TOXICOPHORES AND QUANTITATIVE STRUCTURE -TOXICITY RELATIONSHIPS FOR SOME ENVIRONMENTAL POLLUTANTS

N. N. Gorinchoy[a], I. Ya. Ogurtsov[a], A. Tihonovschi[a], I. Balan[a], I. B. Bersuker[a,b], A. Marenich[b] and J.Boggs[b]

[a] *Institute of Chemistry, Academy of Sciences of Moldova, Academiei str. 3, MD 2028 Kishinev, Republic of Moldova*
[b]*Department of Chemistry & Biochemistry, University of Texas at Austin, USA*
**E-mail:* ngorinchoy@yahoo.com*; Phone 373 22 739649*

Abstract: The electron-conformational (EC) method is employed to reveal the toxicophore and to predict aquatic toxicity quantitatively using as a training set a series of 51 compounds that have aquatic toxicity to fish. By performing conformational analysis (optimization of geometries of the low-energy conformers by the PM3 method) and electronic structure calculations (by *ab initio* method corrected within the SM54/PM3 solvatation model), the Electron-Conformational Matrix of Congruity (ECMC) was constructed for each conformation of these compounds. The toxicophore defined as the EC sub-matrix of activity (ECSA), a sub-matrix with matrix elements common to all the active compounds under consideration within minimal tolerances, is determined by an iterative procedure of comparison of their ECMC's, gradually minimizing the tolerances. Starting with only the four most toxic compounds, their ECSA (toxicophore) was found to consists of a 4x4 matrix (four sites with certain electronic and topologic characteristics) which was shown to be present in 17 most active compounds. A structure-toxicity correlation between three toxicophore parameters and the activities of these 17 compounds with R^2=0.94 was found. It is shown that the same toxicophore with larger tolerances satisfies the compounds with les activity, thus explicitly demonstrating how the activity is controlled by the tolerances quantitatively and which atoms (sites) are most flexible in this respect. This allows for getting slightly different toxicophores for different levels of activity. For some active compounds that have no toxicophore a bimolecular mechanism of activity is suggested. Distinguished from other QSAR methods, no arbitrary descriptors and no statistics are involved in this EC structure-activity investigation.

Keywords: aquatic toxicity, electron-conformational method, QSAR

Abbreviations: Tph –toxicophore
EC – electron-conformational
ECMC – electron-conformational matrix of congruity
ECSA – electron-conformational sub-matrix of activity
QSAR – quantitative structure-activity relationships

Introduction

Many chemicals (organic environmental pollutants) and drugs, in addition to their useful properties, possess different types of toxicity to environmental living organisms. The QSAR methodology is usually used to reveal relationships between the chemical structure of the compound and the toxicity under consideration in order to predict the latter in new chemicals. The main problem in this approach is to choose the molecular features (descriptors) that properly represent the possible interaction of the toxicant with the bioreceptor to produce the toxicity, and to correlate the descriptors with the toxicity by means of some regression relationships. There are many monographs, review articles and original works devoted to this problem (see e.g. [1-3] and references therein). One of the latest versions of the QSAR approach to this problem is the so-called support vector machine (SVM) model [4, 5] in which the hypersurface in multi- dimensional space of the descriptors separates two classes of chemicals, toxic and not toxic. A rule to classify chemicals into four classes was proposed by Verhaar *et al.* [6].

In all these approaches to QSAR problems there is a common shortcoming: the choice of descriptors is not directly based on first principles, meaning it is arbitrary, and their weight is evaluated by means of statistical comparison with the activities. Since such a set of descriptors is necessarily incomplete and they are non-orthogonal to each other with unknown overlap, the comparison with activities may lead to chance correlations, some (or all) descriptors being thus artifacts with no physical meaning implied in their initial choice (see also [7]).

In the present paper we explore some problems of chemical toxicity, more precisely aquatic toxicity to fish, using the electron-conformational (EC) method of pharmacophore identification and quantitative bioactivity predictions which is free from the above shortcomings (see the reviews of this method in [8], and also in [1], p. 455; an earlier less accurate qualitative version of this method was reviewed in [9]). Distinguished from the traditional QSAR approaches, the EC method does not employ arbitrary descriptors and statistics in its evaluation. Instead one (a unique) descriptor based on first principles is used, the electronic structure and topology of the molecule, calculated by means of quantum-chemical methods and presented in computer friendly digital-matrix form (see below). The comparison of these matrices with the

activities allows us to reveal a group of matrix elements that are common to the active compounds under consideration and represent the numerical picture of pharmacophore, the necessary condition of activity (no statistics is involved in this process). This approach has been applied successfully to study several types of biological activities (see, e.g., [10-14]).

There are at least two recent publications where some approaches for prediction of aquatic toxicity were proposed. A rule-based system to classify chemicals into four classes was proposed by Verhaar *ei al* [6]. Using this approach it is possible to calculate the toxicity of chemicals belonging to one of those four classes based on the octanol-water partition coefficient *(P)* of the compound. The other approach applies CODESSA descriptors to correlate with toxicity through the use of regression relationships [15]. The latter publication is based on a data set comprising 293 diverse chemicals with toxicity to *Poecilia reticulata* (guppy). A significant contribution of HYBOT descriptors in modeling polar and non-polar narcosis was reported earlier [16]. These publications represent important progress in the field of construction of structure-ecotoxicity models. Nevertheless the creation of stable predictive models of ecotoxicological properties of organic chemicals is still a long way from realization.

In this paper we intend to construct an alternative (to [15, 16]) approach for the toxicity to *Poecilia reticulata* (guppy) based on the EC method (ECM). We apply the ECM only to one of four groups (Class 3) discussed in [15, 16], which consist of 51 compounds (Table 1, numbers 212-262 in [15, 16]).

Table 1

Name and experimental toxicity -log(LC_{50}^{exp}) of the compounds in the data set

No	Name	-logLC_{50}^{exp}	No	Name	-log LC_{50}^{exp}
1	4-Dinitrobenzylbromide	0.30	27	1,3-Butadienediepoxide	-1.49
2	Hexachlorobutadiene	0.20	28	Benzaldehyde	-1.57
3	1-Chloro-2,4-dinitrobenzene	0.19	29	1,2,7,8-Diepoxyoctane	-1.67
4	1,4-Dichloro-2-butene	0.16	30	Styrene oxide	-1.77
5	α,α'-Dichloro-m-xylene	0.16	31	Octanal	-1.79
6	2,4,α-Trichlorotoluene	-0.08	32	l-Chloro-2-butene	-1.82
7	2-sec-Butyl-4,6-dinitrophenol	-0.17	33	3-Chloro-l-butene	-1.85
8	Pentachlorophenol	-0.22	34	2-Ethylbutenal	-1.89
9	Benzyl chloride	-0.49	35	Heptanal	-1.89
10	2,3,4,6-Tetrachlorophenol	-0.67	36	1,2-Epoxyoctane	-1.91
11	Epibromhydrin	-0.77	37	Cyclohexanecarboxaldehyde	-1.91
12	1,2-Epoxydodecane	-0.78	38	Hexanal	-1.99
13	Epichlorohydrin	-0.85	39	2-Furaldehyde	-2.04
14	Chloroacetone	-0.88	40	Pentanal	-2.18
15	2-Butenal	-0.90	41	3-Methylbutanal	-2.19
16	1,3-Dichloropropene	-0.98	42	1,2-Epoxyhexane	-2.27
17	2,5-Dinitrophenol	-1.00	43	Butanal	-2.28
18	3,4,5,6-Tetrachloro-2-hydroxyphenol	-1.00	44	Propanal	-2.41
19	2,3-Dichloropropene	-1.01	45	2,2-Dichlorodiethyl ether	-2.54
20	3-Cyclohexene-1-carboxaldehyde	-1.01	46	2-Methylpropanal	-2.57
21	3,4,5-Trichloro-2-methoxyphenol	-1.03	47	1,2-Epoxybutane	-2.66
22	3,4,5-Trichloro-2,6-dimethoxyphenol	-1.12	48	Propylene Oxide	-2.74
23	Allyl chloride	-1.20	49	Glycidol	-2.83
24	Decanal	-1.31	50	Acetaldehyde	-2.90
25	1,2-Epoxydecane	-1.32	51	Methanal	-2.96
26	4,5- Dichloro-2-methoxyphenol	-1.40			

Application of the Electron-Conformational Method

The electron-conformational method in its general form consists of the following consecutive steps [8]:

1) Evaluation of the low-energy conformations of the compounds in the training set.

2) Electronic structure calculation of all conformers.

3) Construction of ECMC for each conformation.

4) Identification of the ECSA, the toxicophore, by multiple comparisons of the ECMC of the active compounds.

5) Estimation of the influence of pharmacophore flexibility, anti- toxicophore shielding groups, and other out-of-toxicophore groups by means of a proper parameterization and least-squares regression analysis;

6) Use of the obtained toxicophore and out-of- toxicophore influence parameters for screening new potentially active compounds and prediction of their activity.

These steps were consecutively realized for the training set of 51 compounds.

Step 1. Using the methods of molecular mechanics with the Merck force field [17] and the Monte-Carlo randomized search method the low-energy conformers were determined for all compounds in Table 1.

Step 2. The energies of the conformers were calculated including the aqueous solvation effect by means of the SM5.4/P model [18].

Conformational analysis and electronic structure calculations were performed using the SPARTAN package [19]. The geometries of conformers were optimized with the semi-empirical PM3 method [20]. For optimized geometries single-point calculations of the total energies were carried out using the *ab initio* method in Restricted Hartee-Fock-Roothaan approximation with the 6-31G* basis sets. Only conformers with the lowest total energy within 1 kcal/mole were kept for further consideration.

Step 3. *Computation of electron-conformational matrices of congruity (ECMC)*

The results of single-point calculations of the conformer's electronic structure (molecular orbital population analysis, Mulliken atomic charges, and bond orders) were used for construction of the ECMC for all the compounds in Table 1. In the ECMC the diagonal elements are atomic interaction indices (*II*) [8], a measure of electron-donor properties of the corresponding atoms in the molecule (numerical coefficients are chosen to compensate dimensionalities):

$$II^A = g^A \exp\left(-R_0\sqrt{2\,\mathrm{VOIP}^A}\right) \quad , \qquad (1)$$

where g^A is the Mulliken electron population of the outermost orbital of the atom A (for *n*p-elements the g^A is equal to a third of the total occupancy of valence p-orbitals, p_x, p_y, and p_z, of the atom), and VOIP^A in Hartree units refers to the valence orbital ionization potential of this atom-in-molecule orbital calculated as a function of the Mulliken charge and the electronic configuration of the atom using the reference data [21]. The value of R_0 = 1.51 Bohr radii (0.8 Å) is conventional. It is approximately equal to the distance between the points of the maximum electronic density (R_1) of the outmost orbital of the atom in the molecule and of the maximum overlap (R_2) of this atomic orbital with the wave function of an arbitrary target atom [11]. The value $R_2 - R_1 \sim 0.8$ Å for C, N, O, or S can be obtained directly from the values of the van der Waals radii [22] for R_2 and Slater atomic radii [23] for R_1. Off-diagonal matrix elements represent Mulliken bond orders for chemically bonded atoms and interatomic distances for non-bonded pairs.

Figure 1 illustrates an example of the ECMC calculated for the lowest energy conformation of the compound **1**. For simplicity, the hydrogen atoms are excluded from consideration hereafter.

	C1	C2	C3	C4	C5	C6	C7	Br	N1	N2	O1	O2	O3	O4
C1	0,22	1,42	2,43	2,81	2,43	1,41	0,94	2,72	4,31	2,55	3,48	3,04	5,01	5,01
C2		0,19	1,40	2,42	2,79	2,41	2,54	3,52	3,80	0,73	2,34	2,35	4,19	4,76
C3			0,29	1,43	2,41	2,78	3,81	4,65	2,51	2,48	2,91	3,41	2,80	3,59
C4				0,20	1,41	2,40	4,28	5,11	0,77	3,77	4,26	4,63	2,34	2,34
C5					0,29	1,44	3,75	4,60	2,51	4,30	4,98	5,01	3,59	2,80
C6						0,32	2,46	3,45	3,78	3,82	4,66	4,36	4,75	4,18
C7							0,41	0,96	5,79	3,04	3,97	3,11	6,48	6,44
Br								0,42	6,52	3,84	4,28	4,22	7,22	7,11
N1									0,14	4,99	5,28	5,87	1,50	1,50
N2										0,14	1,52	1,48	5,10	6,06
O1											0,44	2,12	5,22	6,40
O2												0,45	5,98	6,93
O3													0,45	2,12
O4														0,45

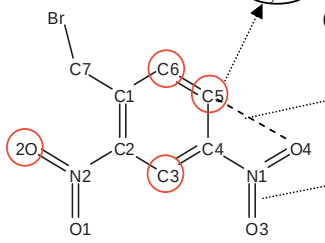


Fig.1. The electron-conformational matrix of congruity for molecule 1. The diagonal elements refer to the atomic (C_5 in the picture) interaction indices calculated by eq 1, while the off-diagonal elements reproduce Mulliken bond orders for chemically bonded pairs of atoms (N_1-O_3) and interatomic distances for non-bonded pairs (C_5-O_4).

Step 4. *Identification of the toxicophore (Tph)*

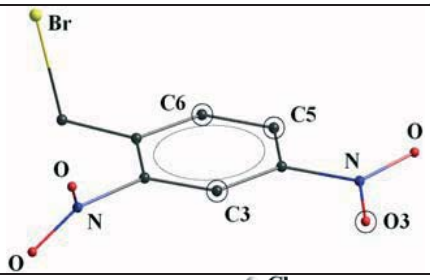
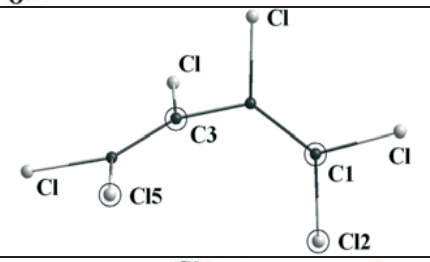
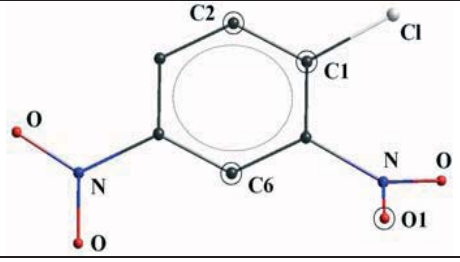
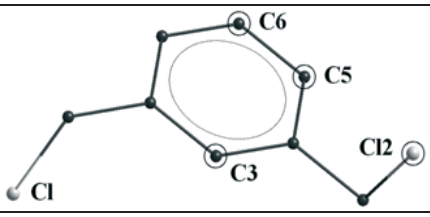
The ECSA sub-matrix that represents the Tph was obtained by comparing the ECMC's of the first four most toxic compounds in the training set. Applying the procedure described earlier [8] assuming reasonable tolerances we got the following common 4x4 submatrix of activity:

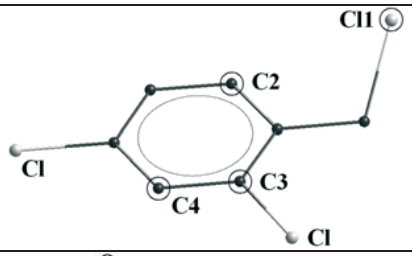
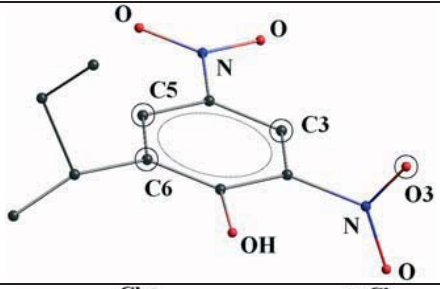
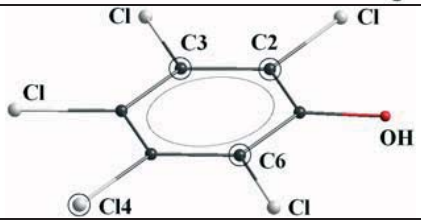
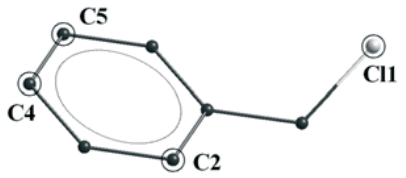
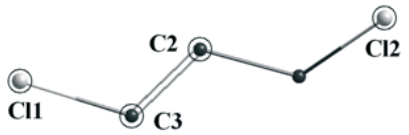
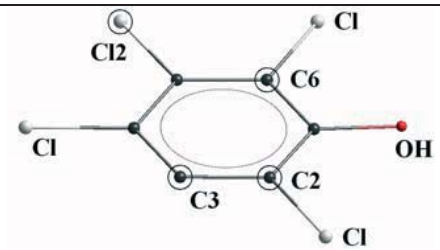
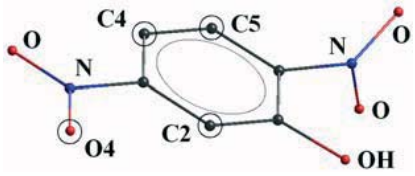
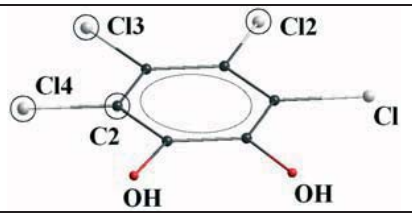
$$\begin{array}{c} \\ O3 \\ C6 \\ C5 \\ C3 \end{array}\begin{pmatrix} 0.457 & 4.753 & 3.598 & 2.804 \\ & 0.328 & 1.448 & 2.780 \\ & & 0.295 & 2.418 \\ & & & 0.296 \end{pmatrix} \qquad (2)$$

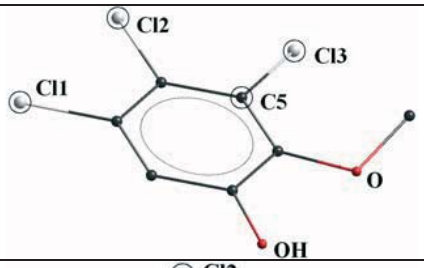
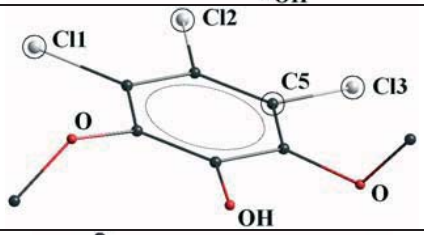
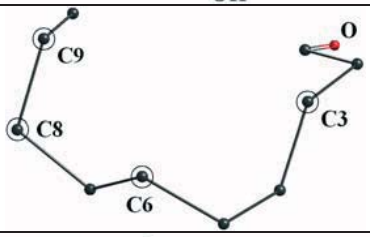
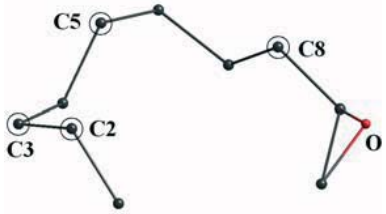
Analyzing the ECMC of the active compounds in the training set we found that the same ECSA obtained from the four most active compounds satisfies the next 17 compounds in some conformation assuming some reasonable tolerances for the matrix elements (see below). These compounds and their corresponding ECSA are shown in Table 2 with the circled four toxicophore atoms corresponding to these 4x4 matrices.

Table 2

The electron-conformational submatrices of toxicity and molecular structures for 17 compounds containing the *Tph*. The four toxicophore atoms are marked by circles.

1		$\begin{array}{c} \\ O3 \\ C6 \\ C5 \\ C3 \end{array}\begin{pmatrix} 0.458 & 4.753 & 3.598 & 2.804 \\ & 0.328 & 1.448 & 2.780 \\ & & 0.295 & 2.418 \\ & & & 0.296 \end{pmatrix}$
2		$\begin{array}{c} \\ Cl5 \\ Cl2 \\ C1 \\ C3 \end{array}\begin{pmatrix} 0.332 & 4.299 & 3.788 & 2.657 \\ & 0.332 & 1.059 & 3.075 \\ & & 0.307 & 2.471 \\ & & & 0.281 \end{pmatrix}$
3		$\begin{array}{c} \\ O1 \\ C2 \\ C1 \\ C6 \end{array}\begin{pmatrix} 0.452 & 4.724 & 3.560 & 2.835 \\ & 0.321 & 1.388 & 2.788 \\ & & 0.278 & 2.423 \\ & & & 0.294 \end{pmatrix}$
4		$\begin{array}{c} \\ Cl2 \\ C1 \\ C2 \\ C3 \end{array}\begin{pmatrix} 0.401 & 4.951 & 3.762 & 2.653 \\ & 0.396 & 0.965 & 2.469 \\ & & 0.295 & 1.943 \\ & & & 0.295 \end{pmatrix}$
5		$\begin{array}{c} \\ Cl2 \\ C6 \\ C5 \\ C3 \end{array}\begin{pmatrix} 0.399 & 4.647 & 3.467 & 3.437 \\ & 0.314 & 1.451 & 2.784 \\ & & 0.318 & 2.415 \\ & & & 0.320 \end{pmatrix}$

6		$\begin{matrix} Cl1 \\ C4 \\ C3 \\ C2 \end{matrix}\begin{pmatrix} 0.396 & 4.948 & 3.838 & 3.069 \\ & 0.304 & 1.404 & 2.790 \\ & & 0.299 & 2.408 \\ & & & 0.307 \end{pmatrix}$
7		$\begin{matrix} O3 \\ C5 \\ C6 \\ C3 \end{matrix}\begin{pmatrix} 0.452 & 4.992 & 4.513 & 3.208 \\ & 0.305 & 1.426 & 2.425 \\ & & 0.253 & 2.822 \\ & & & 0.296 \end{pmatrix}$
8		$\begin{matrix} Cl4 \\ C2 \\ C3 \\ C6 \end{matrix}\begin{pmatrix} 0.333 & 4.467 & 3.953 & 2.657 \\ & 0.321 & 1.368 & 2.433 \\ & & 0.278 & 2.802 \\ & & & 0.312 \end{pmatrix}$
9		$\begin{matrix} Cl1 \\ C4 \\ C5 \\ C2 \end{matrix}\begin{pmatrix} 0.400 & 5.129 & 4.638 & 3.454 \\ & 0.316 & 1.450 & 2.410 \\ & & 0.315 & 2.783 \\ & & & 0.318 \end{pmatrix}$
10		$\begin{matrix} Cl2 \\ Cl1 \\ C3 \\ C2 \end{matrix}\begin{pmatrix} 0.397 & 5.096 & 3.749 & 2.652 \\ & 0.3687 & 1.011 & 2.626 \\ & & 0.347 & 1.903 \\ & & & 0.282 \end{pmatrix}$
11		$\begin{matrix} Cl2 \\ C2 \\ C3 \\ C6 \end{matrix}\begin{pmatrix} 0.336 & 4.461 & 3.957 & 2.658 \\ & 0.321 & 1.405 & 2.431 \\ & & 0.295 & 2.808 \\ & & & 0.310 \end{pmatrix}$
17		$\begin{matrix} O4 \\ C5 \\ C4 \\ C2 \end{matrix}\begin{pmatrix} 0.459 & 4.764 & 3.603 & 2.791 \\ & 0.303 & 1.428 & 2.800 \\ & & 0.319 & 2.423 \\ & & & 0.348 \end{pmatrix}$
18		$\begin{matrix} Cl2 \\ Cl4 \\ C2 \\ Cl3 \end{matrix}\begin{pmatrix} 0.338 & 5.308 & 3.961 & 3.066 \\ & 0.335 & 1.119 & 3.048 \\ & & 0.309 & 2.661 \\ & & & 0.338 \end{pmatrix}$

21		$\begin{matrix} Cl1 \\ Cl3 \\ C5 \\ Cl2 \end{matrix}\begin{pmatrix} 0.347 & 5.304 & 3.956 & 3.060 \\ & 0.340 & 1.098 & 3.053 \\ & & 0.296 & 2.663 \\ & & & 0.342 \end{pmatrix}$
22		$\begin{matrix} Cl1 \\ Cl3 \\ C5 \\ Cl2 \end{matrix}\begin{pmatrix} 0.341 & 5.295 & 3.956 & 3.048 \\ & 0.340 & 1.107 & 3.044 \\ & & 0.302 & 2.663 \\ & & & 0.341 \end{pmatrix}$
24		$\begin{matrix} C3 \\ C8 \\ C9 \\ C6 \end{matrix}\begin{pmatrix} 0.365 & 4.726 & 4.407 & 3.050 \\ & 0.349 & 0.962 & 2.522 \\ & & 0.351 & 3.059 \\ & & & 0.355 \end{pmatrix}$
25		$\begin{matrix} C8 \\ C3 \\ C2 \\ C5 \end{matrix}\begin{pmatrix} 0.369 & 4.716 & 4.388 & 3.050 \\ & 0.349 & 0.963 & 2.522 \\ & & 0.361 & 3.060 \\ & & & 0.355 \end{pmatrix}$

Averaging the corresponding matrix elements in these 17 sub-matrices we obtained the following averaged matrix of toxicity $\hat{T}_{av}$:

$$\hat{T}_{av} = \begin{matrix} T1 \\ T2 \\ T3 \\ T4 \end{matrix}\begin{pmatrix} 0.395 & 4.803 & 4.053 & 3.053 \\ & 0.350 & 1.207 & 2.743 \\ & & 0.302 & 2.481 \\ & & & 0.318 \end{pmatrix} \quad (3)$$

with the matrix of tolerances (in relative values of the matrix elements):

$$\Delta\hat{T}_{rel} = \begin{pmatrix} 0.162 & 0.105 & 0.144 & 0.131 \\ & 0.133 & 0.202 & 0.121 \\ & & 0.160 & 0.233 \\ & & & 0.116 \end{pmatrix} \quad (4)$$

resulting in following ECSA, the digital toxicophore for the 17 compounds with the toxicity in interval $-1.32 < -\log LC_{50} < 0.30$:

$$T = \begin{pmatrix} 0.395 \pm 0.064 & 4.803 \pm 0.504 & 4.053 \pm 0.585 & 3.053 \pm 0.401 \\ & 0.350 \pm 0.046 & 1.207 \pm 0.244 & 2.743 \pm 0.332 \\ & & 0.302 \pm 0.049 & 2.481 \pm 0.578 \\ & & & 0.318 \pm 0.037 \end{pmatrix} \quad (5)$$

Moving to the compounds with lower activity one may find out that this ECSA (revealed for the most active compounds) does not work for them, but they can be accommodated within this ECSA by allowing larger tolerances. In this way one can get slightly different ECSA's and hence *pharmacophores/toxicophores for different levels of activity*. The activity is thus quantitatively a function of the tolerances. The dynamics of change of the tolerances of different atoms when moving from more active to less active compounds reveals also the role of their flexibility in the change of activity.

Let us illustrate this important feature by an example. The compound **benzaldehyde** with $\log LC_{50} = 1.57$ has the following ECSA:

$$\begin{pmatrix} O1 & 0.51 & 4.79 & 5.06 & 2.89 \\ C3 & & 0.35 & 1.45 & 2.79 \\ C4 & & & 0.31 & 2.41 \\ C6 & & & & 0.31 \end{pmatrix} \quad (6)$$

We see that although this matrix does not fit to the above ECSA for the most active compounds (5), it fails just by some increase of the tolerances, mainly in the distance 5.06 Å between the oxygen atom and the carbon C4, which is out of the limits of the corresponding distance 4.05±0.58 in the most active compounds. This feature comprises the next 7 less active compounds with toxicities $1.9 > \log LC_{50} > 1.4$. It demonstrates the important role of the most active heteroatom (chlorine, oxygen, etc.) in the toxicity. The falling toxicity with the increase of the distance of this active atom to the main skeleton of the molecule may be due to either its out of the limits of efficient docking to the bioreceptor or its poor bonding and hence decay through the metabolic processes. The dependence of the quantitative activity on the tolerances is of general importance in the EC method, and it is used to estimate the activity quantitatively (see below).

As mentioned above, this picture of activity presented by ECSA as the pharmacophore or toxicophore and the dependence of the quantitative activity on the tolerances may be complicated by the presence of out-of-pharmacophore groups that influence the activity either by shielding of or competing with the pharmacophore in its interaction with the bioreceptor [8]. Therefore the presence of the pharmacophore should be considered as just a *necessary condition of activity*, but not a sufficient one. Since it is based on first principles (quantum-chemical calculations) and not involving statistics, the prediction of the pharmacophore in the EC method has a potential of the same reliability as that of the experimental data of the training set, meaning 100% reliability if based on experimental data of highest accuracy. The latter are thus most important in getting the absolutely reliable pharmacophore.

Another important feature that influences the pharmacophore in the EC method is the diversity of the training set. The ECSA (the pharmacophore/ toxicophore) is valid for levels of activity and classes of molecules the ECMC of which were used in its calculation. For the same mechanism of substrate-receptor interaction different classes of compounds should have the same pharmacophore, but if the latter is obtained from one (or a limited number) of classes with active compounds, it may happen that it includes more active sites (atoms) than the minimal necessary, which thus may be absent in active molecules from other classes.

To screen new compounds for the activity under consideration, their ECMC should be calculated, and then the presence of the ECSA should be checked taking into account the above class and activity level limits. If only very active compounds are looked for, the ECSA (5) of the highly active ones should be tried, while for less active compounds another ECSA (with enlarged tolerances) should be involved. We thus have a very flexible system of searching for leads of active compounds. Presently these procedures of matrix comparison are very fast.

The high reliability of pharmacophore identification by the EC method may serve as a basis for revealing novel knowledge [7]. Indeed, if the pharmacophore is a necessary condition of activity, it should be present in all the active compounds. What if there are active compounds that have no pharmacophore (as in some of the above compounds in Table 1)? If the structure of the molecules, as well as the experimental measurement of activity are beyond doubt and the compound under consideration is within the diversity of the training set, the situation calls for a special investigation. The first time we encountered such a case was in the search for the pharmacophore in musk odorant activity [24]. The pharmacophore that was present in several hundred compounds with musk odor from a variety of rather different classes failed in the case of a patented musk tibeten (a benzene derivative). The solution of this controversy was found in the suggestion that this molecule exhibits its odor properties in the form of a dimer formed by two stacking substituted benzene rings; the dimer has the pharmacophore.

In fact bimolecular activity does not require a priori dimerization before the interaction with the bioreceptor. Indeed, it is widely accepted that the drug-receptor interaction, like any other substrate(S)-enzyme(E) interaction, follows the Michaelis-Menten mechanism [25] with pre-equilibrium in the formation of the complex SE before the transformation to the product P:

$$S + E \Leftrightarrow SE \Rightarrow P \quad (7)$$

This mechanism can be extended to the simultaneous action of two molecules. Indeed, if the first molecule S has no pharmacophore, there will not be transformation to the product. Then the equilibrium (Boltzman distribution) allows for the second molecule to enter the intermediate complex:

$$\begin{aligned} &S + E \Leftrightarrow SE \\ &SE + S \Leftrightarrow SSE \Rightarrow P \end{aligned} \qquad (8)$$

If two molecules in a bimolecular docking to the receptor posses the pharmacophore, they may produce the necessary action that triggers the drug activity, albeit with lower probability than a single molecule with the pharmacophore. In the problem of aquatic toxicity to fish under consideration, for instance, the molecule allyl chloride (**II**) with $\log LC_{50} = 1.20$ has no pharmacophore, while a 1,4-dichloro-2-butene (**I**) with $\log LC_{50} = -0.16$ has the pharmacophore. The structure of these two molecules in Fig. 2 shows how two molecules of **II** produce the structure **III** which is similar to **I** and has the pharmacophore; hence the activity of **II** can be explained in this way.

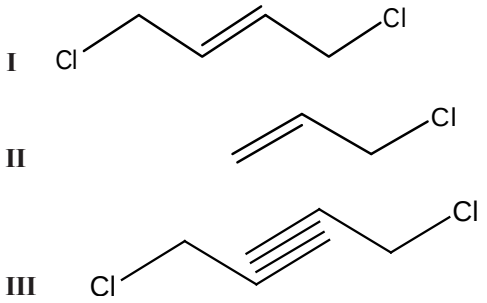


Fig. 2. Illustration of bimolecular activity: two active molecules of allyl chloride (II), which separately have no pharmacophore, by stacking along the double bond produce a bimolecular structure (III) which is similar to 1,4-dichloro-2-butene (I) and has the pharmacophore.

Step 6. In accordance with the EC method techniques [8], the presence of toxicophore is only a necessary condition of activity; the evaluation of activity quantitatively involves the regression analysis mentioned above. For the 17 most active molecules the following relations for biological activity was employed [7, 8]:

$$\log(LC_{50})_i = \log(LC_{50})_{ref} - 2.30259\left\{\frac{E_i - E_{ref}}{kT} + S_i[R]\right\} \qquad (9)$$

where (LC_{50}), and $(LC_{50})_{ref}$ stand for numerical values of activity of the *i*-th compound and the reference compound, respectively, E_i is the relative energy of the lowest energy conformer of the *i*-th compound that contains toxicophore, and $S_i[R]$ is a function of the electronic and geometric parameters of the substrate molecule.

The parameters in the function $S_i[R]$ in Eq. (9) should be obtained from the condition of minimum difference between the calculated toxicity $LC_{50})_i$ and the experimental values $(LC_{50})_i^{\exp}$. The regression analysis was performed over the 17 toxicophore-containing molecules and the correlation coefficient as good as $R^2 = 0.94$ and $R^2 = 0.92$ in a cross-validated procedure was obtained with just three parameters of their electronic structure:

$$S_i[R] = k_1(II^i_{\max} - II^{ref}_{\max}) + k_2(q^i - q^{ref}) + k_3(C^i_{13} - C^{ref}_{13}), \qquad (10)$$

where $II^i_{\max}$ is the largest "interaction index" *II* in *i*-th molecule, $\boldsymbol{q}^i$ is a Mulliken charge of the most electron-negative atom near to the toxicophore atom $T2^i$ (within the radius of R < 2Å of the typical chemical bond length) and C^i_{13} is the matrix element of the ECSA corresponding to the distance (in Å) between the toxicophore atoms $T1^i$ and $T3^i$; T1, T2, T3, and T4 denote consecutively the four atoms in their ECSA (see eq. 3). The following values of the coefficients minimize the expression (10):

$$k_1 = 7.58; \qquad k_2 = -14.74; \qquad k_3 = 1.36. \qquad (11)$$

The theoretical values of toxicities calculated with this set of coefficients and those received in the leave-one-out cross-validation procedure are given in Table 3. The correlation between the experimental and calculated values of -log (LC_{50}) is presented graphically in Fig. 3.

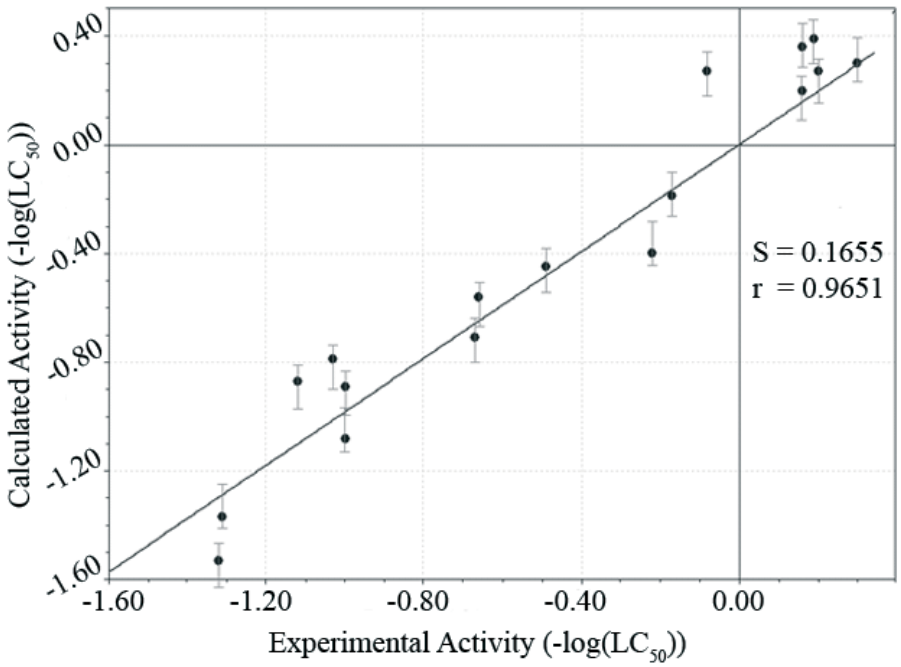


Fig. 3. Calculated vs. experimental -log(LC_{50}) ratio for 17 most active compounds of the training set

Table 3

Calculated with and without cross-validation (C-V) and experimental values of -log(LC_{50}) of 17 most active compounds

	-log(LC_{50})^{exp}	-log(LC_{50})^{calc}	
		Without C-V	With C-V
1(REFER)	0.30	0.30	
2	0,20	0.27	0.28
3	0.19	0.39	0.39
4	0.16	0.36	0.37
5	0.16	0.21	0.21
6	-0.08	0.28	0.29
7	-0.17	-0.20	-0.23
8	-0.22	-0.40	-0.43
9	-0.49	-0.45	-0.42
10	-0.66	-0.56	-0.53
11	-0.67	-0.71	-0.72
17	-1.00	-0.89	-0.85
18	-1.00	-1.08	-1.10
21	-1.03	-0.79	-0.75
22	-1.12	-0.87	-0.83
24	-1.31	-1.38	-1.41
25	-1.32	-1.53	-1.62

With the parameters of the weakly active compounds included a quantitative relationships between the toxicophore parameters and the toxicity can be obtained from a regression calculation that yields the following new correlation formula:

$$\log(LC50)_i = a_0 + a_1 II^i_{\max} + a_2 q^i_{\max} + a_3 C^i_{13} + a_4 C^i_{12} \quad (12)$$

where a_i are the fitting coefficients, $II^i_{\max}$ and $q^i_{\max}$ are the largest values of the ECSA atomic interaction index and atomic charge, respectively, and C^i_{1j} are the corresponding first row matrix elements of the ECSA in Table 2 (the distance between the most active atom and another one in the toxicophore).

The results are presented in Table 4: **I** and **II** are the fittings obtained over the 17 most active compounds with three and five regression parameters, respectively, and **III** is a fitting received over 21 compounds that include the weakly active ones. While correlation **II** is somewhat better than **I** due to the additional two parameters, correlation **III** allows one to screen and predict also weak toxicities with $\log LC_{50} > 1.4$. For strong toxicity with $\log LC_{50} < 1.4$ correlation **I** is quite reasonable. This indicates the significance of the parameters of the active atom **1** in reducing the toxicity.

Table 4

Regression coefficients for three types of correlation (eqs. 10, 12)

Regression	a_0	a_1	a_2	a_3	a_4	R^2
I	-	7.58	14.74	1.36	-	0.94
II	5.18	2.83	5.82	0.39	0.31	0.97
III	1.6	3.29	6.49	0.46	0.54	0.86

Conclusions

Using the electron-conformational method we have shown that aquatic toxicity to fish is due to a special functional group, the toxicophore, a combination of four atomic sites with special electronic and topology characteristics represented numerically as the sub-matrix of activity, ECSA. An important feature of the latter is the set of tolerances in the values of the matrix elements which control the quantitative value of activity. It was shown that by varying the tolerances one can obtain different toxicophores for different levels of activity. A distinguished feature of this approach is that it does not employ arbitrary descriptors and statistics at the level of toxicophore identification, thus making the latter void of artifacts and chance correlations and fully reliable (at the level of reliability of the experimental data in the training set). This led us to suggest a bimolecular mechanism of activity for allyl chloride. The toxicophore for 17 most active compounds yields a good correlation ($R^2 = 0.94$) with the experimental activities using only three electronic structure parameters.

References

[1] Guner O.F. Ed. Pharmacophore Perception, Development, and Use in Drug Design, International University Line: La Jolla, 2000.

[2] Kubinyi, H.; Folkers, G.; Martin, Y. C. 3D QSAR in Drug Design. Volume 3, Recent Advances; Kluwer Academic Publishers: New York, 2002.

[3] Marshall G.R. In: Wolff M.E., Ed. Burger's medicinal chemistry and drug discovery: principles and practice. 5th ed. Vol. 1. New York: John Wiley, 1995, 573-659.

[4] O. Ivanciuc, Internet Electron. J. Mol. Design, **2,** 195–208 (2003).

[5] Zhong–Sheng Yi, Shu–Shen Liu, Internet Electron. J. Mol. Design, **4,** 835–849 (2005).

[6] Verhaar, H.J.M.. Leeuwen, C.J, Hemiens. J.L.M., Chemosphere, **25**, 471 -491(1992).

[7] I. B. Bersuker, J. Comp.-Aid. Mol. Des., submitted for publication.

[8] I. B. Bersuker. Current Pharmaceutical Design., **9,** 1575-1606 (2003).

[9] Bersuker I. B., Dimoglo A. S., In Reviews in Computational Chemistry; Lipkowitz K.B., Boyd D.B. Eds.; VCH: New York, 1991; vol.2, pp.423-60.

[10] Bersuker, I. B.; Bahceci, S.; Boggs, J. E.; Pearlman, R. S. , J. Computer-aided Mol. Design **13,** 419–434 (1999).

[11] Bersuker, I. B.; Bahceci, S.; Boggs, J. E., J. Chem. Inf. Comput. Sci., **40** (2000)1363–1376.

[12] Rosines, E.; Bersuker, I. B.; Boggs, J. E., Quan. Struct.-Act. Relat. **2001**, *20*, 327–334.

[13] A. H. Makkouk, I. B. Bersuker and J. E. Boggs, Int. J. Pharm. Med., **18**, 81-89 (2004).

[14] A.V. Marenich, Pei-Han Yong, I. B. Bersuker and J. E. Boggs, J. Chem. Inf. Mod., submitted.

[15] Katritzky, A.R., Tatham, D.B. and Maran U., J. Chem. Inf. Comput. Sci.,**41,** 1162-1176 (2001).

[16] Dearden. J.C. and Raevsky, O.A., SAR and QSAR in Environm. Research, **15,** 433-448 (2004).

[17] Halgren T. A., J. Comp. Chem. **1996**, *17*, 490–519.

[18] Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. J. Phys. Chem. **1996**, *100*, 16385–16398.

[19] *Spartan'02*; Wavefunction, Inc.: Irvin, CA. Except for molecular mechanics and semi-empirical models, the calculation methods used in Spartan'02 have been documented in: Kong, J.; White, C. A.; Krylov, A. I.; Sherrill, D.; Adamson, R. D.; Furlani, T. R.; Lee, M. S.; Lee, A. M.; Gwaltney, S. R.; Adams, T. R.; Ochsenfeld, C.; Gilbert, A. T. B.; Kedziora, G. S.; Rassolov, V. A.; Maurice, D. R.; Nair, N.; Shao, Y.; Besley, N. A.; Maslen, P. E.; Dombroski, J. P.; Daschel, H.; Zhang, W.; Korambath, P. P.; Baker, H.; Byrd, E. F. C.; Voorhis, T. V.; Oumi,

M.; Hirata, S.; Hsu, C.-P.; Ishikawa, N.; Florian, J.; Warshel, A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M.; Pople, J. A. Q-Chem 2.0: A High-Performance Ab Initio Electronic Structure Program Package. J. Comput. Chem. 2000, 21, 1532–1548.
[20] Stewart, J. J. P. , J. Comp. Chem. **1989**, *10*, 209–220.
[21] Basch, H.; Viste, A.; Gray, H. B., Theoret. chim. Acta (Berl.) **1965**, *3*, 458–464.
[22] Bondi, A., J. Phys. Chem. **1964**, *68*, 441–451).
[23] Slater, J. C., J. Chem. Phys. **1964**, *41*, 3199–3204).
[24] Bersuker I.B., Dimoglo A.S., Gorbachov M.Yu., Vlad P.F., Pesaro M. *New J. Chem.,* **15**, 307-20 (1991).
[25] Metzler, D. E. Biochemistry: The Chemical Reactions of Living Cells; Academic Press: New York,1977.