communications in mathematical
match
and in computer chemistry

no. 44, October 2001

MATCDY (44) 201 - 214 (2001)

ISSN 0340 - 6253

QSAR MODELING OF THE ANTICONVULSANT ACTIVITY OF PHENYLACETANILIDES WITH PRECLAV (PROPERTY EVALUATION BY CLASS VARIABLES)[#]

Laszlo Tarko[a,*] and Ovidiu Ivanciuc[b,§,*]

[a] Institute of Organic Chemistry, Splaiul Independentei 202B, PO Box 15-258, Bucharest, Romania

[b] EST Lead Informatics, AstraZeneca R&D Mölndal, S-43183 Mölndal, Sweden
E-mail: ltarko@cco.ro ; ivanciuc@netscape.net

Abstract. Quantitative structure-activity relationships (QSAR) models for the anticonvulsant activity of 30 monosubstituted phenylacetanilides were established with the PRECLAV (PRoperty Evaluation by CLAss Variables) program. The QSAR models developed with PRECLAV allow accurate computation of the physical, chemical, biological and toxicological properties of organic compounds using simple constitutional, topological, electrostatic, quantum, and grid (field) descriptors; standard chemistry packages are used to optimize the molecular geometry and for semi-empirical quantum computations. A heuristic algorithm selects the best multiple linear regression (MLR) equation according to the highest leave-one-out cross-validation correlation coefficient. The best MLR equation with six theoretical descriptors has $s = 0.14$, $r^2 = 0.789$ for calibration and $r_{LOO}^2 = 0.700$ for cross-validation, while the best QSAR model with ten descriptors has $s = 0.06$, $r^2 = 0.962$ and $r_{LOO}^2 = 0.901$. Both these QSAR models are much better than previous MLR models published for the same data set, indicating that the PRECLAV method is very efficient in detecting structure-activity correlations with good predictive power.

[#] Dedicated on the occasion of the 70th birthday to Professor Alexandru T. Balaban, one of the early developers and promoters of topological indices as QSAR and QSPR descriptors.

[§] To whom correspondence should be addressed.

INTRODUCTION

During the last twenty years quantitative structure-property relationships (QSPR) and quantitative structure-activity relationships (QSAR) models have gained an extensive recognition in physical, organic, analytical, pharmaceutical and medicinal chemistry, biochemistry, chemical engineering and technology, toxicology, and environmental sciences [1-9]. The widespread use of QSPR and QSAR models comes from the development of novel structural descriptors and statistical equations relating various physical, chemical, and biological properties to the chemical structure. The main hypothesis in the QSPR and QSAR approach is that all properties (physical, chemical, and biological) of a chemical substance are statistically related to its molecular structure. The success of the QSPR and QSAR approach can be explained by the insight offered into the structural determination of chemical properties and biological activities, and the possibility to estimate the properties of new chemical compounds without the need to synthesize and test them. These molecular design techniques, which significantly reduce the cost and time involved in obtaining compounds with desired properties, were applied to a wide range of properties, such as melting and boiling temperature, molar heat capacity, standard Gibbs energy of formation, vaporization enthalpy, refractive index, density, aqueous solubility, 1-octanol/water partition coefficient, solvation free energy, receptor binding affinities, pharmacological activities, and enzyme inhibition constants.

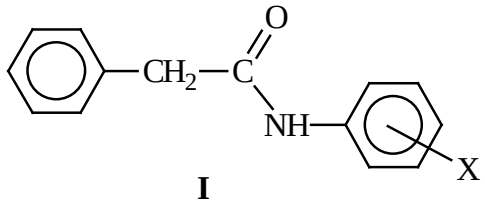
The investigation of large and diverse molecular data bases was made possible by the advent of general QSPR/QSAR programs [8,10,11], such as CoMFA [5], SOMFA [6], HINT [7], CoRSA [9], ADAPT [12,13], OASIS [14,15], SciQSAR [16], CODESSA [17-23], Cerius^2 [24], TSAR [25], and PRECLAV [26,27], which integrate the computation of structural descriptors with the generation of structure-property models. These programs compute more than one thousand structural descriptors from five classes: constitutional, graph theoretic and topological indices, geometrical, electrostatic, and quantum-chemical descriptors. Using statistical methods, such as multiple linear regression (MLR), PCA, PLS, or neural networks, the best descriptors are selected in the final structure-property model. In this paper we present an application of PRECLAV (PRoperty Evaluation by CLAss Variables) approach

for generating QSAR models of the anticonvulsant activity for monosubstituted phenylacetanilides.

MOLECULAR DATA BASE AND QSAR METHOD

Data Base

The data base used as input by PRECLAV consists of 30 monosubstituted phenylacetanilides with the general formula **I**, presented in Table 1 together with their anticonvulsant activity log $1/ED_{50}$ taken from the literature [28]. The anticonvulsant activity ED_{50} (mol kg^{-1}) is evaluated by the maximal electroshock seizure method in mice.



I

TABLE 1. Structure of monosubstituted phenylacetanilides, experimental, calculated, and residual (log $1/ED_{50,exp}$ – log $1/ED_{50,calc}$) anticonvulsant activity computed with Eqs. (4) and (5).

No	X	log $1/ED_{50}$				
		exp	calc Eq. (4)	res Eq. (4)	calc Eq. (5)	res Eq. (5)
1	H	3.77	3.57	0.20	3.64	0.13
2	*m*-Me	3.75	3.65	0.10	3.71	0.04
3	*m*-Et	3.67	3.58	0.08	3.60	0.07
4	*m*-F	3.34	3.52	–0.19	3.35	–0.01
5	*m*-Cl	3.40	3.37	0.03	3.41	–0.01
6	*m*-Br	3.32	3.23	0.09	3.32	–0.00
7	*m*-I	2.64	2.64	0.00	2.64	0.00
8	*m*-CF_3	2.84	2.83	0.01	2.85	–0.01
9	*m*-OH	3.58	3.64	–0.06	3.54	0.04
10	*m*-NH_2	3.81	3.75	0.06	3.93	–0.12
11	*m*-NHMe	4.03	3.82	0.21	4.00	0.03
12	*m*-NHEt	3.91	3.83	0.08	3.85	0.06
13	*m*-OMe	3.55	3.58	–0.03	3.53	0.02
14	*m*-CN	3.44	3.46	–0.02	3.51	–0.07
15	*m*-NO_2	3.62	3.68	–0.06	3.63	–0.01
16	*m*-COMe	3.95	3.65	0.30	4.00	–0.04
17	*m*-OAc	3.48	3.58	–0.10	3.49	–0.01

TABLE 1 (continued)

18	*m*-OEt	3.42	3.56	–0.14	3.46	–0.04
19	*m*-OSO_2Me	3.77	3.83	–0.06	3.74	0.03
20	*p*-Me	3.26	3.67	–0.41	3.42	–0.17
21	*p*-F	3.49	3.49	–0.00	3.52	–0.03
22	*p*-OH	3.72	3.64	0.08	3.73	–0.01
23	*p*-OMe	3.78	3.70	0.08	3.79	–0.01
24	*p*-COMe	3.51	3.71	–0.21	3.44	0.06
25	*o*-F	3.48	3.46	0.02	3.41	0.07
26	*o*-OH	3.33	3.42	–0.09	3.34	–0.01
27	*o*-NH_2	3.40	3.46	–0.06	3.40	–0.00
28	*o*-OMe	3.43	3.41	0.02	3.44	–0.01
29	*o*-NO_2	3.29	3.29	–0.00	3.24	0.05
30	*o*-COMe	3.41	3.34	0.07	3.44	–0.03

Previous QSAR Models

A Hansch-type [1-3] QSAR analysis of these 30 compounds was performed with the aid of the following descriptors [28]: log *P*, the octanol-water partition coefficient; σ, the Hammett electronic constant; I_p, an indicator variable which takes the value 1 for *p*-derivatives and 0 for other compounds; E_S, the Taft steric constant for *o*-derivatives; *R*, the electronic parameter for *o*-derivatives. The Hansch equations with four, five, and six independent variables are:

$$\log 1/ED_{50} = 2.280 + 0.264(\log P)^2 + 1.222(\log P) - 0.161\sigma - 0.079I_p \quad (1)$$

$$n = 30 \quad r = 0.700 \quad s = 0.228 \quad F = 5.99$$

$$\log 1/ED_{50} = 2.311 + 0.290(\log P)^2 + 1.309(\log P) - 0.135\sigma - 0.157I_p + 0.404E_S \quad (2)$$

$$n = 30 \quad r = 0.800 \quad s = 0.195 \quad F = 10.05$$

$$\log 1/ED_{50} = 2.478 + 0.276(\log P)^2 + 1.229(\log P) - 0.353\sigma - 0.223I_p + 0.278E_S + 0.621R \quad (3)$$

$$n = 30 \quad r = 0.855 \quad s = 0.172 \quad F = 7.83$$

Molecular Modeling

In the present investigation, the chemical structures were generated with HyperChem [29], the geometry optimization was performed with MOPAC 7 using the semiempirical quantum method AM1 [30] and the QSAR models were computed with PRECLAV [26,27].

Structural Descriptors

The MOPAC 7 output files are used by PRECLAV program to compute five classes of structural descriptors: constitutional, topological indices and molecular graph invariants, geometrical, quantum, bond indices, and field (grid) descriptors. We will briefly present the main structural descriptors used in the QSAR analysis, with more emphasis on those that are specific to PRECLAV.

Constitutional descriptors represent a simple but efficient way of discriminating between chemical structures: molecular mass; number of H, C, N, O, halogens, S, and P; percentage of H, C, N, O, halogens, S, and P; cyclomatic number; number of bonds sorted, according to the quantum bond order, into weak, single, aromatic, double, and triple. The structural information from the molecular graph is used by PRECLAV to compute various topological indices [8,31-35].

The molecular geometry, optimized with MOPAC, is the source of several important geometric descriptors: van der Waals molecular volume and surface area; spherical shape indices; gravitational indices; positive, neutral, and negative charged surface area. The MOPAC output files offer the possibility to compute a large number of quantum indices: minimum, maximum, and average atomic partial charges; dipole moment; polarity parameter; electronic topological indices; HOMO and LUMO energies; Fukui reactivity indices; free valence indices.

When the QSAR modeling is performed for congeneric series of compounds, PRECLAV uses the atoms in the common skeleton to compute bond and field (grid) descriptors. All molecules are aligned by superimposing the common atoms. The atomic mass, volume, and quantum bond order are used to compute several bond indices for the bonds from the common skeleton. In order to calculate the field (grid) descriptors, a three-dimensional

regularly spaced grid is superposed over the aligned set of molecules in such a way that the grid box dimensions are extended several Å in each direction from the coordinates of every molecule, in order to include the van der Waals surface of all molecules. Each grid axis is divided into five equal segments, in such a way that the entire box is divided into 125 cells. The probe atom (H^+ in our present computations) is successively positioned in the center of each cell, and electrostatic forces (attractive, repulsive, and sum) are computed by considering the AM1 atomic charges of all atoms in the investigated molecule. For each cell center one computes also a set of steric parameters representing the maximum and average angle (parallax) between two lines that connect the cell center with certain atom pairs.

Multiple Linear Regression Model

Before generating the multiple linear regression (MLR) models, PRECLAV makes a descriptor selection by discarding those that are poorly correlated with the investigated property. The removing of all descriptors having a squared correlation coefficients with log $1/ED_{50}$ less than 0.063 reduced the initial pool to 304 significant structural indices; the use of a higher value, namely 0.1261, decreases to 127 the number of structural descriptors that are tested in MLR equations. These descriptors were used in the following heuristic algorithm to generate the best MLR equation:

(*a*) All orthogonal pairs of structural descriptors are selected by PRECLAV from the initial set. Two descriptors are considered orthogonal if their squared intercorrelation coefficient r_{ij}^2 is lower than 0.126.

(*b*) PRECLAV uses the pairs of orthogonal descriptors to compute all possible biparametric regression equations. The best 400 such MLR equations are retained for further use.

(*c*) To an MLR model containing *n* descriptors a new descriptor is added to generate a model with *n*+1 descriptors if the new descriptor is not significantly correlated with the previous *n* descriptors (squared intercorrelation coefficient lower than 0.64).

Step (*c*) is repeated until MLR models with a given maximum number of descriptors are obtained.

Model Validation

QSAR correlations can be observed not only because a causal relationship exists between a set of descriptors and a property, but also due to statistical bias resulting from errors in determining structural descriptors, experimental errors in measuring the property, or even due to chance alone. Model validation techniques are needed in order to distinguish between true and random correlations and to estimate the predictive power of the model. Although the QSAR equations developed with PRECLAV are obtained by selection of descriptors from a large pool, several descriptor selection techniques are used in order to minimize the possibility of chance correlations. In a first step, from the initial pool of descriptors, PRECLAV eliminates descriptors poorly correlated with the anticonvulsant activity ED_{50}, thus greatly reducing the dimensionality of the problem - that of finding a QSAR equation with a good predictive power. Then, as described in the previous section, a heuristic algorithm selects only quasi-orthogonal groups of descriptors that are tested for correlation with the anticonvulsant activity. This selection algorithm ensures that the probability of obtaining a chance correlation is low, and maintains a reasonable searching time. Finally, the leave-one-out (LOO) cross-validation procedure is applied to each and every MLR equation in order to estimate the prediction power of anticonvulsant activity QSAR equations. The predictive ability of a QSAR equation is estimated with the LOO Pearson and rank (Kendall) correlation coefficients r_{LOO} and r^{K}_{LOO}. The equation with the highest predictive power is considered to be the one with the highest value for the product $r_{LOO}^{2} \times r^{K}_{LOO}$; this QSAR model is selected to predict the activity of novel, not yet tested compounds.

The present investigation uses all 30 monosubstituted phenylacetanilides to develop QSAR models. However, the main utility of QSAR equations is in making predictions for new compounds that are not used in model calibration. PRECLAV offers a unique feature, the Class function, used for descriptor selection in order to choose only those structural descriptors that are relevant both for calibration and prediction molecules. When PRECLAV is used for prediction, both calibration and prediction sets of molecules are entered into the program and their structural descriptors are computed; we have to mention that while for the calibration set of compounds the experimental value of the investigated property is known, for the molecules in the prediction set the experimental value is not known and PRECLAV uses

the best QSAR equation to predict it. During the selection of structural descriptors a Class value is computed for each descriptors *X* from the average and standard deviation values of *X* for the calibration and prediction compounds. A structural descriptor is selected for further use in QSAR models depending on its Class function value.

RESULTS AND DISCUSSION

After computing the structural descriptors for the 30 monosubstituted phenylacetanilides, PRECLAV performs the descriptor selection and generation of QSAR equations. In Table 2 we present the notations and definitions of all structural descriptors that are significant for our investigation.

TABLE 2. Notation and definition of the structural descriptors from the QSAR models generated with PRECLAV

Notation	Definition of structural descriptors
mam	molecular mass / number of atoms
nri	number of I atoms
phi	mass percent of H
dns	molecular mass / molecular volume
vlm	molecular volume / molecular mass
vln	molecular volume / number of atoms
spm	molecular surface area / molecular mass
ifu	symmetry index
xsi	maximum net charge for a I atom
nsi	minimum net charge for a I atom
hia	mass percent of H × average charge for H atoms
xsc	maximum net charge for a C atom
mas	molecular mass
poz	total positive charged molecular surface area
nsh	minimum net charge for a H atom
xsf	maximum net charge for a F atom
avi	average free valence of I atoms
cns	number of CN single bonds and weak bonds / number of bonds
fln	mass percent of F × minimum net charge for a F atom
ift	roughness index

TABLE 2 (continued)

pio	mass percent of I
lgp	approximate logP computed with the formula lgp = 0.577 pte + 0.001 igd – 0.989 nro – 5.935 paz
pte	average absolute net atomic charge × average bond order
igd	gravitation index for all pairs of bonded atoms
nro	number of O atoms
paz	mass percent of N
iox	mass percent of I × maximum net charge for a I atom
nsf	minimum net charge for a F atom
A(*n*)	sum of attractive electrostatic forces for grid point *n*
F(*n*)	sum of all (attractive and repulsive) electrostatic forces for grid point *n*
M(*n*)	average parallax for grid point *n*
P(*n*)	maximum parallax for grid point *n*
or_n	maximum bond order for single bonds connected to atom *n*
sr_n	maximum atomic charge difference for single bonds connected to atom *n*

Because it is important to have a reference for the evaluation of MLR models, we give here the squared correlation coefficient in monolinear regression equations for a selected set of descriptors: hia, 0.500; mam, 0.455; dns, 0.442; vlm, 0.437; spm, 0.423; vln, 0.418; phi, 0.408; ifu, 0.319; nri, 0.311; xsi, 0.311; nsi, 0.311; iox, 0.311; avi, 0.311.

TABLE 3. Structural descriptors and statistical indices for the best QSAR equations (according to the calibration correlation coefficient) selected by PRECLAV for the set of 127 significant structural indices

Structural descriptors	*r*	*s*	F	r_{LOO}^2	r^K_{LOO}
hia F(88)	0.778	0.18	21.5	0.163	0.430
dns xsc A(92)	0.839	0.16	21.3	0.581	0.531
mas M(2) A(92) xsc	0.870	0.14	20.3	0.657	0.595
mas poz A(92) vln nsh	0.882	0.14	17.4	0.659	0.582
mas poz A(92) vln nsh xsf	0.895	0.13	16.1	0.684	0.559

The first experiment was made by considering only the structural descriptors having a squared correlation coefficients with log $1/ED_{50}$ less than 0.1261, which reduces the initial pool to 127 significant structural indices. The structural descriptors and statistical indices for the best QSAR equations (according to the calibration correlation coefficient) with up to six descriptors are presented in Table 3. Compared with the correlation coefficients of the

individual descriptors, even the biparametric QSAR from Table 3 represents a significant improvement; also, its statistical indices are better than those of the Hansch-type models from Eqs. (1) and (2). With three descriptors, the second QSAR from Table 3 surpasses Eq. (3) demonstrating that the PRECLAV approach is able to generate good QSAR models; also, the Pearson and Kendall LOO coefficients are high (with the exception of the biparametric model), showing that these equations can provide reliable predictions. We have to mention that r_{LOO}^2 steadily increases with the number of descriptors, showing that the predictive power of the QSAR models increases, and there is no overfiting. The equation with the highest predictive power (highest value for the product $r_{LOO}^2 \times r^K_{LOO}$) is:

$$\log 1/ED_{50} = 10.1446 - 4.0793\ \text{xsc} - 6.1638\ \text{avi} - 1.4332\ \text{A(92)} \quad (4)$$
$$- 0.0094\ \text{mas} - 9.2412\ \text{M(2)} + 0.1434\ \text{mam}$$
$$s = 0.14 \quad r^2 = 0.789 \quad F = 15.0 \quad r^K = 0.660 \quad r_{LOO}^2 = 0.700 \quad r^K_{LOO} = 0.595$$

The QSAR model from Eq. (4) has much better statistics than Eq. (3), that has also six variables. In Table 1 we present the log $1/ED_{50}$ values computed with Eq. (4) together with the corresponding residuals. For the majority of compounds, the residuals are small, showing that Eq. (4) has a fairly good statistical quality. Only two compounds have residuals larger than 2*s*, namely **16** (*m*-COMe) with 0.30 and 20 (*p*-Me) with –0.41; however, they are not outliers, since the commonly accepted characterization for outliers defines them as compounds with absolute residuals greater than 3*s*.

Table 4. Structural descriptors and statistical indices for the best QSAR equations (according to the calibration correlation coefficient) selected by PRECLAV for the set of 304 significant structural indices

Structural descriptors	*r*	*s*	F	r_{LOO}^2	r^K_{LOO}
hia A(58)	0.801	0.17	25.1	0.084	-0.025
mam xsc A(58)	0.840	0.16	21.6	0.588	0.563
dns F(88) cns fln	0.880	0.14	22.2	0.664	0.545
dns F(88) cns fln ift	0.897	0.13	20.5	0.699	0.664
hia A(58) pio xsc F(113) P(64)	0.914	0.12	20.3	0.785	0.614
mam xsc A(58) F(113) P(64) nri M(70)	0.937	0.10	23.5	0.810	0.655
hia A(58) pio xsc F(113) P(64) M(68) F(109)	0.965	0.08	36.7	0.875	0.655
hia A(58) pio xsc F(113) P(64) M(68) A(68) lgp	0.973	0.07	41.9	0.891	0.715
hia A(58) pio xsc F(113) P(64) M(67) F(109) or_3 sr_4	0.982	0.06	52.6	0.902	0.706

The second experiment was made by considering only the structural descriptors having a squared correlation coefficients with log $1/ED_{50}$ less than 0.063, which reduces the initial pool to 304 significant structural indices. The structural descriptors and statistical indices for the best QSAR equations (according to the calibration correlation coefficient) with up to ten descriptors are presented in Table 4. As a general trend, both F and r_{LOO}^2 increase with the number of descriptors indicating a significant increase in the predictive power of the QSAR models. It is interesting to mention that the descriptors from the QSAR with six descriptors, namely hia, A(58), pio, xsc, F(113), and P(64), are used also in the QSAR models with eight, nine, and ten descriptors, indicating that this group of indices is important in modeling the anticonvulsant activity of monosubstituted phenylacetanilides. The equation with the highest value for the product $r_{LOO}^2 \times r^K_{LOO}$, that has also the highest predictive power, is:

$$\log 1/ED_{50} = 0.9172 - 0.174\ \text{mam} - 3.2783\ \text{A(58)} - 1.3221\ \text{xsc} \qquad (5)$$
$$+ 5.1735\ \text{F(113)} - 7.5902\ \text{iox} + 1.7525\ \text{P(64)} - 3.406\ \text{M(72)}$$
$$- 35.7998\ \text{F(109)} + 1.987\ \text{nsf} - 17.7329\ \text{F(96)}$$
$$s = 0.06 \quad r^2 = 0.962 \quad F = 51.2 \quad r^K = 0.899 \quad r_{LOO}^2 = 0.901 \quad r^K_{LOO} = 0.821$$

Compared with Eq. (4), the above QSAR model exhibits significant improvements of all statistical indices, especially for those related to the LOO cross-validation. Despite the larger number of structural descriptors in Eq. (5), there is no evidence of overfiting, as indicated by the high values of r_{LOO}^2 and r^K_{LOO}. In the last two columns of Table 1 we present the log $1/ED_{50}$ values computed with Eq. (5) together with the corresponding residuals. While compound **16** has a small residual, –0.04, **20** has a fairly large one, –0.17, at the border between 2*s* and 3*s*. However, with the exception of **20**, the remaining compounds have very small residuals, indicating that this QSAR model can be confidently used to predict the anticonvulsant activity for novel monosubstituted phenylacetanilides.

CONCLUSIONS

Using the PRECLAV (PRoperty Evaluation by CLAss Variables) program we have obtained multiple linear regression QSAR the anticonvulsant activity of 30 monosubstituted phenylacetanilides using only theoretical descriptors computed from the molecular structure. Approximately one thousand constitutional, topological, electrostatic, quantum, and grid (field) descriptors are computed by PRECLAV using the output offered by MOPAC 7 quantum chemistry software. The PRECLAV heuristic algorithm for descriptor selection proved to be very efficient in generating QSAR models with a high predictive power, as indicated by the leave-one-out cross-validation statistics. The best QSAR equation with six theoretical descriptors has s = 0.14, r^2 = 0.789 for calibration and r_{LOO}^2 = 0.700 for cross-validation, while the best QSAR model with ten descriptors has s = 0.06, r^2 = 0.962 and r_{LOO}^2 = 0.901. Both these QSAR models are much better than previous QSAR models published for the same data set, indicating that PRECLAV is very efficient in detecting structure-activity correlations with good predictive power.

REFERENCES

[1] C. Hansch and T. Fujita, *J. Am. Chem. Soc.*, **1964**, *86*, 1616-1626.

[2] C. Hansch, *Acc. Chem. Res.*, **1969**, *2*, 232-239.

[3] C. Hansch, and A. Leo, *Exploring QSAR. Fundamentals and Applications in Chemistry and Biology*, American Chemical Society, Washington, DC, 1995.

[4] H. Kubinyi, *Drug Discovery Today*, **1997**, *2*, 457-467.

[5] R. D. Cramer, D. E. Patterson, and J. D. Bunce, *J. Am. Chem. Soc.*, **1988**, *110*, 5959-5967.

[6] D. D. Robinson, P. J. Winn, P. D. Lyne, and W. G. Richards, *J. Med. Chem.*, **1999**, *42*, 573-583

[7] G. E. Kellogg, S. F. Semus, and D. J. Abraham, *J. Comput.-Aided Mol. Design*, **1991**, *5*, 545-552.

[8] O. Ivanciuc, 3D QSAR Models. In *QSPR/QSAR Studies by Molecular Descriptors* (M. V. Diudea, Ed.), Nova Science, Huntington, N.Y., 2001.

[9] O. Ivanciuc, T. Ivanciuc, and D. Cabrol-Bass, *Analysis*, **2000**, *28*, 637-642.

[10] O. Ivanciuc and J. Devillers, Algorithms and Software for the Computation of Topological Indices and Structure-Property Models. In: *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds.: J. Devillers and A. T. Balaban. Gordon and Breach Science Publishers, The Netherlands, 1999, pp 779-804.

[11] A. R. Katritzky, U. Maran, V. S. Lobanov, and M. Karelson, *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 1-18.

[12] A. J. Stuper and P. C. Jurs, *J. Chem. Inf. Comput. Sci.*, **1976**, *16*, 99-105.

[13] E. S. Goll and P. C. Jurs, *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 974-983.

[14] O. Mekenyan, S. Karabunarliev, and D. Bonchev, *Computers Chem.*, **1990**, *14*, 193-200.

[15] O. G. Mekenyan, S. H. Karabunarliev, J. M. Ivanov, and D. N. Dimitrov, *Comput. Chem.*, **1994**, *18*, 173-187.

[16] SciQSAR, SciVision, Inc., 200 Wheeler Road, Burlington, MA 01803, U.S.A., www http://www.scivision.com.

[17] R. Murugan, M. P. Grendze, J. E. Toomey Jr., A. R. Katritzky, M. Karelson, V. S. Lobanov, and P. Rachwal, *CHEMTECH*, **1994**, *24*, 17-23.

[18] A. R. Katritzky, V. S. Lobanov, and M. Karelson, *Chem. Soc. Rev.*, **1995**, 279-287.

[19] M. Karelson, V. S. Lobanov, and A. R. Katritzky, *Chem. Rev.*, **1996**, *96*, 1027-1043.

[20] A. R. Katritzky, L. Mu, V. S. Lobanov, and M. Karelson, *J. Phys. Chem.*, **1996**, *100*, 10400-10407.

[21] A. R. Katritzky, V. S. Lobanov, and M. Karelson, *J. Chem. Inf. Comput. Sci.*, **1998**, *38*, 28-41.

[22] O. Ivanciuc, T. Ivanciuc, and A. T. Balaban, *Tetrahedron*, **1998**, *54*, 9129-9142.

[23] O. Ivanciuc, T. Ivanciuc, P. A. Filip, and D. Cabrol-Bass, *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 515-524.

[24] Cerius^2 3.0 QSAR+, Molecular Simulations Inc., 9685 Scranton Road, San Diego, CA 92121-3752.

[25] TSAR, Oxford Molecular Ltd., The Medawar Centre, Oxford Science Park, Oxford, OX4 4GA, U.K., www http://www.oxmol.com.

[26] PRECLAV (PRoperty Evaluation by CLAss Variables), L. Tarko and S. Calafeteanu, Institute of Organic Chemistry, Splaiul Independentei 202B, PO Box 15-258, Bucharest, Romania.

[27] L. Tarko and S. Calafeteanu, *Rev. Chim.* (*Bucharest*), **1998**, *49*, 169.

[28] C. Yamagami, N. Takao, M. Tanaka, K. Horisaka, S. Asada, and T. Fujita, *Chem. Pharm. Bull.*, **1984**, *32*, 5003-5009.

[29] HyperChem 5.1, Hypercube, Inc., Florida Science and Technology Park, 1115 N.W. 4th Street Gainesville, Florida 32601, U.S.A., www http://www.hyper.com.

[30] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc*., **1985**, *107*, 3902-3909.

[31] L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure-Activity Analysis*, Research Studies Press, Letchworth, 1986.

[32] O. Ivanciuc and A. T. Balaban, Graph Theory in Chemistry. In: *The Encyclopedia of Computational Chemistry*, Eds.: P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, and P. R. Schreiner. John Wiley & Sons, Chichester, 1998, pp. 1169-1190.

[33] O. Ivanciuc and A. T. Balaban, The Graph Description of Chemical Structures. In: *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds.: J. Devillers and A. T. Balaban. Gordon and Breach Science Publishers, The Netherlands, 1999, pp. 59-167.

[34] O. Ivanciuc, T. Ivanciuc, and A. T. Balaban, Vertex- and Edge-Weighted Molecular Graphs and Derived Structural Descriptors. In: *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds.: J. Devillers and A. T. Balaban. Gordon and Breach Science Publishers, The Netherlands, 1999, pp. 169-220.

[35] O. Ivanciuc and T. Ivanciuc, Matrices and Structural Descriptors Computed from Molecular Graph Distances. In: *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds.: J. Devillers and A. T. Balaban. Gordon and Breach Science Publishers, The Netherlands, 1999, pp. 221-277.