QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP (QSAR). V. ANALYSIS OF THE TOXICITY OF ALIPHATIC ESTERS BY MEANS OF MOLECULAR COMPRESSIBILITY DESCRIPTORS

VICENȚIU VLAIA[1*] TUDOR OLARIU[1], LAVINIA VLAIA[2], MARIUS BUTUR[3], CIPRIAN CIUBOTARIU[4], MIHAI MEDELEANU[5], DAN CIUBOTARIU[1]

University of Medicine and Pharmacy "Victor Babeş", Faculty of Pharmacy
[1]Department of Organic Chemistry,
[2]Department of Pharmaceutical Technology,
[3]Department of Mathematics and Biostatistics,
[4]Technical University "Politehnica" Timisoara, Department of Computer Sciences
[5]Technical University "Politehnica" Timisoara, Department of Organic Chemistry
[*]corresponding author: vlaiav@yahoo.com

Abstract

The present work represents an attempt to correlate toxicity of aliphatic esters using calculated molecular descriptors based on properties of the molecular van der Waals (vdW) space considered compressible in some extent. Assuming that a molecule can be compressed from the greatest sphere, corresponding to the vdW surface S^w, to the smallest sphere, corresponding to the vdW volume, V^w, we developed three ***i***-dimensional structural parameters C_{iD}, i=1,2,3, which are measures of compressibility in ***i***D molecular vdW space.

C_{iD} compressibility measures of the vdW molecular space were tested with good results on a series of 56 aliphatic esters exhibiting toxicity to ciliate protozoan *Tetrahymena pyriformis*. A reduced series of 49 esters was obtained by outlier statistical analysis. The quality and the robustness of all the QSAR models were analyzed. For the QSAR model wherein C_{2D} was used as predictor variable, the external cross-validation was performed. The goodness of fit and the predictive power of the QSAR models developed in this study show that accurately describing the toxicity of aliphatic esters with a single correlation equation is quite feasible if the physical meaning of the molecular descriptors is clear.

Rezumat

Această lucrare reprezintă o tentativă de a corela toxicitatea esterilor alifatici utilizând descriptori moleculari calculați pe baza spațiului molecular van der Waals (vdW), considerat, într-o oarecare măsură, compresibil. Presupunând că o moleculă poate fi comprimată de la sfera cea mai mare, care corespunde suprafeței vdW, notată S^w, la sfera cea mai mică, care corespunde volumul vdW, notat V^w, am dezvoltat trei parametri structurali ***i***-dimensionali, notați C_{iD}, i=1,2,3, care sunt măsuri ale compresibilității în spațiul molecular vdW ***i***-dimensional. Măsurile de compresibilitate ale spațiului molecular

vdW, C_{iD}, au fost testate cu bune rezultate pe o serie de 56 de esteri alifatici care îşi manifestă toxicitatea asupra protozoarului ciliat *Tetrahymena pyriformis*. O serie redusă de 49 de esteri a fost obținută în urma analizei statistice a punctelor care se abat semnificativ de la dreapta de regresie (autlaier: „outliers"). Au fost analizate calitatea şi robustețea tuturor modelelor QSAR. Pentru modelul QSAR în care s-a utilizat variabila predictor C_{2D}, a fost făcută validarea externă. Capacitatea de a reproduce datele experimentale (capacitatea de fitare) şi puterea predictivă a modelelor QSAR obținute în acest studiu demonstrează că descrierea cu acuratețe a toxicității esterilor alifatici cu o singură (unică) ecuație corelațională este într-adevăr posibilă dacă semnificația fizică a descriptorilor moleculari este clară.

Keywords: QSAR, toxicity, *Tetrahymena pyriformis*, aliphatic esters, compressibility molecular descriptors (C_{iD}), cross-validation (CV).

Introduction

The quantitative structure-activity relationship (QSAR) is an important research field in toxicology, which deals with the prediction of the toxicities of new compounds using mathematical relationships based on structural and physicochemical properties of previously tested chemical compounds, usually on microorganisms, algae, and fishes. QSAR in toxicology may be seen as an environmental approach that has been efficiently used for the study of toxicity mechanisms of various reactive chemicals. In fact, this is a powerful quantitative technique, which makes an attempt to relate the variations in biological activity as a function of the changes in molecular properties. The QSAR method attempts to link toxicity data with molecular descriptors derived from physicochemical properties or theoretical models of the molecular structure of chemical compounds [1].

The toxicity of aliphatic esters results from interaction between the toxicant molecule and its biological targets. They exhibit a narcosis mode of toxic action, producing a non-covalent and reversible alteration at the site of action – lipid and/or protein components within biological membranes [2,3]. Interactions with receptors are typically a non-covalent "lock-and-key-type" interface. Such exchanges need 3D conformational requirements for binding/activation, which are governed by stereo electronic molecular properties [4].

High accuracy in prediction of toxicity is impossible because the toxic activity is defined not only by the molecular structure of toxin, but also depends on the characteristics of the living organisms. Nevertheless, predictive models for the toxicity are very useful [3,5,6]. The majority of the applied models were based on the octanol/water partition coefficient (log P

or log K_{ow}), and on molecular structural descriptors developed by means of the mathematical theory of graphs [7-10].

In this study we decided to use the compressibility descriptors developed on the basis of the molecular van der Waals (vdW) space [11-15] and multiple linear regression (MLR) techniques to build and validate QSARs modeling the toxic activity of a series of 56 aliphatic esters, which exhibit toxic effects on the ciliate protozoan *Tetrahymena pyriformis*. Assuming that a molecule can be characterized by two spheres corresponding, respectively, to its vdW volume (V^W) and surface (S^W), we developed three compressibility measures of the molecular vdW space, denoted by C_{iD}, where iD (i=1,2,3) represents the dimensionality of the vdW space [15]. Thus, we supposed that a molecule can be compressed with a specific quantity C_{iD} from the greater sphere, corresponding to S^W, to a smaller one, corresponding to V^W. The values of C_{iD} structural parameters can be easily computed, and their physical meaning is clear. The good correlation results prove that the vdW molecular compressibility descriptors are valuable tools for modeling the toxicity of chemical compounds, the aliphatic esters in this case.

Van der Waals measures of molecular compressibility

Generally, a molecule M can be viewed as a solid into Cartesian-3D space. A molecular van der Waals envelope, Γ, can be defined in the "hard-spheres" approximation as the external surface resulted from the overlapping of all vdW spheres corresponding to the atoms of M. The points (x,y,z) inside the envelope Γ satisfy at least one of the following inequalities:

$$(X_i - x)^2 + (Y_i - y)^2 + (Z_i - z)^2 \le (r_i^w)^2 \quad , \; i = \overline{1,m} \qquad (1)$$

In relation (1) *m* represents the number of atoms in a given M molecule and X_i, Y_i, Z_i are the Cartesian coordinates of *i* atom. Consequently, the total volume embedded by the envelope Γ is the molecular vdW volume of M, noted by V^w. Obviously, this envelope is a surface, and the area of this surface is noted by S^w. For calculating V^W and S^W we developed some algorithms based on the Monte Carlo method [16,17].

Assuming that a molecule can be characterized by two spheres, corresponding to the vdW volume, V^w, and vdW surface, S^w, respectively, we developed three compressibility measures of molecular vdW space, C_{iD}, i=1,3. This hypothesis is based on the known conformational flexibility of the molecules and on the fact that the molecules are relatively compressible [13,14]. Therefore, a molecule should be compressed from the greatest sphere

(S_G), corresponding to the vdW surface area of Γ equal to S^w, to the smallest sphere (S_S), concordant to the vdW volume embedded by Γ, equal to V^w [15].

The vdW radius, noted r_S^w, and the vdW volume, V_S^w, of the molecular S_G sphere are calculated as follows,

$$r_S^w = [S^w / 4\pi]^{1/2} \quad (2)$$

$$V_S^w = 4\pi(r_S^w)^3 / 3 \quad (3)$$

The molecular S_G sphere can be compressed to a molecular S_S sphere, which has a volume equal to the molecular van der Waals volume, V^w. The radius, r_V^w, and the surface area S_V^w of this molecular S_S sphere, are calculated with the following relations:

$$r_V^w = [3V^w / 4\pi]^{1/3} \quad (4)$$

$$S_V^w = 4\pi(r_V^w)^2 \quad (5)$$

Thus, the following two triplets characterize the molecular S_G sphere and the molecular S_S sphere:

$$\{S_G\}: \left(r_S^w, S^w, V_S^w\right) \quad (6)$$

$$\{S_S\}: \left(r_V^w, S_V^w, V^w\right) \quad (7)$$

The molecular vdW compressibility measures, C_{iD}, can be easily defined from the triplets (6) and (7), as the difference between the values of vdW radius, surface area, and volume of S_G sphere and S_S sphere, as follows,

$$C_{1D} = r_S^w - r_V^w \quad (8)$$

$$C_{2D} = S^w - S_V^w \quad (9)$$

$$C_{3D} = V_S^w - V^w \quad (10)$$

The figure 1 shows the 3D physical model of these molecular compressibility descriptors. One may easily observe the outer sphere, corresponding to S^W (with its radius, r_S^w), and the inner sphere, which corresponds to V^w (with its radius, r_V^w). A molecule can be compressed with the quantities C_{1D}, C_{2D}, and C_{3D}, measured in the one-(1D), bi-(2D), and three-dimensional (3D) vdW space, respectively. The compressibility capacity may be important in the interaction of molecules with their biological targets. In such cases, these vdW molecular descriptors can be used with good results in QSARs in order to model the biological activity of the chemical compounds.

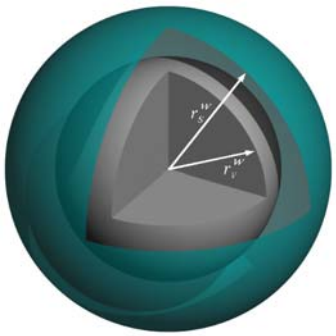

Figure 1
The physical model (3D) of the molecular compressibility measures C_{iD}, i=1,2,3.

Materials and methods

Toxicity Data

Reliable data are required to build reliable predictive QSAR models. In terms of toxicity, such data should ideally be measured by a single protocol, even the same laboratory and by the same workers. High quality biological data will have lower experimental error associated with them. Biological data should ideally be from well-standardized assays, with a clear endpoint [5].

The population growth impairment is one with significance in the development and validation of QSARs in aquatic toxicity to the ciliated protozoan *Tetrahymena pyriformis*. This data set, TETRaTOX [18], lists relative toxicity on a wide variety of industrial chemicals including aliphatic esters [2]. These data originated from a single laboratory using a standard protocol [19] which has been validated [20].

The data set used here was collected from literature [2,18]. Each ester was tested in three replicate assays to the ciliate *T. pyriformis*. Each test replicate consisted of six to ten different concentrations. The reported toxicological activity was the 50% growth inhibitory concentration (IGC_{50}), expressed in millimols [2,18].

We used as experimental biological activity, A, the logarithm of the inverse of concentration that produces 50% growth inhibition to *T. pyriformis*. The values of $A=Log(1/IGC_{50})$ for a series of 56 esters used in this QSAR study are presented in Table I.

Molecular Descriptors

The molecular compressibility descriptors C_{iD}, i=1,2,3 were used in this QSAR study. They were evaluated as described in the above section by relations (8), (9), and (10), corresponding to one-, bi-, and tri-dimensional compressibility measures of molecular vdW space – C_{1D}, C_{2D}, and C_{3D},

respectively. The vdW volume (V^w) and the area of vdW surface (S^w) were calculated with in house algorithm developed on the basis of the *Monte Carlo* method [21]. The geometry of molecules was optimized with MM+ and AM1 algorithms from the HyperChem software package. The values of C_{1D}, C_{2D}, and C_{3D} compressibility descriptors for the series of 56 esters are systematized in Table I.

Table I

The data used in QSAR analysis. The toxic activity of esters to *Tetrahymena pyriformis* is expressed by A=log(1/IGC$_{50}$), where IGC$_{50}$ is the 50% inhibition growth concentration (in mM). Molecular descriptors used in correlations are 1-octanol/water partition coefficient (log K_{OW}) and compressibility molecular descriptors C_{1D} (in Å), C_{2D} (in Å^2), and C_{3D} (in Å^3)

No.	ester name	A	C_{1D}	C_{2D}	C_{3D}
1	decyl acetate	1.8794	1.133	123.73	270.27
2	methyl undecanoate	1.4248	1.126	122.74	267.62
3	methyl decanoate	1.3778	1.039	109.60	231.27
4	octyl acetate	1.0570	0.964	98.39	200.68
5	vinyl 2-ethylhexanoate	1.0462	0.890	89.00	177.88
6	methyl nonanoate	1.0419	0.956	97.40	198.29
7	allyl heptanoate	0.7282	0.922	92.74	186.44
8	methyl octanoate	0.5358	0.868	85.08	166.64
9	butyl butyrate	0.5157	0.782	73.67	138.51
10	allyl hexanoate	0.2128	0.828	80.08	154.63
11	butyl propionate	0.1704	0.694	62.38	111.83
12	amyl acetate	0.1625	0.699	62.88	112.81
13	methyl heptanoate	0.1039	0.775	72.87	136.77
14	ethyl hexanoate	0.0637	0.775	72.85	136.77
15	propyl valerate	0.0094	0.766	71.85	134.63
16	hexyl acetate	-0.0087	0.789	74.32	139.74
17	amyl propionate	-0.0431	0.777	73.05	137.13
18	2-ethylbutyl acetate	-0.1202	0.724	67.66	126.10
19	ethyl valerate	-0.3580	0.692	62.21	111.56
20	n-hexyl formate	-0.3824	0.705	63.47	114.06
21	vinyl butyrate	-0.3825	0.546	45.60	75.89
22	tert butyl propionate	-0.4095	0.632	56.19	99.69
23	propyl butyrate	-0.4138	0.683	61.27	109.72
24	butyl acetate	-0.4864	0.602	51.35	87.41
25	isopropenyl acetate*,#,@	-0.4892	0.432	33.65	52.30
26	ethyl butyrate	-0.4903	0.596	50.91	86.66
27	methyl hexanoate	-0.5611	0.692	62.21	111.59
28	isobutyl isobutyrate*,#	-0.5908	0.757	70.96	132.80
29	allyl butyrate	-0.6355	0.647	57.07	100.47

(continued)

Table I (continued)

30	vinyl propionate*,#	-0.6530	0.448	35.13	54.89
31	propargyl propionate	-0.6554	0.512	41.95	68.46
32	sec-butyl acetate	-0.6794	0.569	48.41	82.08
33	isobutyl propionate	-0.6935	0.683	61.31	109.81
34	ethyl isovalerate	-0.7231	0.674	60.40	108.06
35	n-amyl formate	-0.7826	0.609	52.10	88.86
36	propyl propionate	-0.8148	0.600	51.19	87.11
37	methyl valerate	-0.8448	0.541	45.72	77.00
38	vinyl acetate*,#	-0.8595	0.353	25.73	37.34
39	allyl propionate	-0.8791	0.546	45.55	75.72
40	2-butynyl-acetate	-0.8834	0.518	42.37	69.14
41	ethyl-2-methylbutyrate*,#	-0.8893	0.667	59.73	106.66
42	butyl formate	-0.9336	0.513	41.35	66.47
43	ethyl propionate	-0.9450	0.501	40.27	64.57
44	propyl formate	-1.0221	0.415	31.25	46.90
45	methyl-2-methylbutyrate	-1.1650	0.569	48.31	81.87
46	propargyl acetate	-1.1664	0.417	31.95	48.82
47	propyl acetate	-1.2382	0.502	40.42	64.84
48	methyl butyrate	-1.2463	0.492	39.49	63.17
49	ethyl isobutirate*,#	-1.2709	0.592	50.51	85.92
50	ethyl acetate	-1.2968	0.408	30.63	45.85
51	isobutyl formate	-1.3081	0.500	40.25	64.56
52	tert butyl formate	-1.3719	0.460	36.78	58.56
53	methyl formate*	-1.4982	0.214	13.46	16.80
54	isopropil acetate#	-1.5900	0.486	39.05	62.47
55	methyl acetate	-1.5954	0.306	21.17	29.19
56	methyl propionate	-1.6092	0.402	30.20	45.17

#outliers when using C_{1D} (eq. 11), *C_{2D} (eq. 12), and @C_{3D} as predictor variables

Statistical Analysis

All QSAR calculations were performed using the MobyDigs software [22], using the linear regression and the LOO (leave-one-out method), bootstrapping and external cross validation methods. Log $(IGC_{50})^{-1}$ was used as independent variable, and C_{1D}, C_{2D}, and C_{3D} compressibility measures acted as dependent (predictor) variables (see the values of C_{iD} in Table I).

The resulted QSAR models were measured to fit by correlation coefficient (r) and coefficient of determination (r^2), adjusted for the degree of freedom (r^2_{adj}),also called explained variance, (EV). The uncertainty in the model was noted as the standard error (s). The reliability in the model was

expressed by the F (Fisher) and t (Student) statistics. The t-test was used to determine the 95% confidence limits of the obtained QSAR models. Statistical fit should not be confused with the ability of a model to make predictions. Therefore, we used the leave-one-out (LOO), bootstrapping, and a leave-half-out (LHO) cross-validation (CV) method to estimate the predictive ability of the obtained QSAR model, using the corresponding CV coefficient (coefficient of predictions), q^2, and the standard deviation error in prediction (SDEP).

Results and discussion

The aliphatic ester compounds were sorted by activity $A=Log(IGC_{50})^{-1}$, in its decreasing order (see Table I). Toxicity values varied roughly 3.5 orders of magnitude on a logarithmic scale (from 1.879 to -1.609). The molecular compressibility indices were calculated using relations (2)-(5) and (8)-(10). These descriptors exhibited a linear variation on each dimension of the molecular vdW space in which they were estimated. C_{1D}, C_{2D}, and C_{3D} are measures of the degree of compression along one-dimension (1D), on a surface (2D), and of a volume (3D), respectively. The range of variation for C_{1D} was from 1.133Å to 0.214Å; C_{2D} varied from 123.73Å^2 to 13.46Å^2 and C_{3D} from 270.27Å^3 to 16.80Å^3. One can observe from Table I that the degree of compressibility increases when the molecular size increases. The variation of molecular shape seems also passible to be measured by the compressibility descriptors, but the analysis of this subject is in progress.

The linear QSAR models obtained by correlating toxicity (A) *versus* compressibility descriptors (C_{iD}, i=1-3) are the following:

$$\hat{A} = -2.8957(\pm 0.1520) + 3.8486(\pm 0.2263)\cdot C_{1D}$$
$$n = 56, s = 0.330; r = 0.918; r^2_{adj} = 0.840; F = 289 \quad (11)$$

$$\hat{A} = -2.2754(\pm 0.1075) + 0.0319(\pm 0.0017)\cdot C_{2D}$$
$$n = 56, s = 0.305, r = 0.930, r^2_{adj} = 0.863; F = 347 \quad (12)$$

$$\hat{A} = -1.8848(\pm 0.0854) + 0.0138(\pm 0.0007)\cdot C_{3D}$$
$$n = 56, s = 0.296, r = 0.935, r^2_{adj} = 0.871, F = 373.5 \quad (13)$$

where $\hat{A}$ stands for the calculated value of experimental inhibitory activity with the QSAR models (11)-(13), n represents the number of data, s is the standard error, r stands for correlation coefficient and r^2_{adj} is the

coefficient of determination adjusted for the degree of freedom. The statistical tests F and t are used at the 95% reliability degree.

The goodness of fit of the QSARs (11)–(13) is satisfactory, as expressed by the values of r, r^2_{adj}, s, and F statistics and by the 95% confidence limits of the linear model parameters. One can see that the obtained results are similar for all compressibility descriptors; for example, the explained variances of A values are 84% (for C_{1D}), 86% (for C_{2D}), and 87% (for C_{3D}). The reliability in the all QSAR models is very close – see the values of the Fisher test, F, and the confidence limits.

The predictive ability of the above models was estimated by means of CV leave-one-out (LOO) and bootstrapping (BOOT) methods, using the coefficients of prediction, q^2, the standard deviation error in calculation, SDEC, and standard deviation error in prediction, SDEP. The values of these statistical indicators are systematized in Table II.

Table II

Values of the statistics used to asses the predictive power of the QSAR models A vs. compressibility descriptors (CDs).[#]

QSARs	CDs	q^2_{LOO}	q^2_{BOOT}	SDEP	SDEC	r^2_{Y-s}	q^2_{Y-s}
(11)	C_{1D}	0.829	0.831	0.337	0.324	-0.025	-0.104
(12)	C_{2D}	0.856	0.859	0.310	0.300	-0.029	-0.106
(13)	C_{3D}	0.865	0.868	0.300	0.290	-0.027	-0.106

[#] q^2_{LOO} - coefficient of prediction obtained by LOO-CV method; q^2_{BOOT} - coefficient of prediction obtained by Bootstrapping-CV procedure; the subscript "Y-s" refers to the Y-scrambling technique; for the significance of SDEP and SDEC, see the text.

The predictive capacity (or goodness of prediction) of these models is good, taking into account the commonly accepted values for a satisfactory QSAR model, $q^2 > 0.500$. CV techniques allowed the assessment of internal prediction, in addition to the robustness (stability of QSAR model parameters). Bootstrapping simulates what would happen if the population was resampled by randomly resampling the data set from Table I. The risk of chance correlation was verified by Y-scrambling procedure, in which the dependent variable A (toxic activities of esters on *Tetrahymena pyriformis,* $\log IGC_{50}^{-1}$) was randomly shuffled and a new QSAR model was developed using the C_{iD}, i=1,2,3 independent variables. The process was repeated several times and the resulting QSAR models are the expected low r^2_{Y-s} and low q^2_{Y-s} values.

Outliers in QSAR model generation present their own problems. If they are not well fit by the model (off by more than 2 standard deviations), they should be dropped from the data set. Their aberrant behavior may be attributed to inaccuracies in the testing procedure or unusual actions. In this work, the presence of outliers was based on *a posteriori* examination of the prediction error, by comparing predicted and observed responses for each ester from Table I. The outliers showing exceptional prediction errors – with a CV standardized residual greater than two standard deviation units – were removed from QSARs (11), (12), and (13), in order to improve the model's reliability. The results of our analysis are summarized in Table III.

Table III

The improved statistical results obtained by elimination of outliers from QSAR models (11), (12), and (13).*

QSAR	CD	r	r^2_{adj}	s	F	q^2_{LOO}	q^2_{BOOT}	SDEP	SDEC	r^2_{Y-s}	q^2_{Y-s}
(14)	C_{1D}	0.959	0.918	0.244	535	0.913	0.915	0.249	0.239	-0.037	-0.126
(15)	C_{2D}	0.962	0.924	0.233	586	0.919	0.922	0.239	0.229	-0.041	-0.129
(16)	C_{3D}	0.941	0.884	0.283	413	0.878	0.880	0.288	0.278	-0.024	-0.104

* The meaning of the statistical indicators is the same as in Table II.

The best QSAR models were obtained when using C_{1D}, and C_{2D} as predictor variables, slightly better for C_{2D}. Therefore, we present below only the external validation of the model (15).

The external validation procedure consists in splitting the available data set (from Table I) into a training (calibration) set, used to develop the QSAR model, and a validation (test) set, used only for predictions [23]. At this point it is important to ensure that both the training and the validation sets overlap the whole descriptor space occupied by the entire data set and the chemical domain in two data sets is not too different. For that reason an ideal splitting leads to a validation set in which each of its members is close to at least one point of the training set [24].

To assess the predictive ability of the statistical QSAR model (15), the data set sorted by toxicity values in decreasing order (outliers were omitted – see Table I) was splitted into test set and training set, assigning the compounds alternately to test set and training set, and *vice versa*. Thus, 50% of the compounds were used for training and 50% for testing. That is, the QSAR model obtained for the set composed of odd ranking compounds was used to calculate toxic activities, of the pair ranking subset, and, inversely, the QSAR model developed for the pair ranking subset was used to estimate the toxic activities of the compounds belonging to odd ranking

subset. The procedure described above will be referred below as the Leave odd-pair Out (Lo-pO) cross-validation technique (Lo-pO-CV) [16,25,26].

When we used the odd ranking subset as a training set we obtained the following QSAR model:

$$\hat{A} = -2.3416(\pm 0.1295) + 0.0339(\pm 0.0020) \cdot C_{2D}$$
$$n = 25, s = 0.252, r = 0.962, r^2_{adj} = 0.922; F = 283 \qquad (17)$$

If the pair ranking subset was used for training, the QSAR model was as follows,

$$\hat{A} = -2.3594(\pm 0.1210) + 0.0328(\pm 0.0019) \cdot C_{2D}$$
$$n = 24, s = 0.213, r = 0.966, r^2_{adj} = 0.931; F = 310 \qquad (18)$$

Obviously, the QSAR equations (17) and (18) were used to predict the toxicity of the esters in test sets – the pair ranking subset, and the odd ranking subset, respectively.

The results obtained by application of the Lo-pO-CV technique are given in Table IV.

Table IV

Values of the statistics used to asses the predictive power of the QSAR models (17) and (18).*

QSAR	q^2_{LOO}	q^2_{BOOT}	q^2_{ext}	SDEP	SDEC	r^2_{Y-s}	q^2_{Y-s}
(17)	0.912	0.916	0.921	0.261	0.242	-0.033	-0.242
(18)	0.920	0.926	0.914	0.224	0.204	-0.040	-0.221

* q^2_{ext} is the coefficient of prediction in external validation, obtained by comparing the square error of predicted test set toxicities with the variance of experimental training set toxicities.

The Lo-pO-CV procedure used in this work can be considered a pseudorandom division because the actual values of activities, A, are scattered by measurement errors. The method has the advantage that the activity distribution of corresponding training sets and test sets are very similar, and it should allow assessing the ability of the model to interpolate [26].

Conclusions

We presented here three molecular descriptors, C_{iD}, i=1,3, developed on the basis of molecular vdW space supposed isotropic, homogeneous, and compressible in some extent. C_{iD} measure the i-dimensional relation between the packed and the extended vdW size of a molecule within its environment, during physical and chemical interaction.

It has been shown in this study that accurately describing the toxicity of aliphatic esters with a single correlation is quite feasible if the molecular descriptors have clear physical meaning, and they are related to the physical and chemical interacting forces among molecules, and particularly between biological receptors of biological membranes and the toxin molecules. The reason is that the mechanism of action of esters, i.e. a reversible accumulation of the ester within the cell membrane that results in distortion and disruption of function, is well modeled by the compressibility descriptors.

References

1. Toropov A. A., Rasulev B. F. and Leszczynski J, QSAR modeling of acute toxicity by balance of correlations, *Bioorg. Med. Chem.* 2008, 16, 5999-6008.
2. Schultz T. W., Cronin M. T. D., Netzeva T. I. and Aptula, A. O., Structure – Toxicity Relationships for Aliphatic Chemicals Evaluated with *Tetrahymena pyriformis*. *Chem. Res. Toxicol.* 2002, 15, 1602-1609.
3. Schultz T. W., Cronin M. T. D., Walker J. D., and Aptula A. O., Quantitative structure – activity relationships (QSARs) in toxicology: a historical perspective. *J. Mol. Struct. (THEOCHEM)* 2003, 622, 1-22.
4. Crisan O., Oprean Luminita, Marchand J., Influence de l'isomerie sur les effets toxiques, *Farmacia* 2007, LV, 2, 144-156.
5. Cronin M. T. D., Schultz T. W., Pitfalls in QSAR, *J. Mol. Struct. (THEOCHEM)* 2003, 622, 39-51.
6. Schultz T. W., Cronin M. T. D., Netzeva T. I., The present status of QSAR in toxicology. *J. Mol. Struct. (THEOCHEM)* 2003, 622, 23-38.
7. Randić M., Basak S. C., On the Use of the Variable Connectivity Index $^{1}\chi^{f}$ in QSAR: Toxicity of Aliphatic Ethers, *J. Chem. Inf. Comput. Sci.* 2001, 41, 614-618.
8. Hawkins D. M., Basak S. C., Kraker J., Geiss K. T. and Witzmann F. A., Combining Chemodescriptors and Biodescriptors in Quantitative Structure – Activity Relationship Modeling, *J. Chem Inf. Model,*. 2006, 46, 9-16.
9. Vlaia V., Olariu T., Ciubotariu C., Medeleanu M., Dragos D., Ursica L., Ciubotariu C., Quantitative Structure-Activity Relationship (QSAR). III. Generalized Topological Descriptors Based on Reciprocal Distance Matrix for development of toxicological QSARs, *Farmacia,* 2007, LV, 3, 284-296.
10. Toropov A. A. and Schultz T. W., Prediction of Aquatic Toxicity: Use of Optimization of Correlation Weights of Local Graph Invariant. *J. Chem. Inf. Comput. Sci,*. 2003, 43, 560-567.
11. Ciubotariu D., Gogonea V., Medeleanu M., Van der Waals Molecular Descriptors. Minimal Steric Difference, in QSPR/QSAR Studies by Molecular Descriptors, M. V. Diudea (Ed.), Nova Science Publishers, Inc., Huntington, New York, 2001, 281-361.
12. Ciubotariu D., Medeleanu M., Vlaia V., Olariu T., Ciubotariu C., Dragos D. and Seiman C., Molecular van der Waals Space and Topological Indices from The Distance Matrix, *Molecules*, 2004, 9, 1053-1078.
13. Medeleanu M., Dragos D., Olariu T., Vlaia V., Ciubotariu C., Seiman C., and Ciubotariu D., Molecular Descriptors in QSPR/QSAR Studies. I. Molecular van der Waals Space, *Chem. Bull. "Politehnica" Univ. Timisoara*, 2004, 49, 8-13.

14. Chiriac A., Ciubotariu D., Timofei-Funar S., Kurunczi L., Mracec M., Szabadai Z., Seclaman E., Simon Z., QSAR and 3D-QSAR in Timisoara. 1972-2005, *Rev. Roum. Chim*., 2006, 51, 79-99.
15. Ciubotariu D., Vlaia V., Olariu T., Ciubotariu C., Medeleanu M., Ursica L., Dragos D., Molecular van der Waals Descriptors for Quantitative Treatment of Toxicological Effects, 12th International Workshop on Quantitative Structure–Activity Relationships in Environmental Toxicology, 8-12 May 2006, Lyon, France, 90.
16. Ciubotariu D., Structure – Reactivity Relationships in the Series of Carbonic Acid Derivatives, PhD Thesis (in Romanian language), Polytechnic Institute of Bucharest, 1987.
17. Niculescu-Duvaz I., Ciubotariu D., Simon Z. and Voiculetz N., Quantitative Approach to Carcinogenesis Inhibition: a QSAR Analysis, in Modeling of Cancer Genesis and Prevention, N. Voiculetz, A. T. Balaban, I. Niculescu-Duvaz and Z. Simon (Eds.), CRC Press, Boca Raton, 1991, 157-214.
18. http://www.vet.utk.edu/TETRaTOX/
19. Schultz T. W., TETRATOX: The Tetrahymena pyriformis population growth impairment endpoint. A surrogate for fish lethality. *Toxicol. Methods,* 1997, *7*, 289-309.
20. Larsen J., Schultz T. W., Rasmussen L., Hoofman R., Pauli W., Progress in an ecotoxicological standard protocol with protozoa: Results from a pilot ring test with *Tetrahymena pyriformis*, *Chemosphere,* 1997, 35, 1023-1041.
21. Ciubotariu C., Medeleanu M., Ciubotariu D., IRS – a Computer Program Package for QSAR and QSPR Studies, *Chem. Bull. "Politehnica" Univ. Timisoara*, 2006, 51, 13-16.
22. MobyDigs software package
23. Konovalov D. A., Llewellyn L. E., Heyden Y. V. and Coomans D., Robust Cross-Validation of Linear Regression QSAR Models. *J. Chem. Inf. Model.* 2008, 48, 2081-2094.
24. Tropsha A, Gramatica P., Gombar V. K., The importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models, *QSAR Comb. Sci.* 2003, 22, 69-77.
25. Ciubotariu D., Deretey E., Oprea T. J., Sulea T., Simon Z., Kurunczi L., Chiriac A., Multiconformational Minimal Steric Difference. Structure – Acetylcholinesterase Hydrolysis Rates Relations for Acetic Acid Esters. *Quant. Struct.-Act. Relat.,* 1993, 12, 367-372.
26. Gedeck P., Rohde B., Bartels C. QSAR – How Good Is It in Practice? Comparison of Descriptor Sets on an Unbiased Cross Section of Corporate Data Sets. *J. Chem. Inf. Model.,* 2006, 46, 1924-1936.

Manuscript received: 30.04.2009