) was also performed ( $Q^2_{L(5)O}$ ).

In bootstrap validation technique  $K$  n-dimensional groups are generated by a randomly repeated selection of  $n$ -objects from the original data set. The model obtained on the first selected objects is used to predict the values for the excluded sample, and then  $Q^2$  is calculated for each model. The bootstrapping was repeated 8000 times for each validated model.

By using the selected model the values of the response for the test objects are calculated and the quality of these predictions is defined in terms of  $Q^2_{ext}$ , which is defined as

$$Q^2_{ext} = 1 - \frac{\sum_{i=1}^{n_{ext}} (\hat{y}_{i/i} - y_i)^2 / n_{ext}}{\sum_{i=1}^{n_{tr}} (y_i - \bar{y}_{tr})^2 / n_{tr}} = 1 - \frac{PRESS / n_{ext}}{TSS / n_{tr}} \quad (5)$$

Here  $n_{ext}$  and  $n_{tr}$  are the number of objects in the external set (or left out by bootstrap) and the number of training set objects, respectively.

The data set randomly was divided into a training set (20 objects) used to develop the QSAR models and a validation set (10 objects), used only for statistical external validation.

Other useful parameters are  $R^2$ , calculated for the validation chemicals by applying the model developed on the training set, and external standard deviation error of prediction ( $SDEP_{ext}$ ), defined as:

$$SDEP_{ext} = \sqrt{\frac{1}{n_{ext}} \sum_{i=1}^{n_{ext}} (y_i - \bar{y})^2} \quad (6)$$

where the sum runs over the test set objects ( $n_{ext}$ ).

### 3. RESULTS AND DISCUSSION

The best one dimensional non-logP model was obtained using the average distance-distance degree (ADDD) index. It encodes information on the molecular folding [40, 41] information about molecular diffusion easiness through biological barriers like membranes. Distance/distance matrices, denoted as  $D/D$ , were defined as quotient matrices in terms of geometric  $r_{ij}$  distances and topological distance  $d_{ij}$ :

$$[G/D]_{ij} = \begin{cases} \frac{r_{ij}}{d_{ij}} & \text{if } i \neq j \\ 0 & \text{if } i=j \end{cases} \quad (7)$$

The row sums of these matrices contain information on the molecular folding; in effect, in highly folded structures, they tend to be relatively small as the inter-atomic distances are small while the topological distances increase as the size of the structure increases.

Therefore, the average row sum is a molecular invariant called average distance-distance degree, that is:

$$ADDD = \frac{1}{A} \sum_{i=1}^A \sum_{j=1, j \neq i}^A \frac{r_{ij}}{d_{ij}} \quad (8)$$

$A$  being the number of molecule atoms. Table 1 lists the CAS number,  $-\log IGC_{50}$ ,  $AlogP$ ,  $MlogP$ ,  $ClogP$  and ADDD values of the selected aliphatic alcohols and amines.

Table 1. Relative toxicity and molecular descriptors data for the selected aliphatic alcohols and amines

Compound	CAS number <sup>(a)</sup>	-logIGC <sub>50</sub>	AlogP	MlogP	ClogP	ADDD
Methanol	67-56-1	-2.77	-0.358	-0.814	-0.764	4.964
Ethanol	64-17-5	-2.41	0.009	-0.172	-0.235	8.108
1-propanol	71-23-8	-1.84	0.515	0.347	0.294	11.098
1-pentanol	71-41-0	-1.12	1.427	1.209	1.352	17.194
1-hexanol	111-27-3	-0.47	1.883	1.587	1.881	20.305
1-heptanol	111-70-6	0.02	2.339	1.940	2.410	23.451
1-nonanol	143-08-8	0.77	3.252	2.591	3.468	29.859
1-decanol	112-30-1	1.1	3.708	2.894	3.997	33.118
1-dodecanol	112-53-8	2.07	4.620	3.467	5.055	39.716
1-tridecanol	112-70-9	2.28	5.077	3.739	5.584	43.053
2-propanol	67-63-0	-1.99	0.368	0.347	0.074	10.987
2-methyl-1-butanol	137-32-6	-1.13	1.290	1.209	1.222	16.810
3-methyl-1-butanol	123-51-6	-1.13	1.223	1.209	1.222	16.833
3-methyl-2-butanol	598-75-4	-1.08	1.211	1.209	1.002	16.622
(tert)pentanol	75-85-4	-1.27	1.097	1.209	1.002	16.640
1-propylamine	107-10-8	-0.85	0.225	0.347	0.394	11.968
1-hexylamine	11-26-2	-0.34	1.594	1.587	1.981	21.226
1-heptylamine	111-68-2	0.1	2.050	1.940	2.510	24.393
1-octylamine	111-86-4	0.51	2.506	2.274	3.039	27.602
1-undecylamine	7307-55-3	2.26	3.875	3.186	4.626	37.408
1-butanol*	71-36-3	-1.52	0.971	0.800	0.823	14.138
1-otanol*	111-87-5	0.5	2.796	2.274	2.939	26.640
1-undecanol*	112-42-5	1.87	4.164	3.186	4.526	36.401
2-pentanol*	6032-29-7	-1.25	1.348	1.209	1.132	16.888
3-pentanol*	584-02-1	-1.33	1.416	1.209	1.132	16.908
(neo) pentanol*	75-84-3	-0.96	1.108	1.209	1.092	16.710
1-butylamine*	109-73-9	-0.7	0.681	0.800	0.923	15.015
1-amylamine*	110-58-7	-0.61	1.137	1.209	1.452	18.101
1-nonylamine*	112-20-9	1.59	2.962	2.591	3.568	30.837
1-decylamine*	2016-57-0	1.95	3.418	2.894	4.097	34.110

(a): Chemical Abstract Services registry number ; (\*): validation set compound.

Intercepts ( $\beta_0$ ) and slopes ( $\beta_1$ ) of the calculated one dimensional models are shown in table 2

Table 2. Coefficients for the ordinary least squares calculated models.

X	AlogP	MlogP	ClogP	ADDD
$\beta_0$	-2.144 ( $\pm 0.139$ )	-2.230 ( $\pm 0.134$ )	-1.998 ( $\pm 0.088$ )	-3.333 ( $\pm 0.134$ )
$\beta_1$	0.939 ( $\pm 0.057$ )	1.191 ( $\pm 0.068$ )	0.815 ( $\pm 0.033$ )	0.138 ( $\pm 0.006$ )

Relevant statistical parameters reported in table 3 below clearly show the difference in fitting and prediction performances for the selected logP descriptors, ClogP appear as the best ones.

Another remarkable fact observed is that ClogP and ADDD theoretical molecular descriptor can be interchanged without relevant variations in the statistical results.

Table 3. Summary statistics for the one dimensional calculated models.

X	R <sup>2</sup>	Q <sup>2</sup>	Q <sup>2</sup> <sub>L(5)O</sub>	Q <sup>2</sup> <sub>boot</sub>	Q <sup>2</sup> <sub>ext</sub>	SDEC	SDEP	SDEP <sub>ext</sub>	F <sub>(p=0.000)</sub>	SE
Alog P	0.9372	0.9188	0.9165	0.9063	0.8584	0.365	0.415	0.554	268.4	0.385
Mlog P	0.9444	0.9291	0.9270	0.9160	0.8998	0.344	0.388	0.461	305.52	0.362
Clog P	0.9713	0.9630	0.9621	0.9580	0.9352	0.247	0.280	0.371	610.02	0.260
ADDD	0.9711	0.9624	0.9616	0.9566	0.9326	0.248	0.283	0.378	603.82	0.261

Carbö –Dorca *et al.* [42] reported a QSAR study where the same data was examined, these authors constructed a predictive model using, as a molecular descriptor, the expectation value of the inter electronic repulsion energy operator presented as a kind of quantum self-similarity measure (QS-SM). The correlation results reached R<sup>2</sup> = 0.9240, Q<sup>2</sup> = 0.9090 and SE = 0.415, which are inferior than the present approach.

The value of R<sup>2</sup> attests the good fitting performances of the model. In general, the larger the magnitude of the F ratio, the better the model predicts the property values in the training set. The large F ratio of 603.82 indicates that the model does an excellent job of predicting the -log IGC<sub>50</sub> values. The model is robust, the difference between R<sup>2</sup> and Q<sup>2</sup> is small (<1%). Figure 1 shows a plot contrasting experimental and cross-validated -log IGC<sub>50</sub>. The point dispersion is small, although there is one point a little bit far away from the rest (1-propylamine). SDEP is similar to SDEC, so this model has internal predictivity not so dissimilar from fitting power.

The model demonstrates a very good stability in internal validation (difference between Q<sup>2</sup> and Q<sup>2</sup><sub>L(5)O</sub> is 0.33% ), while bootstrapping confirms the internal predictivity and stability of the model.

Though small sized the data set underwent statistical validation by preliminary random splitting of the chemicals into training (20 chemicals) and validation (10 chemicals) sets. The small size of the published experimental

data set [26] did not allow a more drastic splitting. The information obtained by Q<sup>2</sup><sub>ext</sub> is

somewhat optimistic. In fact with small data sets (20-30 chemicals), completely new chemicals external predictivity can only be verified *a posteriori*, case -by-case.

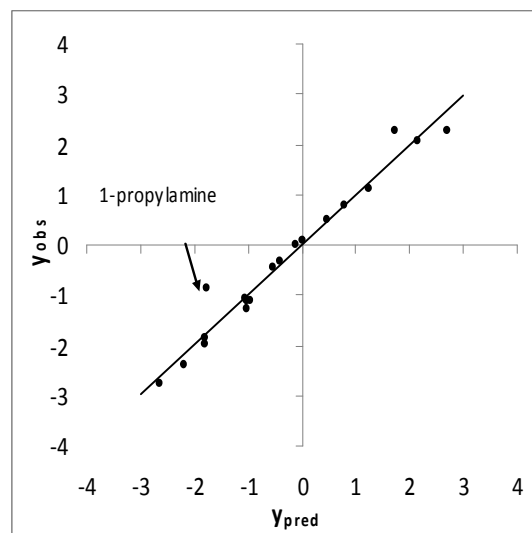


Figure-1 Experimental versus cross-validation activity for the training set objects.

#### 4. CONCLUSION

From the results and discussion above we conclude that:

1. Among the logP descriptors selected to model the inhibition of *Tetrahymena pyriformis*

growth by aliphatic alcohols and amines, Clog P is the best.

2. Geometrical descriptor ADDD (Average Distance/Distance Degree), a theoretical molecular descriptor, and Clog P can be interchanged without relevant variations in the statistical results.

3. The non- logP model obtained in this study has very good fitting performances, is robust and with acceptable predictive power. The internal validation parameters ( $Q^2_{LOO}$ ,  $Q^2_{L(5)O}$  and bootstrap ) are similar to the fitting parameters.

Notice that chemical 19 (1-propyl amine) with cross validated standardized residuals (not reported here) greater than 3 standard deviation units, is an heavy outlier in all the models considered in this paper.

## REFERENCES

- [1] Zeeman M., Aver C.M., Clements R.G., Nabholz J.V. & Boethling R.S., 1995. U.S. EPA Regulatory Perspectives on the use of QSAR for new and existing chemical evaluations SAR QSAR, *Environmental Research*, Vol. 3(3),179-201.
- [2] Walker J.D., 2003. Applications of QSARs in toxicology: a US Government perspective, *Journal of Molecular Structure - Theochem*, Vol. 622(1-2), 167-184.
- [3] Bradbury S.P., Russon C.L., Ankley G.T., Schultz T.W. & Walker J.D., 2003. Overview of data and conceptual approaches for derivation of Quantitative Structure –Activity Relationships, for ecotoxicological effects of organic chemicals, *Environmental Toxicology & Chemistry*, Vol. 22 (8), 1789-1798.
- [4] European Commission. White Paper on a strategy for a future Community Policy for Chemicals., 2001.[http:// europa .eu.int / comm / enterprise / reach /](http://europa.eu.int/comm/enterprise/reach/).
- [5] Toussaint M.W., Shedd T.R., Van der Schalie W.H. & Leather G.R., 1995. A comparison of standard acute toxicity tests with rapid screening toxicity tests. *Environmental Toxicology & Chemistry*, Vol. 14(5), 907-915.
- [6] Kubinyi H., 2002. From Narcosis to Hyperspace: The History of QSAR, *Quantitative Structure.-Activity Relationships.*, Vol. 21(4), 348-356.
- [7] [http:// e c b . j r c . i t / Q S A R /](http://ecb.jrc.it/QSAR/).
- [8] Schultz T.W., Cronin M.T.D., Walker J.D. & Aptula A.O., 2003.Quantitative structure –activity relationships (QSAR<sub>s</sub>) in toxicology: a historical perspective, *Journal of Molecular Structure –Theochem*, Vol.622(1-2), 1-22.
- [9] Posthumus R. & Slooff W., 2001. Implementation of QSARs in ecotoxicological risk assessments RIVM report 601516003.
- [10] Dearden J.C., 2002. Prediction of Environmental Toxicity and Fate Using Quantitative Structure –Activity Relationships (QSARs), *Journal of Brazilian Chemical Society*, Vol . 13 (6), 754-762.
- [11] Schultz T.W., Cronin M.T.D. & Netzeva T.I., 2003. The present status of QSAR in toxicology, *Journal of Molecular Structure -Theochem*, Vol. 622(1-2), 23-38.
- [12] Cronin M.T.D. & Dearden J.C., 1995. QSAR in toxicology .1. Prediction of Aquatic Toxicity, *Quantitative Structure.-Activity. Relationships.*, Vol.14(1), 1-7.
- [13] Mannhold R. & van de Waterbeemd H., 2001. Substructure and whole molecule approaches for calculating logP, *Journal of Computer- Aided Molecular Design*, Vol. 15(4), 337-354.
- [14] Mannhold R. & Rekker R.F., 2000. The hydrophobic fragmental constant approach for calculating log P in octanol/water and aliphatic hydrocarbon/water systems. *Perspectives in. Drug Discovery & Design*, Vol.18(1), 1-18.
- [15] Benfenati E., Gini G., Piclin N., Roncaglioni A. & Vari M.R ., 2003.Predicting log P of pesticides using different software, *Chemosphere*, Vol.53(9), 1155-1164.
- [16] Manhold R. & Petrauskas A., 2003. Substructure versus Whole-molecule Approaches for Calculating Log P, *QSAR & Combinatorial Science*, Vol. 22(4), 466-475.1
- [17] <http://clogP.pomona.edu/medchem/chem/clogP/index.html>.
- [18] Klopman G., Li J.K., Wang S. & Dimayuga M., 1994.Computer Automated log P calculations based on an extended group contribution approach, *Journal of Chemical. Information. Computer Sciences*, Vol.34(4), 752-781.
- [19] Kaiser K.L.E., 2003. The use of neural networks in QSARs for acute aquatic toxicological endpoints, *Journal of.Molecular Structure. Theochem*, Vol .622(1-2), 85-95.
- [20] Papa E., Villa F. & Gramatica P., 2005. Statistically Validated QSARs Based on Theoretical Descriptors , for Modeling Aquatic Toxicity of Organic Chemicals in *Pemiphales promelas* (Fathead Minnow ), *Journal of Chemical Information & Modeling*, Vol.45(5), 1256-1266.
- [21] Roy K. & Ghosh G., 2009.QSTR with extended topochemical atom (ETA) indices. 12. QSAR for the toxicity of diverse aromatic compounds to *Tetrahymena pyriformis* using chemometric tools. *Chemosphere*, Vol. 77(7), 999-1009.
- [22] Zhao Y.H., Zhang X.J.,WEN Y., Sun F.T., Guo Z., Qin W.C., Qin H.W.,Xu J.L., Sheng L.X. & Abraham M.H., 2010. Toxicity of organic chemicals to *Tetrahymena pyriformis* : Effect of polarity and ionization on toxicity. *Chemosphere*, Vol. 79(1), 72-77.
- [23] Roy K. & Das R.N., 2010.QSTR with extended topochemical atom (ETA) indices.14. QSAR modeling of toxicity of aromatic aldehydes to *Tetrahymena pyriformis*. *Journal of Hazardous Materials*, Vol. 183(1-3), 913-922.
- [24] Bouaoune A., Lourici L., Haddag H. & Messadi D., 2012. Inhibition of Microbial Growth by anilines: A QSAR study, *Journal of Environmental Science and Engineering.*, A1, Vol. 1(5A), 663-671.
- [25] Hill D.L., 1972. The biochemistry and physiology of *Tetrahymena*. Academic Press, New York & London, 230p
- [26] Schultz T.W., Lin D.T., Wilke T.S. & Arnold L.M., 1990. Quantitative structure-activity relationships for the

- Tetrahymena pyriformis* population growth endpoint : a mechanism of action approach. In Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology W. Karcher, J. Devillers (Eds) Kluwer Academic Publishers. Dordrecht, Vol.1., 241-262.
- [27] Hansch C. & Leo A., 1979. Substituent Constants for Correlation Analysis in Chemistry and Biology. Wiley & Sons Inc. New York (NY), 339p.
- [28] Nys C.G. & Rekker R.F., 1973. Statistical Analysis of a series of Partition Coefficients with Special Reference to the Predictability of Folding of Drug Molecules. The introduction of Hydrophobic Fragmental Constants (f values), *Chimica. Therapeutica* Vol.8, 521-535.
- [29] Kleinöder T., Yan A & Spycher S., 2008. Estimation of Octanol/Water Partition coefficient ( $\log P_{ow}$ ). In *Cheminformatics J. Gasteiger, T. Engel (Eds)*, WILEY-VCH GmbH & Co. kgaA, Weinheim, 492-494.
- [30] Todeschini R., Consonni V., Mauri A. & Pavan M. 2005 DRAGON Software for the Calculation of Molecular Descriptors—version 5.3 for Windows, Talete s.r.l., Milano, Italy.
- [31] Ghose A.R. & Crippen G.M., 1986. Atomic Physicochemical Parameters for Three-Dimensional-Structure-Directed-Quantitative Structure -activity relationships .I-Partition Coefficients as a Measure of Hydrophobicity, *Journal of Computational Chemistry*, Vol. 7(4), 565-577.
- [32] Moriguchi I., Hirono S., Liu O., Nakagome I. & Matsushita Y., 1992. Simple Method of Calculating Octanol/Water Partition Coefficient, *Chemical & Pharmaceutical Bulletin*, Vol.40(1), 127-130.
- [33] Moriguchi I., Hirono S., Nakagome I. & Hirano H., 1994. Comparison of Reliability of LogP Values for Drugs Calculated by Several Methods, *Chemical & Pharmaceutical Bulletin*, Vol. 42(4), 976-978.
- [34] HYPERCHEM/CHEMPLUS ver.7.03 for Windows Autodesk Inc., Sausalito, CA, U.S.A. 2002.
- [35] MOBYDIGS – Models BY Descriptors In Genetic Selection – ver. 1.1 for Windows, Talete S.r.l., Milano, Italy.
- [36] Eriksson L., Jaworska J., Worth A., Cronin M Mc., Dowell R.M. & Gramatica P., 2003. Methods for Reliability, uncertainty assessment , and applicability evaluations of regression based and classification QSARs, *Environmental Health Perspectives*, Vol. 111(10),1361-1375.
- [37] Tropsha A., Gramatica P. & Grombar V.K., 2003.The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models, *QSAR & Combinatorial Science*, Vol. 22(1), 69-77.
- [38] Kubinyi H., Hamprecht F.A. & Mietzner T., 1998. Three- dimensional quantitative similarity-activity relationships (3D QSAR ) from SEAL similarity matrices, *Journal of Medicinal Chemistry*, Vol.41(14), 2553-2564.
- [39] Golbraikh A. & Tropsha A., 2002.Beware of  $q^2$ , *Journal of Molecular.Graphics*, Vol.20(4), 269-276.
- [40] Randic M., Kleiner A.F. & De Alba L.M., 1994. Distance Matrices, *Journal of Chemical. Information. Computer Sciences*, Vol. 34(2), 277-286.
- [41] Randic M. & Krilov G.,1999.On a characterization of the Folding of Proteins, *International Journal of Quantum Chemistry*, Vol. 75(6), 1017-1026.
- [42 ] Carbö-Dorca R., Robert D., Amat Li., Gironés X. & Besalu E., 2000. Molecular Quantum Similarity in QSAR and Drug Design. Springer-Verlag Berlin Heidelberg. Vol. 42, 123p.