QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP (QSAR). VI. MODELING THE TOXICITY OF ALIPHATIC ESTERS BY MEANS OF MOLECULAR OVALITY DESCRIPTORS

TUDOR OLARIU[1], VICENŢIU VLAIA[1*] , LAVINIA VLAIA[2], CIPRIAN CIUBOTARIU[3], DAN CIUBOTARIU[1]

[1]Laboratory of Organic Chemistry, University of Medicine and Pharmacy "Victor Babeş", Faculty of Pharmacy, Eftimie Murgu No. 2, Timisoara
[2]Laboratory of Pharmaceutical Technology, University of Medicine and Pharmacy "Victor Babeş", Faculty of Pharmacy, Eftimie Murgu No. 2, Timisoara, Romania
[3]Technical University "Politehnica" Timisoara, Department of Computer Sciences, Romania
**corresponding author: vlaiav@umft.ro*

Abstract

Quantitative structure – activity relationships were developed for the toxicity of 56 aliphatic esters to the protozoan ciliate *Tetrahymena pyriformis*. The toxicity was measured as A=Log(1/IGC$_{50}$), where IGC$_{50}$ is the concentration which inhibits a 50% growth of *T. pyriformis*. The ovality van der Waals descriptors of molecular shape, Θ_{iD}, *i=1,2,3* were used as predictor variables. They were calculated as the ratios of the radii (Θ_{1D}), surfaces (Θ_{2D}), and volumes (Θ_{3D}) of the greatest molecular sphere, corresponding to the vdW surface area of a molecule, and those of the smallest molecular sphere, corresponding to the vdW volume of the same molecule. The goodness of fit was estimated by the coefficient of determination adjusted for the degree of freedom (r^2_{adj} > 0.805 for all three Θ_{iD} shape descriptors) and the predictive ability by bootstrapping (q^2_{BOOT} > 0.777) and LOO (q^2_{LOO} > 0.789) cross-validation statistical procedures. The best model was obtained for the Θ_{3D} predictor variable: *r=0.911,* $r^2_{adj} = 0.826$, $q^2_{BOOT} = 0.804$ $q^2_{LOO} = 0.968$. An external cross-validation procedure based on odd-even series was also applied with good predictive results (q^2 > 0.728). The ovality molecular descriptors Θ_{iD}, i=1,2,3 can be easily calculated for any molecule and their physical meaning is clear.

Rezumat

În prezenta lucrare au fost dezvoltate relaţii cantitative structură chimică-activitate biologică pentru cuantificarea toxicităţii unei serii de 56 esteri alifatici asupra ciliatului *Tetrahymena pyriformis.* Toxicitatea este exprimată ca A=Log(1/IGC$_{50}$), unde IGC$_{50}$ reprezintă concentraţia ce inhibă cu 50% creşterea *T. pyriformis*. Ca variabile predictor au fost utilizaţi descriptorii de ovalitate van der Waals ai formei moleculare Θ_{iD}, *i=1,2,3* . Aceştia au fost calculaţi ca rapoarte de raze (Θ_{1D}), suprafeţe (Θ_{2D}), şi volume

(Θ_{3D}) a celei mai mari sfere moleculare, corespunzătoare suprafeţei moleculare vdW a unei molecule, şi cele ale celei mai mici sfere moleculare, corespunzătoare volumului vdW ale aceleiaşi molecule. Capacitatea de fitare a fost exprimată cu ajutorul coeficientului de determinare ajustat pentru gradele de libertate (r^2_{adj} > 0.805 pentru toţi cei trei descriptori ai formei moleculare Θ_{iD}), iar abilitatea predictivă prin procedee statistice de validare încrucişată: *bootstrapping* (q^2_{BOOT} > 0.777) şi LOO (q^2_{LOO} > 0.789). Cel mai bun model a fost obţinut pentru variabila predictor Θ_{3D}: $r=0.911$, $r^2_{adj} = 0.826$, $q^2_{BOOT} = 0.804$ $q^2_{LOO} = 0.968$. A fost aplicat un procedeu de validare încrucişată externă bazat pe serii par-impar, cu rezultate predictive bune (q^2 > 0.728). Descriptorii moleculari de ovalitate Θ_{iD}, i=1,2,3 pot fi calculaţi uşor pentru orice moleculă, iar semnificaţia lor fizică este clară.

Keywords: aliphatic esters, ovality descriptors, *Tetrahymena pyriformis*, QSAR

Introduction

Quantitative structure – activity relationship (QSAR) searches information relating chemical structure to biological and other activities by developing a QSAR model. Using such an approach one could predict the activities of newly designed compounds before a decision is being made whether these compounds should be synthesized and tested [1].

QSAR studies are important tools in environmental and industrial risks assessment allowing the analysis of toxicology data. Since it was noticed that the properties of compounds depend on their structure, QSARs have been used in elucidating the specific mechanisms underlying the toxic effects. At present, predictive QSARs have been recognized by the regulatory authorities as an affordable and safe alternative for toxicological measurements [2].

For the assessment of the environmental impact of toxicants, the unicellular ciliated protozoan, *Tetrahymena pyriformis*, is attractive for its fast growth rates and inexpensive assays. The testing method has been carefully established giving the assurance of very high quality to the data produced. In addition to the environmental safety, toxicity data to this organism have proven useful in estimation of the toxic potencies of compounds to other aquatic organisms [3-7].

Several QSAR models predicting acute chemical toxicity for aquatic environments have been published [3,5,6]. They are based mainly on the logarithm of the octanol-water coefficient (logP, also referred to as $\log K_{ow}$) as this hydrophobicity term reproduces the ability of a substance to enter cells through the lipid membranes and indicates both toxicant uptake and baseline toxicity. Nevertheless, the experimental determination of logP can

be a complex matter, and experimental values can differ greatly even when referred to the same compound [8]. Thus, several approaches have been developed for the theoretical calculation of logP [9-11] but also in these calculations it is not uncommon to have differences of several orders of magnitude.

In recent years, we used with good results various van der Waals molecular descriptors (vdWMDs) in toxicological QSARs for predicting the toxicity of various organic chemical classes on fishes or microorganisms [12-14]. Our previous work showed that the shape of molecules, as quantified by ovality descriptors, plays an important role in predicting toxicity of alcohols [12,13]. These shape molecular descriptors were developed on the basis of molecular van der Waals space, quantified by the vdW molecular volume, V^W, and surface, S^W [15]. They describe the more or less spherical form of organic molecules, and are easy to calculate for various molecules, irrespective of their complexity.

In this paper we present a QSAR study of the toxicity of aliphatic esters based on the ovality shape molecular descriptors, Θ_{iD}, i=1,2,3. This approach for the development of robust toxicological QSAR is based on the large space of theoretically calculated molecular descriptors. Esters serve a wide variety of purposes, including uses as solvents and starting or intermediate compounds in organic synthetic processes, and are therefore present throughout our environment. Their toxicity results from the interaction between the ester molecule and its biological target: the cellular membranes. Baseline toxicity of esters can be understood as a disruption of the functions of biological membranes, although the detailed mechanism remains unclear. 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 [4,5]. 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. Shape is expected to play an important role in the interaction of esters and their biological target membrane [16-18].

Materials and Methods

Experimental data were taken from a series of 500 aliphatic chemicals that include different structural classes such as esters, saturated alcohols, ketones, nitriles, and sulfur-containing compounds [7]. The toxicity of the esters on the protozoan *Tetrahymena pyriformis* are expressed in terms of inhibitory growth concentration, IGC_{50} (measured

millimolar). Protozoa are real eukaryotic cells and ubiquitous in the aquatic and terrestrial environment. Their normal behavior in nature may be related to the presence of pollutants and to air, soil and water quality. This fact has led toxicologists and ecotoxicologists to use protozoa as test systems for studies on xenobiotics and health risk assessment. Among protozoa, *Tetrahymena pyriformis* is the most commonly ciliated model used for laboratory research and QSAR studies. The *T. pyriformis* toxicity data for various chemicals are also available at the Tetratox database Web site [7].

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

Van der Waals (VdW) ovality as shape molecular descriptor

Shape is a very important molecular feature for describing ligand molecules interacting with a receptor, and, also, other various complex chemical and biological processes. Irrespective of the type of definition used, the essence of shape is very useful in describing a molecule or a part of molecule (substituent). Hence, the study of shape in molecular pharmacology has gained importance owing to its applicability in drug design *in silico* techniques widely employed to decrease the costs of drug discovery and development [13].

Isolated atoms show spherical symmetry, and it is obvious to consider a molecule (in the hard sphere approximation) as a collection of atomic spheres centered in the equilibrium positions of the atomic nuclei; each sphere has a radius equal with its vdW radius, r^W. Because the vdW radii of atomic spheres used for representation of molecular space are usually much too large for modeling molecules by simply placing the atomic hard vdW spheres side by side, commonly one generates various "fused sphere" models for molecules. Positions of atomic spheres may be described by their Cartesian coordinates according to the 3D stereochemical bond pattern of a particular molecule. An envelope, Γ, may be defined as the outer surface of the intersected atomic spheres of M. Γ represents the van der Waals surface of a molecule M, which embeds an associated molecular body with a well defined boundary. Van der Waals surfaces are models based on the above assumption and they are exceptionally useful tools for the approximate representations of molecules [19].

The points (x,y,z) inside the envelope Γ satisfy at least one of the following inequalities:

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

where *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. Obviously, this envelope is a surface.

Therefore, the total volume embedded by this envelope Γ is the molecular vdW volume of M, noted by V^w, and the area of this envelope was noted by S^w. V^W and S^W can be estimated by analytical integration, but the algorithms are prohibitively complicated. Therefore, we developed some methods for calculating V^W and S^W with the aid of Monte Carlo integration methods [19,20].

Molecules are dynamic objects undergoing continuous internal motion. Some finite range of possible deformations with respect to the formal equilibrium shape of the molecule is an inseparable aspect of any realistic molecular model. Consequently, it is important to use techniques for molecular shape characterization which can account for the deformability and the dynamic features of molecular shapes. One must be able to distinguish the essential shape deformations from those having little chemical significance [21].

Therefore, we consider that a molecule, M, can be compressed in a range comprised between its maximum and minimum surface area. Consequently, the deformability of a molecule M may be described by two spheres, corresponding, respectively, to the molecular vdW volume V^W – the smallest molecular sphere, S_S, and to molecular vdW surface, S^W – the greatest molecular sphere, S_G [12-14]. The vdW radius, r_S^w, and the vdW volume, V_S^w, of the greatest molecular S_G sphere are calculated using the following relations,

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

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

The vdW radius, r_V^w, and the vdW surface area S_V^w of the molecular S_S sphere are calculated as follows:

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

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

Thus, the molecular S_G and S_S spheres were described by the following two triplets:

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

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

On the basis of the spheres defined by relations (6) and 7 we introduced three molecular vdW ovality descriptors, denoted by Θ_{iD}, i=1,2,3, where D refers to the dimensionality of the vdW molecular space. Thus, taking into account the characteristics of the greatest and the smallest molecular sphere, S_G (relation 6), and S_S (relation 7), respectively, the ovality descriptors have been defined as follows [12],

$$\Theta_{1D} = \frac{r_S^w}{r_V^w} \quad (8)$$

$$\Theta_{2D} = \frac{S^W}{S_V^W} \quad (9)$$

$$\Theta_{3D} = \frac{V_S^W}{V^W} \quad (10)$$

One may observe that the all three descriptors Θ_{1D}, Θ_{2D}, and Θ_{3D} are dimensionless, but they refer, respectively, to one-dimensional (1D), bi-dimensional (2D), and tridimensional (3D) vdW molecular space. The values of Θ_{iD}, i=1,2,3 descriptors systematized in Table I were computed with the aid of the IRS software package [22].

Table I

Toxic activity and ovality molecular descriptors for the esters from study data set

No.	Esters	A	Θ_{1D}	Θ_{2D}	Θ_{3D}
1	decyl acetate	1.8794	1.3000	1.6910	2.1993
2	methyl undecanoate	1.4248	1.2985	1.6874	2.1821
3	methyl decanoate	1.3778	1.2823	1.6458	2.1084
4	octyl acetate	1.0570	1.2695	1.6133	2.0507
5	vinyl 2-ethylhexanoate	1.0462	1.2516	1.5668	1.9584
6	methyl nonanoate	1.0419	1.2675	1.6074	2.0346
7	allyl heptanoate	0.7282	1.2602	1.5889	2.0055
8	methyl octanoate	0.5358	1.2502	1.5632	1.9490
9	butyl butyrate	0.5157	1.2331	1.5214	1.8801
10	allyl hexanoate	0.2128	1.2413	1.5419	1.9126
11	butyl propionate	0.1704	1.2152	1.4774	1.7874
12	amyl acetate	0.1625	1.2167	1.4805	1.7968
13	methyl heptanoate	0.1039	1.2312	1.5172	1.8663
14	ethyl hexanoate	0.0637	1.2310	1.5163	1.8705
15	propyl valerate	0.0094	1.2284	1.5094	1.8513
16	hexyl acetate	-0.0087	1.2354	1.5271	1.8813
17	amyl propionate	-0.0431	1.2318	1.5182	1.8652
18	2-ethylbutyl acetate	-0.1202	1.2160	1.4793	1.8006
19	ethyl valerate	-0.3580	1.2144	1.4762	1.7902

No.	Esters	A	Θ_{1D}	Θ_{2D}	Θ_{3D}
20	n-hexyl formate	-0.3824	1.2183	1.4856	1.8067
21	vinyl butyrate	-0.3825	1.1792	1.3908	1.6432
22	tert butyl propionate	-0.4095	1.1960	1.4315	1.7029
23	propyl butyrate	-0.4138	1.2114	1.4694	1.7819
24	butyl acetate	-0.4864	1.1945	1.4276	1.6970
25	isopropenyl acetate	-0.4892	1.1495	1.3202	1.5133
26	ethyl butyrate	-0.4903	1.1926	1.4230	1.6982
27	methyl hexanoate	-0.5611	1.2143	1.4750	1.7835
28	isobutyl isobutyrate	-0.5908	1.2258	1.5030	1.8416
29	allyl butyrate	-0.6355	1.2029	1.4482	1.7441
30	vinyl propionate	-0.6530	1.1547	1.3345	1.5389
31	propargyl propionate	-0.6554	1.1708	1.3708	1.6045
32	sec-butyl acetate	-0.6794	1.1838	1.4025	1.6619
33	isobutyl propionate	-0.6935	1.2115	1.4683	1.7752
34	ethyl isovalerate	-0.7231	1.2085	1.4615	1.7608
35	n-amyl formate	-0.7826	1.1967	1.4335	1.7118
36	propyl propionate	-0.8148	1.1939	1.4262	1.6973
37	methyl valerate	-0.8448	1.1751	1.3807	1.6240
38	vinyl acetate	-0.8595	1.1297	1.2770	1.4412
39	allyl propionate	-0.8791	1.1795	1.3927	1.6327
40	2-butynyl-acetate	-0.8834	1.1727	1.3774	1.6131
41	ethyl-2-methylbutyrate	-0.8893	1.2068	1.4577	1.7505
42	butyl formate	-0.9336	1.1737	1.3779	1.6164
43	ethyl propionate	-0.9450	1.1697	1.3698	1.6021
44	propyl formate	-1.0221	1.1488	1.3207	1.5110
45	methyl-2-methylbutyrate	-1.1650	1.1836	1.4021	1.6663
46	propargyl acetate	-1.1664	1.1466	1.3149	1.5086
47	propyl acetate	-1.2382	1.1703	1.3713	1.6075
48	methyl butyrate	-1.2463	1.1670	1.3612	1.5890
49	ethyl isobutirate	-1.2709	1.1912	1.4195	1.6931
50	ethyl acetate	-1.2968	1.1464	1.3159	1.5028
51	isobutyl formate	-1.3081	1.1695	1.3679	1.5965
52	tert butyl formate	-1.3719	1.1561	1.3367	1.5448
53	methyl formate	-1.4982	1.0898	1.1872	1.2938
54	isopropil acetate	-1.5900	1.1649	1.3573	1.5791
55	methyl acetate	-1.5954	1.1176	1.2499	1.3964
56	methyl propionate	-1.6092	1.1443	1.3112	1.4983

Molecular modeling

Three-dimensional conversions of constitutional formulas and pre-optimization were performed using the molecular mechanics MM+ algorithm implemented in the HyperChem Package (*hyper.com*).

Final geometry optimization of the ester molecules was carried out by using the semi-empirical quantum-mechanical AM1 parametrization, and the optimized geometries were loaded into our in home developed IRS software (http://irs.cheepe.homedns.org/). In this way, the ovality descriptors from Table I were calculated.

QSAR analysis

Toxicological QSARs (QSTRs) were developed using the regression procedure of MobyDigs software [24]. A=Log(1/IGC$_{50}$) values reported millimolar were used as the independent variable (see Table I). Molecular descriptors quantifying the shape of ester molecules, i.e. the Θ_{iD}, i=1,2,3 values from Table I, were used as predictor variables. Resulting models were measured for goodness of fit by the correlation coefficient (r) and the coefficient of determination adjusted for the degree of freedom (r^2_{adj}). The uncertainty in the model was quantified by standard error (*s*), and the reliability by the *F* (Fisher) and *t* (Student) statistics. The t-test was used to determine the 95% confidence limits of the QSAR models. The predictive ability of QSARs was noted as the cross-validation coefficient q^2 determined by the leave-n-out method (LOO if n=1). The quantity q^2 is also known as coefficient of prediction. Outliers were identified with reference to their residual values being outside the 95% confidence interval of the linear QSAR model.

Finally a chance correlation was checked by scrambling the toxicological response values (Y-scrambling) [24] and trying to build a model using the scrambled data. This procedure is then repeated several times and the r^2 and q^2 values are checked against that for the real QSAR. One expect low r^2_{Y-s} and low q^2_{Y-s} values: if only one of the r^2 (or q^2) values from the scrambled data is as high as that from real QSAR, then there is a risk that the real QSAR is a chance correlation [23].

The statistical calculations used for the development of QSAR models were made with MobyDigs computer program [24].

Results And Discussion

In previous paper [14] we reported QSAR studies on these esters using other structural parameters, such as the vdW compressibility. It is important to use in such studies as predictor variables the structural characteristics of a molecule, which are easy to calculate for molecules with many atoms and various configurations and conformations and, also, to interpret in the terms of molecular features and physical meanings. In this way, one may obtain supplementary information about the interaction between toxicant molecule and its biologic target. To detect the best molecular descriptors, we developed reliable QSAR models for a series of congeners – alcohols, amines, esters, etc.

The aliphatic esters in Table I, sorted in the descending order of their toxicity values, A= $Log(IGC_{50})^{-1}$, agree with Tetratox database [7]. The biological activity A represents the logarithm of the inverse of concentration, measured millimolar, which produces 50% growth inhibition to *T. Pyriformis*.

The ovality molecular descriptors Θ_{iD}, i=1,2,3 have been used in this QSAR study as predictor (dependent) variables. They are theoretically derived structural molecular parameters related to the van der Waals space of ester molecules. Because the reference elements of the two spheres, $\{S_S\}$ and $\{S_G\}$ with respect to units proportional to the size of a molecule M (Å, $Å^2$, $Å^3$), the shape characterization by the three ovality descriptors (relations 8, 9, and 10) is size-invariant, that is, a "pure" shape characterization is obtained [21]. In fact, these descriptors are dimensionless measures of the molecular shape. They measure the intrinsic degree of elongation of a molecule without any reference to another molecule. For molecules whose shape is closest to the shape of a sphere, the index values tend to 1 (unity value). The shape of a molecule is more oblong, the index values are greater than 1. The values of Θ_{iD}, i=1,2,3 – see relations (8), (9), and (10), respectively, were calculated with IRS software package [22] for optimal geometries of ester molecules. These geometries were obtained as was described above – see section of molecular modeling. To reduce the internal strains, the ester molecules have generally adopted an ANTI conformation corresponding to the most elongated shape. Θvalues for the aliphatic esters were systematized in Table I.

The linear QSAR models obtained by correlating toxicity (*A*) *versus* ovality descriptors Θ_{iD}, i=1,2,3, are the following:

$$\hat{A} = -20.622(\pm 2.676) + 16.838(\pm 2.229)\cdot \Theta_{1D}$$
$$n = 56, s = 0.364; r = 0.899; r^2_{adj} = 0.805; F = 228.1 \qquad (11)$$

$$\hat{A} = -10.562(\pm 1.302) + 7.032(\pm 0.901)\cdot \Theta_{2D}$$
$$n = 56, s = 0.354; r = 0.905; r^2_{adj} = 0.815; F = 243.8 \qquad (12)$$

$$\hat{A} = -7.214(\pm 0.843) + 3.920(\pm 0.484)\cdot \Theta_{3D}$$
$$n = 56, s = 0.343; r = 0.911; r^2_{adj} = 0.826; F = 262.7 \qquad (13)$$

In relations (11) – (13) $\hat{A}$ is the calculated value of experimental inhibitory activity A=LogIGC$_{50}^{-1}$, *n* represents the number of data set, *s* stands for the standard error, *r* is the correlation coefficient and r^2_{adj} represents 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 one can see from the values of r, r^2_{adj}, s, and F statistics. The reliability in the all QSAR models is very close – see the values of the Fisher test, F, and the confidence limits.

All ovality descriptors work well, but the best QSAR model (13) corresponds to Θ_{3D} ovality shape molecular descriptor, quantifying the form more or less spherical of ester molecules in the tri-dimensional space (3D). The equation (13) roughly explains 83% of the variance of experimental values. The shape of the ester molecule as quantified by these ovality descriptors interferes in the biological interaction and explains an important part of measured toxicity values.

In order to discriminate the statistical fit from the ability of a model to make predictions, we used the leave-one-out (LOO) and the *leave-n-out* (LnO) cross-validation method to estimate the predictive ability of the obtained QSAR model, via the cross-validation coefficient (also called coefficient of predictions), q^2. In the LOO procedure one compound is removed from the training set, the QSAR is reconstructed using the remaining compounds, and the toxicological activity of the deleted compound is then predicted with the new QSAR model. The deleted compound is then reintroduced in the initial set and the procedure is repeated until each compound in turn has been left out. A value of $q^2>0.5$ is acceptable [23-25]. The results obtained by cross-validation procedure applied to this series of esters exhibiting toxic activity against *T. Pyriformis* are presented in Table II.

Table II

Values of the statistics used to assess the predictive power of the QSAR models A vs. ovality descriptors (Θ_{iD}, *i=1,2,3*).[#]

QSARs	CDs	q^2_{LOO}	q^2_{BOOT}	SDEP	SDEC	r^2_{Y-s}	q^2_{Y-s}
(11)	Θ_{1D}	0.789	0.777	0.375	0.357	-0.021	-0.100
(12)	Θ_{2D}	0.801	0.790	0.364	0.348	-0.026	-0.106
(13)	Θ_{3D}	0.814	0.804	0.352	0.337	-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.

The predictive power of these models is good, the values of cross-validation coefficients being greater than 0.770 (if we take into account the commonly accepted values for a satisfactory QSTR model: $q^2 > 0.500$). Consequently, the QSARs (11) – (13) are sufficiently robust and stable. Bootstrapping simulates what would be 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 alcohols on *Tetrahymena pyriformis*, $\log IGC_{50}^{-1}$) was randomly shuffled and a new QSTR model was developed using the Θ_{iD}, i=1,2,3 independent variables. The process was repeated 300 times and the resulting QSTR models have the expected low r^2_{Y-s} and low q^2_{Y-s} values, which are presented in Table II. SDEC is the square root of the residual sum of squares divided by the number of compounds in the training set (standard deviation error in calculation); SDEP (standard deviation error in prediction) is similar to SDEC, but the residuals are calculated by using the predicted value of the dependent variable when an ester is left out from the training set and put into the test set.

We also applied a L*50%*O procedure, which may be considered as a global external method for validation, because the chemical structures not used in the training set were selected for inclusion in the validation set, and reciprocally. All data points were ordered in the descending order of A-toxicity values, and the series in Table I was separated into two subsets (conditionally denoted as odd and even series) by selection of every second point from the original dataset in order to obtain a similar distribution of the investigated property values for the whole set. The standard QSAR modeling procedure (see the section "QSAR Analysis") was applied to those two datasets. The results presented below are only for external validation of the model (13) obtained with Θ_{3D} as predictor variable when applied the above described leave-odd – even-out (LoeO) cross-validation technique [26-27].

The QSAR model (14) was obtained when we used the odd ranking subset as a training set.

$$\hat{A} = -7.324(\pm 1.395) + 3.986(\pm 0.797)\cdot \Theta_{3D}$$
$$n = 28, s = 0.392, r = 0.896, r^2_{adj} = 0.795; F = 105.8 \qquad (14)$$

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

$$\hat{A} = -7.098(\pm 1.080) + 3.850(\pm 0.623) \cdot \Theta_{3D}$$
$$n = 28, s = 0.301, r = 0.928, r^2_{adj} = 0.856; F = 161.7 \quad (15)$$

The QSAR equations (14) developed from data of odd ranking subset and (15) for even ranking subset were used to predict the toxicity of the esters in test sets, namely the even ranking subset, and the odd ranking subset, respectively. The values of predict toxicity values, A_{pr}, together with their deviations from experimental values, A, are given in Table III.

Table III

The values of predicted toxicity values, A_c and A_{pr}, together with their deviations from experimental values for the 56 esters presented in Table I

No.	A_c [#]	$\Delta = A - A_c$	A_{pr} [*]	$\Delta = A - A_{pr}$
1	1.4073	0.4721	1.3692	0.5102
2	1.3398	0.0850	1.3740	0.0508
3	1.0509	0.3269	1.0192	0.3586
4	0.8247	0.2323	0.8502	0.2068
5	0.4629	0.5833	0.4417	0.6045
6	0.7616	0.2803	0.7861	0.2558
7	0.6476	0.0806	0.6231	0.1051
8	0.4261	0.1097	0.4449	0.0909
9	0.1560	0.3597	0.1403	0.3754
10	0.2834	-0.0706	0.2998	-0.0870
11	-0.2074	0.3778	-0.2166	0.3870
12	-0.1705	0.3330	-0.1617	0.3242
13	0.1019	0.0020	0.0871	0.0168
14	0.1184	-0.0547	0.1320	-0.0683
15	0.0431	-0.0337	0.0294	-0.0200
16	0.1607	-0.1694	0.1751	-0.1838
17	0.0976	-0.1407	0.0829	-0.1260
18	-0.1556	0.0354	-0.1466	0.0264
19	-0.1964	-0.1616	-0.2059	-0.1521
20	-0.1317	-0.2507	-0.1223	-0.2601
21	-0.7727	0.3902	-0.7718	0.3893
22	-0.5386	0.1291	-0.5360	0.1265
23	-0.2290	-0.1848	-0.2378	-0.1760
24	-0.5618	0.0754	-0.5595	0.0731
25	-1.2819	0.7927	-1.2719	0.7827
26	-0.5571	0.0668	-0.5547	0.0644
27	-0.2227	-0.3384	-0.2317	-0.3294
28	0.0051	-0.5959	0.0168	-0.6076
29	-0.3771	-0.2584	-0.3833	-0.2522
30	-1.1815	0.5285	-1.1897	0.5367
31	-0.9244	0.2690	-0.9208	0.2654
32	-0.6994	0.0200	-0.6994	0.0200
33	-0.2552	-0.4383	-0.2636	-0.4299
34	-0.3117	-0.4114	-0.3052	-0.4179

No.	A_c [#]	$\Delta=A-A_c$	A_{pr} [*]	$\Delta=A-A_{pr}$
35	-0.5037	-0.2789	-0.5077	-0.2749
36	-0.5606	-0.2542	-0.5583	-0.2565
37	-0.8479	0.0031	-0.8457	0.0009
38	-1.5645	0.7050	-1.5791	0.7196
39	-0.8138	-0.0653	-0.8122	-0.0669
40	-0.8906	0.0072	-0.8939	0.0105
41	-0.3520	-0.5373	-0.3587	-0.5306
42	-0.8777	-0.0559	-0.8808	-0.0528
43	-0.9338	-0.0112	-0.9300	-0.0150
44	-1.2909	0.2688	-1.3009	0.2788
45	-0.6821	-0.4829	-0.6829	-0.4821
46	-1.3003	0.1339	-1.3104	0.1440
47	-0.9126	-0.3256	-0.9093	-0.3289
48	-0.9851	-0.2612	-0.9900	-0.2563
49	-0.5770	-0.6939	-0.5797	-0.6912
50	-1.3230	0.0262	-1.3335	0.0367
51	-0.9557	-0.3524	-0.9516	-0.3565
52	-1.1584	-0.2135	-1.1661	-0.2058
53	-2.1423	0.6441	-2.1170	0.6188
54	-1.0239	-0.5661	-1.0294	-0.5606
55	-1.7401	0.1447	-1.7220	0.1266
56	-1.3407	-0.2685	-1.3515	-0.2577

[#] Values calculated with Equation (13); [*] Values calculated with Equations (14) for even ranking data set and (15) for odd ranking data set.

The values of the statistics associated to Equations (14) and (15) are given in Table IV.

The LoeO 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 [27].

Table IV

Values of the statistics used to assess the predictive power of the QSAR models (14) and (15)

QSAR	q^2_{LOO}	q^2_{BOOT}	SDEP	SDEC	r^2_{Y-s}	q^2_{Y-s}
(14)	0.759	0.728	0.418	0.378	-0.023	-0.188
(15)	0.837	0.818	0.314	0.290	-0.028	-0.185

Conclusions

Baseline toxicity can be understood as a disruption of the functions of biological membranes, although the detailed mechanism remains unclear.

The aliphatic esters act on the cellular membrane. The shape of the molecules, as measured by Θ_{iD}, seems to be an important factor affecting the integrity of these membranes.

Θ_{iD} are molecular shape parameters, which describe the degree of deviation of a molecule from a spherical (tetrahedral) shape. They model well the toxicity in a series of 56 aliphatic esters on *Tetrahymena pyriformis*.

The values of these ovality molecular structural parameters can be easily computed, and their physical meaning is clear. The good results correlation prove that these shape molecular descriptors are valuable tools to model the toxicity of chemical compounds.

References

1. Chiriac A., Ciubotariu D., Simon Z., Relaţii cantitative structură chimică - activitate biologică (QSAR).Metoda MTD. Editura Mirton, Timişoara, 1996.
2. Schultz T. W., Cronin M. T. D., Walker J. D., Osman A., QSAR in toxicology: a historical perspective, *J. Mol. Struct. (Theochem)*, 2003, 622, 1-22.
3. Schultz T. W., Netzeva T. I., Cronin M. T. D., Selection of data sets for QSARs: Analyses of *Tetrahymena* toxicity from aromatic compounds, *SAR QSAR Environ. Res*., 2003, 14, 59-81.
4. Schultz T. W., Cronin M. T. D., Netzeva T. I., The present status of QSAR in toxicology, *J. Mol. Struct. (Theochem),* 2003, 622, 23-28.
5. Schultz T. W., Cronin M. T. D., Netzeva T. I., Aptula A. O., Structure-Toxicity Relationships for aliphatic chemicals evaluated with *Tetrahymena pyriformis, Chem. Res. Toxicol.*, 2002, 15, 1602-1609.
6. Netzeva T. I., Pavan M., Worth A. C., Review of (Quantitative) Structure-Activity Relationships for acute aquatic toxicity, *QSAR Comb. Sci*., 2008, 27(1), 77-90.
7. http://www.vet.utk.edu/TETRaTOX/
8. Renner R. The Kow controversy. Doubts about the quality of basic physicochemical data for hydrophobic organic compounds could be undermining many environmental models and assessments. *Environ. Sci. Technol*. 2002, 36, 410A-413A.
9. Benfenati E., Gini G., Piclin, N., Roncaglioni A., Vari M. R. Predicting logP of pesticide using different software. *Chemosphere* 2003, 53, 115-1164.
10. Anoiaca P.G., Anoaica M.B., Amzoiu E., Giubelean L., Tătucu M., A CAMD-QSAR methodology for prospecting the potential active antihistaminic substances: class of mandelic acid esters, *Farmacia,* 2010, 4, 515-523.
11. http://clogP.pomona.edu/medchem/chem/clogP/index.html
12. 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 Int. Workshop Quant. Struct.-Act. Relat. Environ. Toxicol. 2006, 8-12 May, Lyon, France, p. 90.
13. Ciubotariu D., Vlaia V., Ciubotariu C., Olariu T., Medeleanu M., Molecular Shape Descriptors: Applications to Structure – Activities Studies, in Carbon Bonding and Structures. Advances in Physics and Chemistry, M. V. Putz (Ed.), Springer, Dordrecht Heidelberg, 2011, pp. 337-378.
14. Vlaia V., Olariu T., Butur M., Vlaia L., Ciubotariu C., Medeleanu M., Ciubotariu D., Quantitative Structure-Activity Relationship (QSAR). V. Analysis of The Toxicity of

Aliphatic Esters by Means of Molecular Compressibility Descriptors., *Farmacia,* 2009, 5, 549-561.
15. Ciubotariu D., Medeleanu M., Vlaia V., Olariu T., Ciubotariu C., Dragos D., Seiman C., Molecular van der Waals Space and Topological Indices from the Distance Matrix, *Molecules*, 2004, 9, 1053-1078.
16. Crisan O., Oprean Luminita, Marchand J., Influence de l'isomerie sur les effets toxiques, *Farmacia* 2007, LV, 2, 144-156.
17. Drăgoi M. C., Mitrea N., Arsene A. L., Ilie M., Nicolae A. C., Jurkat E 6.1 cell line studies regarding the effects of some bio-indols on the membrane fluidity, *Farmacia*, 2012, 60(1), 13-20.
18. Bodoki E., Iacob B. C., Tiuca I., Oprean R., Chemometric interpretation of spectroscopic data for chiral probing: a review, *Farmacia*, 2011, 59(3), 283-302.
19. 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.
20. Ciubotariu D., Structure – Reactivity Relationships in the Series of Carbonic Acid Derivatives, PhD Thesis (in Romanian language), Polytechnic Institute of Bucharest,1987.
21. Mezey P. G., Shape in Chemistry. An Introduction to Molecular Shape and Topology, VCH Publishers, Inc., New York, 1993; 23-103.
22. http://irs.cheepe.homedns.org/
23. Esbensen K. H., Multivriate Data Analysis-In Practice; CAMO Process AS: Oslo, Norway, 2004
24. MobyDigs v.1.1 is available from Talete srl., via V. Pisani, 13-20124, Milano, Italy, http://www.talete.mi.it
25. Dearden J.C., In Computer Applications in Pharmaceutical Research and Development; Ekins, S.; John Wiley & Sons, Inc.: NewYork, 2006, p. 469-494
26. 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
27. 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: February 1st 2012