STUDIA UNIVERSITATIS BABEŞ-BOLYAI, CHEMIA, XLVIII, 1, 2003

CORRELATING ABILITY OF CLUJ TYPE INDICES

GABRIEL KATONA[a] and MIRCEA V. DIUDEA[b]

[a,b] *Faculty of Chemistry and Chemical Engineering*
Babes-Bolyai University, 3400 Cluj, Romania

ABSTRACT. This paper reviewes the most important results in using Cluj type indices in correlating tests on several sets of organic compounds (alkanes, cycloalkanes, dipeptide ACE inhibitors, substituted 3-(phthalimidoalkyl)-pyrazolin-5-ones, aromatase inhibitors, nitrogen-containing compounds and poly-chlorinated bipheniles).

INTRODUCTION

QSPRs/ QSARs (Quantitative Structure-Property Relationships/ Quantitative Structure-Activity Relationships) link in a quantitative manner the physico-chemical or biological properties of chemicals with their molecular structure.[1]

Some molecular properties (*i.e.*, those of which numerical value vary with changes in the molecular structure) such as the normal boiling point, critical parameters, viscosity, solubility, retention chromatographic index, are often used for characterizing chemicals in databases. However, a certain property is not always available in tables or other reference sources. It is just the case of newly synthesized compounds. As a consequence, methods of evaluating physico-chemical properties from the structural features of organic molecules become very important.

In this work several correlating results, both QSPRs and QSARs, by using Cluj type topological indices are reported, with the aim to demonstrate the capability of our indices to model the molecular properties or activities of organic compounds.

Cluj indices are calculated on the ground of the Cluj matrices[2-9].

Cluj type indices

The graph-theoretical descriptors *CJ* and *CF* represent the theoretical ground for counting the fragmental property indices. They are vertex sets defined by:

$$CJ_{i,j,p} = \{v \mid v \in V(G);\ di(G)_{v,i} < di(G)_{v,j};\ \text{and}\ \exists\ w \in W_{v,i},\ V(w) \cap V(p) = \{i\}\} \quad (1)$$

$$CF_{i,j,p} = \{v \mid v \in V(G);\ di(G_p)_{v,i} < di(G_p)_{v,j};\ G_p = G - p \quad (2)$$

In the above relations, $G_p = G - p$ is the spanning subgraph, resulted by deleting the path *p* joining the vertices *i* and *j* (except its endpoints), *di*(*G*) and *di*(G_p) denote the topological distances measured in *G* and G_p, respectively.

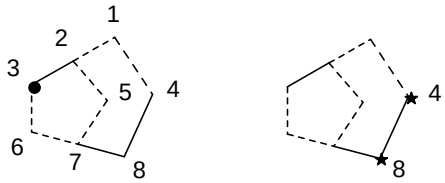
The sets $CJ_{i,j,p}$ and $CF_{i,j,p}$ represent subgraphs (connected or not) in *G,* referred to the endpoint *i* and related to *j* and path *p*.

In defining *Cluj indices*, the *path p* plays the central role in selecting the subgraphs (eqs 1 and 2), particularly in cycle-containing graphs, where more than one path could join the pair (*i,j*). In such graphs, more than one subgraph (*i.e.*, fragment), referred to *i*, can be counted. By this reason, the non-diagonal entries $[\mathbf{UM}]_{ij}$ in Cluj matrices are defined as the *maximum cardinality* of the sets supplied by eq 1 or 2

$$[\mathbf{UM}]_{ij} = \max_{p} \left| V_{i,j,p} \right| \tag{3}$$

where $V_{i,j,p}$ is either $CJ_{i,j,p}$ or $CF_{i,j,p}$ and consists of vertices, *v*, lying *closer* to the vertex *i* than to the vertex *j*. When $p \in Di(G)$, (*i.e.*, the set of all topological distances, or geodesics in *G*) then **M = CJDi** (Cluj-Distance) or **CFDi** (Cluj-Fragmental-Distance). When $p \in De(G)$, (*i.e.*, the set of all topological detours, or the longest distances in *G*) **M = CJDe** (Cluj-Detour) or **CFDe** (Cluj-Fragmental-Detour). The diagonal entries are zero. The Cluj matrices are square arrays, of dimension NxN, usually *unsymmetric* (excepting some symmetric regular graphs).

Figure 1 illustrates the construction of **CJDe** matrix.



Cluj Detour Sets $CJDe_{i,j,p}$; pair (3, 4):

(3, 4) [3, 6, 7, 5, 2, 1, 4]{ 3 } (4, 3) [4, 1, 2, 5, 7, 6, 3]{ 4, 8 }

Cluj-Detour Matrix **UC JDe**

0	1	1	1	1	1	2	1	8
2	0	2	2	2	2	2	3	15
2	1	0	***1***	1	1	1	1	8
1	1	***2***	0	2	2	1	1	10
1	1	1	1	0	1	1	1	7
1	1	1	1	1	0	1	2	8
3	2	2	2	2	2	0	2	15
1	2	1	1	1	1	1	0	8
11	9	10	9	10	10	9	11	79

IP2(*CJDe*) = 56

IE2(*CJDe*) = 15

Figure 1. Construction of Cluj Detour matrix, UCJDe

The unsymmetric matrices can be symmetrized, *e.g.*, by the Hadamard product with their transposes

$$\mathbf{SM}_p = \mathbf{UM} \bullet (\mathbf{UM})^T \quad (4)$$

$$\mathbf{SM}_e = \mathbf{SM}_p \bullet \mathbf{A} \quad (5)$$

The symbol $\bullet$ indicates the Hadamard (pairwise) matrix product (*i.e.*, $[\mathbf{M}_a \bullet \mathbf{M}_b]_{ij} = [\mathbf{M}_a]_{ij}\ [\mathbf{M}_b]_{ij}$). In eq 5, the Hadamard product between the path-defined matrix $\mathbf{SM}_p$ and the adjacency matrix **A (***i.e.*, the matrix having the non-diagonal entries unity for two adjacent vertices and zero otherwise**)** provides the corresponding edge-defined matrix, $\mathbf{SM_e}$, which is a weighted adjacency matrix. For the symmetric matrices, the letter **S** is usually missing.

In trees, **CJDi**, **CJDe**, **CFDi** and **CFDe**, are identical, due to the uniqueness of the path joining a pair of vertices (*i,j*).

The above matrices allow the calculation of indices by relations given for the fragmental property indices[9].

Model Description

Let (*i,j*) be a pair of vertices and $Fr_{i,j}$ any fragment referred to *i* and related to *j*.

Dense Topological Model

Let *v* be a vertex in the fragment $Fr_{i,j.}$ The property descriptor applies to the vertex property p_v and topological distance $d_{Tv,j}$. The fragmental *property descriptor PD*, resulting by the vertex descriptor superposition, gives the interaction of all the points belonging to the fragment $Fr_{i,j}$ with the point *j*:

$$PD(Fr_{i,j}) = \underset{v \in Fr_{i,j}}{\Psi} (\Omega(d_{T\,v,j}, p_v)) \quad (6)$$

The *j* point can be conceived as an *internal probe atom* with no chemical identity.

Rare Topological Model

Within this model, the property descriptor applies to the fragmental property and topological distance $d_{T\ i,j}$. The fragmental property descriptor models the interaction of the whole fragment $Fr_{i,j}$ with the point *j* and looks the global property being *concentrated* in the vertex *i*:

$$PD(Fr_{i,j}) = \Omega(d_{T\,i,j}, \underset{v \in Fr_{i,j}}{\Psi} (p_v)) \quad (7)$$

Dense Geometric Model

The fragmental property descriptor is the vector sum of the vertex descriptor vectors. It applies the property descriptor to the vertex property p_v and the Euclidean distance $d_{E\,v,j}$ in providing a *point of equivalent* (*fragmental*) *property* located at the Euclidean distance $d_{E\,CP,j}$ (with $d_{E\,CP,j}$ being the *distance*

of property). The vector of the fragmental property has the orientation of this *distance vector*. The model simulates the interactions in non-uniform fields (gravitational, electrostatic, etc):

$$PD(Fr_{i,j}) = \left\| \sum_{v \in Fr_{i,j}} \vec{\Omega}\left(d_{E\,v,j}, p_v\right) \right\|;\ \vec{\Omega} = \Omega \cdot \frac{\vec{d}_{E\,v,j}}{d_{E\,v,j}};\ P(Fr_{i,j}) = \underset{v \in Fr_{i,j}}{\Psi}(p_v);$$

$$d_{E\,CP,j} = \Omega_p^{-1}(DG(Fr_{i,j}), P(Fr_{i,j})), \tag{8}$$

where $d_{E\,CP,j}$ is the distance that satisfies: $\Omega(d_{E\,CP,j}, P(Fr_{i,j})) = PD(Fr_{i,j})$

Rare Geometric Model

The scalar fragmental descriptor applies the property descriptor to the *center of fragment property* and Euclidean distance between this center and the vertex *j*.

The model simulates the interactions in uniform fields (uniform gravitational, electrostatic, etc.):

$$PD(Fr_{i,j}) = \Omega(d_{E\,CP_i,j}, \underset{v \in Fr_{i,j}}{\Psi}(p_v));$$

$$CP_i(x_{CP_i,j}, y_{CP_i,j}, z_{CP_i,j});\ x_{CP_i,j} = \sum_{v \in Fr_{i,j}} x_v \cdot p_v \Big/ \sum_{v \in Fr_{i,j}} p_v \tag{9}$$

$$y_{CP_i,j} = \sum_{v \in Fr_{i,j}} y_v \cdot p_v \Big/ \sum_{v \in Fr_{i,j}} p_v;\ z_{CP_i,j} = \sum_{v \in Fr_{i,j}} z_v \cdot p_v \Big/ \sum_{v \in Fr_{i,j}} p_v$$

Fragmental Property Matrices

The fragmental property matrices are non-symmetric square matrices of order *N* (i.e., the number of non-hydrogen atoms in the molecule). The non-diagonal entries in such matrices are fragmental properties corresponding to any pair of vertices (*i,j*) by a chosen model.

In case of Cluj criteria, the fragmentation can supply more than one maximal fragment for the pair (*i,j*). In such cases, the matrix entry is the arithmetic mean of the individual values.

Thus, if $i, j \in V(G)$, $i \neq j$ and $P_{i,j} = \{p_{i,j}^1, p_{i,j}^2, \ldots, p_{i,j}^k\}$ paths joining *i* and *j*, then cf. *CJ* or *CF* definition, the fragments $Fr_{i,j}^1, Fr_{i,j}^2, \ldots, Fr_{i,j}^k$ are generated. Let *m* be the number of maximal fragments among all the *k* fragments, $1 \leq m \leq k$, and let $\sigma_1, \ldots, \sigma_m$ be the index for the maximal fragments.

By applying any of the above models, for all *m* maximal fragments we obtain *m* values, e.g.:

$$PD(Fr_{i,j}^{\sigma_1}),\ PD(Fr_{i,j}^{\sigma_2}),\ \ldots,\ PD(Fr_{i,j}^{\sigma_m})$$

and consequently, the matrix entry associated to the pair (*i,j*) is the mean value:

$$PD_{i,j} = \frac{\sum_{t=1}^{m} PD(Fr_{i,j}^{\sigma_t})}{m} \tag{10}$$

Fragmental Property Indices

Fragmental property indices are calculated at any fragmental property matrices above discussed, by applying four types of index operators: *P_*, *P2*, *E_*, *E2* according to the relations:

$$P\_(M)=½\Sigma\Sigma[\mathbf{M}]_{i,j}\ ;\ P2(M)= ½\Sigma\Sigma[\mathbf{M}]_{i,j}[\mathbf{M}]_{j,i};$$
$$E\_(M)= ½\Sigma\Sigma[\mathbf{M}]_{i,j}\,[\mathbf{A}]_{i,j}\ ;\ E2(M) = ½\Sigma\Sigma[\mathbf{M}]_{i,j}[\mathbf{M}]_{j,i}[\mathbf{A}]_{i,j} \tag{11}$$

where **M** is any property matrix, symmetric or unsymmetric.

Symbolism of the Fragmental Property Matrices and Indices

The name of *fragmental property matrices* is of the general form:

ABcDdEfffffG (12)

where:

A ∈ {**D**, **R**}; D = Dense; R = Rare;
B ∈ {**T**, **G**}; *T* = Topological; *G* = Geometric;
c ∈ {**f**, **j**, **s**}; *f* = *CF*-type; *j* = *CJ*-type; *s* = *Sz*-type;
Dd ∈ {**Di**, **De**}; *Di* = Distance; *De* = Detour;
E ∈ Φ (i.e., **E**∈{**M**, **E**, **C**, **P**}
where *M* = *mass*; *E* = *electronegativity*; *C* = *cardinality*;
P = other atomic property - implicitly, *partial charge*; explicitly,
a property given by manual input);

fffff ∈ Ω (i.e., **fffff**∈{**__p__**, **_1/p_**, **__d__**, **_1/d_**, **_p.d_**, **_p/d_**, **_p/d2**, **p2/d2**}
G ∈ Ψ (i.e., **G**∈{**S**, **P**, **A**, **G**, **H**} with the known meaning (see above).
The name of *fragmental property indices* is of the general form:

ABcDdEfffffGii (13)

where: *ii* ∈ {*P_*, *P2*, *E_*, *E2*} with the known meaning (eq 24).

If an operator, such as *f(x)=1/x* (inverse operator) or *f(x)=ln(x)*, is applied the indices are labeled as follows:

$$lnABcDdEfffffGii := \ln(ABcDdEfffffGii);$$
$$1/ABcDdEfffffGii := \frac{1}{ABcDdEfffffGii} \tag{14}$$

For example, index *lnDGfDeM__p__SP_* is the logarithm of index *DGfDeM__p__SP_* computed on the property matrix **DGfDeM__p__S**. The model used is dense, geometric, on fragment of type *CF*, with the cutting path being detour. The chosen property is the mass, the descriptor for property is even the property (mass) and the sum operator counts the vertex descriptors.

The fragmental indices were calculated by the aid of ***Cluj3Cmd*** 16-bit windows computer programs.

CORRELATING STUDIES

A mathematical model for correlating some biological activities or physical properties with molecular structures can be built up by using *multy linear regression* ***MLR***.

MLR, for *n* observations and *m* independent variables is represented by equation

$$\mathbf{Y_i = b_0 + \sum_{j}^{m} b_{ij} X_{ij}} \tag{15}$$

The regression coefficients $\boldsymbol{b_{ij}}$ can be determined by the least-squares method. Eq (28) can be used for estimating a chosen property in any other sets of chemical structures.

To avoid the chance correlations, it is recommended that the number of descriptors submitted to regression be less than 60 % of the number of observations in the training set.[10]

Physico-chemical properties

Cycloalkanes

A set of 25 cycloalkanes[4] (Table 1) was chosen for testing the correlating ability of some Cluj type indices with viscosity (as log η)[14-15] (Table 1). Topological indices and several properties are presented in ref 4, while the statistics of multilinear regression (MLR) appear in Table 2.

Table 1.

Structural formula for some cycloalkanes.

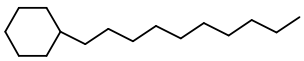
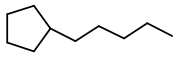
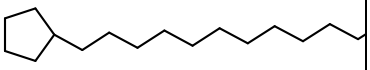
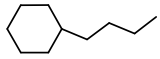
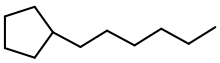
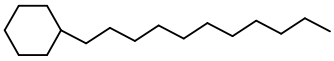
No.	Structural formula	No.	Structural formula
1		14	
2		15	
3		16	

Table 1. (continued)

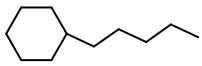
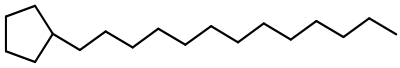
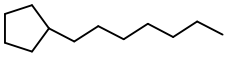
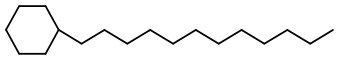
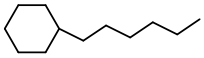
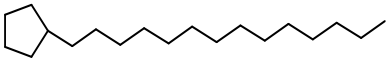
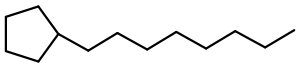
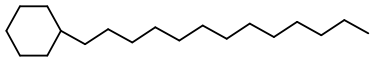
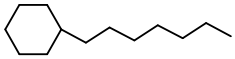
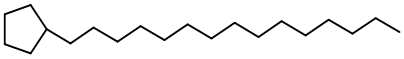
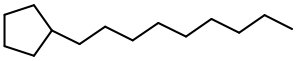
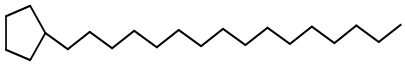
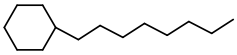
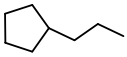
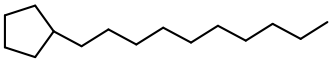
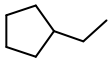
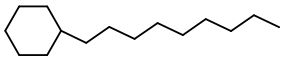
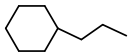
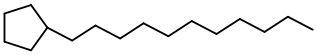
No.	Structural formula	No.	Structural formula
4		17	
5		18	
6		19	
7		20	
8		21	
9		22	
10		23	
11		24	
12		25	
13			

Table 2.

MLR data: $Y = a + \sum_i b_i X_i$; Y = viscosity of structures in Table 1.

No.	X_i	b_i	a	r	s	cv(%)	F
1	CJD_p	0.0001	0.1489	0.8760	0.1721	34.746	75.897
2	CJD_e	0.0008	0.0387	0.9213	0.1389	28.020	129.074
3	$CJ\Delta_p$	0.0001	0.1872	0.8530	0.1864	37.598	61.459
4	$CJ\Delta_e$	0.0008	0.1178	0.8887	0.1638	33.037	86.392
5	$\ln CJD_p$	0.3018	-1.7589	0.9944	0.0377	7.600	2044.310
6	$\ln CJD_e$	0.3905	-1.8722	0.9922	0.0446	9.002	1450.474
7	$\ln CJ\Delta_p$	0.2619	-1.3651	0.9920	0.0451	9.094	1420.770
8	$\ln CJ\Delta_e$	0.3172	-1.3142	0.9887	0.0536	10.824	996.154
9	$\ln CJD_p$	0.2684	-1.5958	0.9969	0.0292	5.879	1137.362
	CJD_p	0.0002					
	$CJ\Delta_p$	-0.0002					
10	$\ln CJ\Delta_p$	0.2169	-1.1728	0.9970	0.0289	5.833	1162.764
	CJD_p	0.0003					
	$CJ\Delta_p$	-0.0003					

The logarithm of the values of both Wiener and Cluj-type indices led to good correlation coefficients (over 0.99, already in single variable regression) and coefficients of variance less than 10 %.

A cross-validation procedure (leave one out – loo) indicated a good predicting ability of our indices:

$lnCJD_p$ (entry 5 –Table 3), $r_{(loo)}$ = 0.9933; s = 0.0414; v% = 8.349;

$lnCJDp\&CJDp\&CJ\Delta_p$ (entry 9), $r_{(loo)}$ = 0.9957; s = 0.0330; v% = 6.661;

$lnCJ\Delta_p\&CJDp\&CJ\Delta_p$ (entry 10), $r_{(loo)}$ = 0.9957; s = 0.0329; v% = 6.638.

N-containing compounds

A set of 90 N-containing compounds (Table 3) of industrial importance was taken from the paper.[23] The tested property was the normal boiling point, B.P. The authors modeled this property by using four categories of molecular descriptors: topological, geometric, electronic and charged-partial surface area descriptors (CPSA).[24,25]

The nitrogen-containig compounds were problematic in modeling a diverse set of organic chemicals, so that the authors excluded such compounds from their initial model.

The best found MLR model involved ten descriptors (1. dipole moment; 2. partial negative surface area; 3. relative negative charge; 4. relative negative charged surface area; 5. number of aromatic bonds; 6. path 2 molecular connectivity index; 7. cluster 3 valence connectivity index; 8. sum of all path weights from heteroatoms; 9. surface area of donatable hydrogens and 10. charge of donatable hydrogens) and showed the following statistics: n = 90 compounds; R = 0.990; s = 10.7 K. The largest pairwise R value of descriptors was 0.83. The modeling was performed by the ADAPT system.[26]

Our aim was to verify the quality of our property descriptors exactly in the same conditions as given in ref.[23] Thus, we extracted from the initial set of 104 N-containing compounds the same subset of 90 structures.

Molecular geometries and partial charges were calculated by the semiempirical AM1 method. The set of 19350 descriptors were reduced to 16383 after the monovariate regression.[19] Our procedure for finding the optimal subset of descriptors led to a subset of 72 descriptors.

The best scores in ten variate regression for the set of 90 compounds of Table 3 are listed in Table 4.

The best model was:

$$BP_{calc} = 225.441 - 59.627*lnDTsDiP\_p/d2SE2 + 316.627*RTsDiPp2/d2AE\_ 1.124*DGfDePp2/d2PP\_ - 1729.562*1/DTsDiE\_p*d\_HE2 - 0.010*1/DTsDePp2/d2SP2 - 49.623*1/DGsDeP\_p*d\_HE\_ + 8.846*lnDGjDiPp2/d2GP\_ - 4.698*1/RGjDeP\_p*d\_GP\_ - 12.188*lnDGjDeP\_p/d\_HP\_ + 33.597*DGjDeE\_\_p\_\_SE2$$

$$R = 0.98543;\ s = 13.149;\ n = 90 \qquad (16)$$

Table 3.

Nitrogen-Containing Compounds and Their Boiling Points.

No.	Compound	BP	No.	Compound	BP
1.	2-ethylpyridine	422.2	46.	n-tetradecylamine	564.5
2.	2-ethylpiperidine	416.2	47.	acridine	619.2
3.	1-ethylpiperidine	404.2	48.	tri-n-butylamine	487.2
4.	2,2-dimethyl-1,3-diaminomethane	426.2	49.	n-dodecylamine	532.4
5.	N,N-dimethyl-1,3-diaminomethane	418.2	50.	diamylamine	476.1
6.	3,3-dimethylpiperidine	410.2	51.	tripropylamine	429.7
7.	p-fluorobenzylamine	456.2	52.	n-nonylamine	475.4
8.	cianogene	252	53.	quinoline	510.8
9.	m-bromoaniline	524.2	54.	acetonitrile	354.8
10.	o-bromoaniline	502.2	55.	isoquinoline	516.4
11.	N-ethylbutylamine	381.2	56.	n-octylamine	452.8
12.	triethylamine	362	57.	indole	526.1
13.	N,N-diethylamin¾	337.2	58.	n-heptylamine	430.1
14.	o-nitrotoluene	498.2	59.	p-nitrotoluene	511.7
15.	nitrocyclopentane	453.2	60.	benzonitrile	464.1
16.	N-alylaniline	492.2	61.	3-nitrobenzotrifluoride	475.9
17.	ethylamine	289.7	62.	di-n-propilamine	382
18.	p-nitrophenole	552.2	63.	nitrohexane	436.8
19.	cyclopentylamine	380.2	64.	phenilhidrazine	516.7
20.	2-methylbutylamine	368.7	65.	methylamine	266.8
21.	N-methylbutylamine	364.2	66.	3-methylpyridine	417.3
22.	benzylamine	457.7	67.	aniline	457.2
23.	p-methoxyaniline	514.7	68.	p-chloroaniline	503.7
24.	m-methoxyaniline	524.2	69.	m-chloroaniline	501.7
25.	o-methoxyaniline	498.2	70.	n-pentylamine	377.6
26.	t-pentylamine	350.2	71.	isobutylamine	340.9
27.	dimethylamine	280	72.	diethylamine	328.6
28.	1-(2-aminoethyl)-piperidine	459.2	73.	tert-butylamine	317.5
29.	1-(2-aminoethyl)-piperidine	493.2	74.	n-butylamine	350.6
30.	9-methyl carbazole	616.8	75.	Pirolidine	359.7
31.	carbazole	627.8	76.	nitromethane	374.4
32.	4-methylaniline	473.6	77.	isobutyronitrile	376.8
33.	3-methylaniline	476.5	78.	n-butyronitrile	390.8
34.	2-methylaniline	473.5	79.	cis-crotonitrile	380.6
35.	2-propylamine	304.9	80.	trimethylamine	276
36.	1-naphtylamine	573.8	81.	2-nitropropane	393.4
37.	nitroethane	387	82.	1-nitropropane	404.3
38.	piperidine	376.4	83.	propionitrile	370.5
39.	4-methylpyridine	418.5	84.	acrylonitrile	350.5
40.	2-methylpyridine	402.5	85.	N-methylhexylamine	414.2
41.	pyridine	388.4	86.	n-heptylamine	428.2
42.	pyrole	402.9	87.	N-t-butyl-i-propylamine	371.2
43.	2-butylamine	335.9	88.	2-aminoheptane	416.2
44.	triamylamine	516.2	89.	malononitrile	491.5
45.	ethylenimine	329	90.	hydrogen cyanide	298.8

Table 4.
The Best Multivariate Regressions for the 90 Structures of Table 14.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	R
1	7	4959									0.9446
3	5	4959									0.9487
5	5	4959	10990	9671							0.9628
6	5	4959	10990	7206							0.9671
8	5	4959	10990	9671	3320	6422					0.9761
9	5	4959	10990	9671	3528	6422					0.9766
11	5	4959	10990	9671	3528	6422	16148	6895			0.9798
12	5	4959	10990	9671	3528	6422	6895	15789			0.9800
14	5	4959	10990	9671	3528	6422	16158	16225	15060	15789	0.9843
12	6	6895	4	16275	963	841	163	13920	1	4727	0.9854

The plot corresponding to eq 16 is given in Figure 1.

Our result is slightly lower (s = 13.149 K) than that reported in ref.[2] (s = 10.7 K). It is possible to further improve the model by mining the whole descriptor pool not only within the limits of a heuristic procedure. Another possibility is to use different training subset selection and outlier elimination. Such procedures will be reported in a future paper.

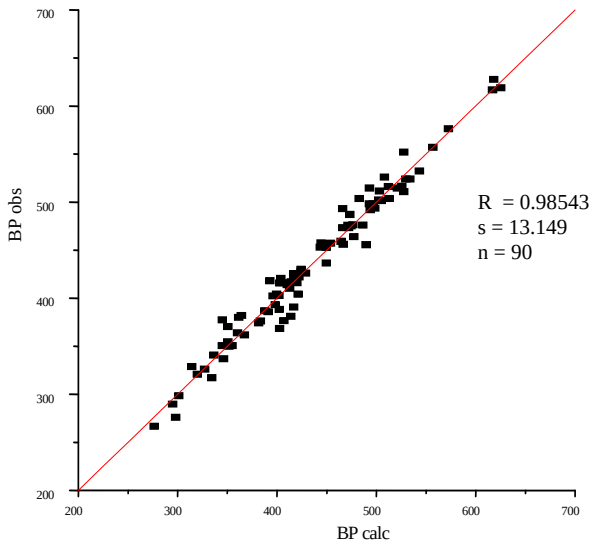


Figure 1. The plot of calculated vs observed normal boiling points (the set of Table 4).

Polychlorinated Biphenils

Polychlorinated biphenils, PCBs, have been synthesized and used as dielectric fluids in electrotechnics, fire retardents, plasticiers or pesticides.[39-41] The occurrence of PCBs in the environment[42] is detrimental for the reproduction[43] of several animal species and is hazardous to humans.[44]

By these reasons, PCBs were monitorized in the environment[45] and their biological activity was modeled.[46-48]

In the present work the *vapor pressure* of PCBs (as logVP), taken from the Rouvray's report[46]) at 25 ^{o}C (VP25) and 100 ^{o}C (VP100), respectively are modeled by using *FPIF*. The property is important in connection with PCBs spread and toxicity. Tables 5 lists the VP25 values and the corresponding TIs (showing the best scores in mono and bivariate regression).

Table 5.

Polychlorinated Biphenils PCBs, logVP25, logVP100 and *FPIF* Descriptors

	PCBs	logVP 25 (^{o}C)	logVP 100 (^{o}C)	lnDGfDeP_p/d2HP2 (1)[a]	lnDGjDeP_p/d2HP2 (2)	1/RGjDeCp2/d2GE_ (3)	1/RTjDiC_p*d_HP2 (13325)	1/DGjDiE_p/d_GP2 (13060)	1/RTjDiM_p*d_AP2 (14595)
1	Biphenyl	0.0043	-	5.6402	5.6449	0.1846	0.0179	0.0123	0.0011
2	2-Chloro-	0.1847	2.4553	5.8422	5.8669	0.1856	0.0104	0.0078	0.0006
3	3-Chloro-	-0.1409	-	6.0050	6.0077	0.1876	0.0103	0.0068	0.0005
4	4-Chloro-	-0.757	-	6.1627	6.1660	0.1883	0.0088	0.0059	0.0005
5	22'-Dichloro-	-0.8729	-	6.0255	6.0105	0.1865	0.0066	0.0056	0.0003
6	33'-Dichloro-	-1.5889	1.3257	6.3264	6.3291	0.1908	0.0059	0.0041	0.0003
7	44'-Dichloro-	-2.58	1.0667	6.6288	6.6313	0.1924	0.0044	0.0030	0.0002
8	25'-Dichloro-	-1.1107	1.679	6.1390	6.1595	0.1889	0.0063	0.0042	0.0003
9	2'34-Trichloro-	-1.8601	1.0366	6.3540	6.3721	0.1924	0.0038	0.0030	0.0002
10	246-Trichloro-	-1.9066	1.4201	6.4185	6.4252	0.1903	0.0042	0.0029	0.0003
11	22'55'-Tetrachloro-	-2.3036	0.9128	6.5843	6.5947	0.1934	0.0027	0.0020	0.0001
12	22'455'-Pentachloro-	-2.9547	0.3892	6.9262	6.9279	0.1976	0.0018	0.0013	0.0001
13	22'44'66'-Hexachloro-	-2.762	0.4518	7.0418	7.0554	0.1963	0.0014	0.0010	0.0001
14	22'33'55'66'-Octachloro-	-4.5391	-	7.7351	7.7474	0.2018	0.0010	0.0007	0.0001
15	22'33'44'55'66'-Decachloro-	-7.2757	-2.9914	8.9277	8.9613	0.2127	0.0005	0.0004	0.0001
				mono	*mono*	*mono*	*bi*	*bi*	*bi*
							lnDGfDeP_p/d2HP2	lnDGfDeP_p/d2HP2	lnDGfDeP_p/d2HP2
			n = 15				(1)	(1)	(1)
			r	-0.9869	-0.9863	-0.9852	0.9889	0.9890	0.9881
			s	0.323	0.330	0.343	0.314	0.315	0.320
			F	486.528	465.831	429.922	258.661	256.989	249.513
			b_0	13.003	12.912	48.234	11.844	11.813	12.135
			b_1	-2.284	-2.266	-260.973	35.455	51.568	468.873
			b_2				-2.138	-2.135	-2.174

[a] score in monovariate regression.

The best *monovariate* regression reveals the fact that the vapor pressure of PCBs is a function of the molecular geometry (G - four of six best descriptors - Table 5 - are of geometric model), that further control the distribution of partial charges (P) and, ultimately, the molecular polarity. Other important local properties are electronegativity (E) and atomic mass (M).

Recall that, in biphenils, the torsion angle between the two benzene rings depends on the number and nature of attached substituents. It is involved in the extend of aromatic conjugation and thereafter in the charge distribution.

$$\log PV25 = 13.003 - 2.284 * \ln DGfDeP\_p/d2HP2 \quad (17)$$
$$n = 15;\ r = -0.98690;\ s = 0.323;\ F = 486.528$$

In *bivariate* regression, the model is slightly better (Table 5, columns 6-8).

We compared the models supplied by *FPIF* with those given by some graph theoretical descriptors: IP(Di) = Wiener index W, IP2(CJDi), IE2(CJDi), IP2(CJDe), IE2(CJD2) and IP2(CFDi). The values of descriptors are included in Table 8. The drop in correlation coefficient *r* of the best models is of 1.7 % (with IP2(CJDi), column 3, Table 6) and 0.4 % (with IE2(CJDi) & IE2(CJDe), column 5) in monovariate and bivariate regression, respectively. Clearly the computational cost is far more less for the graph theoretical descriptors. This result is not surprising since the rotation of the two rings around the joining bond is quite hindered in substituted biphenils (see the occurrence of atropisomery in this class of organic compounds).

Table 6.

Polychlorinated Biphenils PCBs and Graph Theoretical Descriptors *vs*. logVP25.

	PCBs / Index	IP(Di)	IP2(CJDi)	IE2(CJDi)	IP2(CJDe)	IE2(CJDe)
1	Biphenyl	198	1169	360	381	72
2	2-Chloro-	240	1406	426	513	100
3	3-Chloro-	246	1501	438	508	94
4	4-Chloro-	252	1545	450	489	94
5	22'-Dichloro-	287	1679	499	678	133
6	33'-Dichloro-	301	1906	527	663	119
7	44'-Dichloro-	315	2008	555	616	119
8	25'-Dichloro-	294	1792	513	670	126
9	2'34-Trichloro-	358	2236	618	810	155
10	246-Trichloro-	348	2031	600	750	156
11	22'55'-Tetrachloro-	412	2511	700	1003	192
12	22'455'-Pentachloro-	488	3038	824	1164	226
13	22'44'66'-Hexachloro-	555	3325	927	1335	279
14	22'33'55'66'-Octachloro-	702	4291	1152	1855	364
15	22'33'44'55'66'-Decachloro-	907	5706	1483	2274	459
	n = 15	*mono*	*mono*	*mono*	*mono*	*mono*
	r	0.9627	0.9703	0.9647	0.9470	0.9443
	s	0.543	0.485	0.528	0.648	0.660
	F	164.349	209.448	173.327	112.982	106.953

Table 6. (continued)

PCBs / Index	IP(Di)	IP2(CJDi)	IE2(CJDi)	IP2(CJDe)	IE2(CJDe)
b_0	1.705	1.633	2.018	1.067	0.932
b_1	-0.009	-0.002	-0.006	-0.003	-0.017
	bi	*bi*	*bi*	*bi*	
n = 15	IP(Di) IE2(CJDe)	IP2(CJDi) IP2(CJDe)	IP2(CJDi) IP2(CFDi)	IE2(CJDi) IE2(CJDe)	
r	0.9846	0.9822	0.9833	0.9848	
s	0.364	0.392	0.380	0.362	
F	190.879	164.390	175.340	193.354	
b_0	3.710	2.284	2.043	4.445	
b_1	-0.036	-0.004	-0.009	-0.021	
b_2	0.048	0.005	0.007	0.041	

The vapor pressure at 100 ^{0}C, as logVP100, is modeled using the descriptors shown in Table 6. The best monovariate model is:

$$\text{logPV100} = 7.713 - 1.625^* \text{ lnRGsDeM\_p/d2SE2} \quad (18)$$
$$n = 10;\ r = -0.98318;\ s = 0.281;\ F = 231.873$$

Looking at eq (18) reveals that: the geometry is again important in modeling the property but the local property governing VP100 is the atomic mass (M). It is correlated with the loss in the electrostatic interactions of molecules in liquid phase and increase of gravitational interactions (see the rare geometric model. Figure 2 shows the plot of lnRGsDeM_p/d2SE2 vs logVP100.

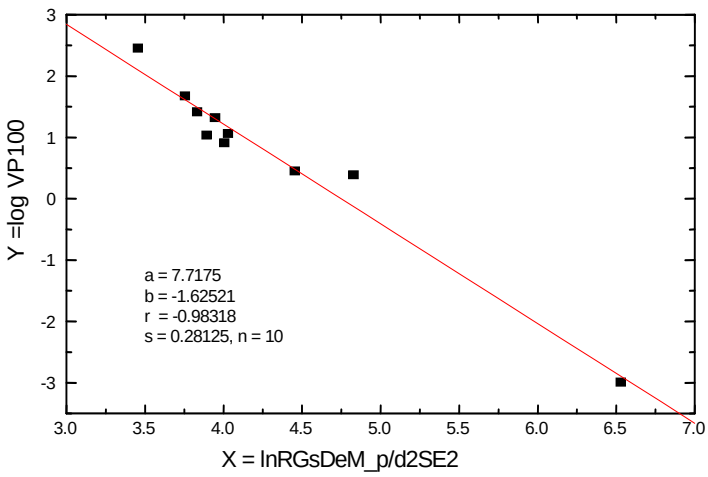


Figure 2. The plot: lnRGsDeM_p/d2SE2 vs logVP100.

In *bivariate* regression, the model is still better. The best model gives an additional support of the conclusion that vapor pressure of PCBs is better modeled by *FPIF* descriptors including information on geometry, partial charges and atomic mass:

$$\log VP100 = -15.005 - 1.926*\ln RGsDeM\_\_p/d2SE2 + 15.095*\ln RGjDeP\_1/d\_HE2$$

$$n = 10;\ r = 0.9956;\ s = 0.155;\ F = 390.790 \qquad (19)$$

By using the classical descriptors (Table 6), other two excellent correlations were found:

$$\log VP100 = 2.8123 + 0.0137\ IE2(CjDi) - 0.0045\ IP2(CjDi) \qquad (20)$$

$$n = 10;\ r = 0.9934;\ s = 0.188;\ F = 265.920$$

$$\log VP100 = 3.5364 - 0.0044\ IP2(CjDi) + 0.0207\ IP2(Di) \qquad (21)$$

$$n = 10;\ r = 0.9941;\ s = 0.179;\ F = 291.910$$

The best reported data in literature are as follows: Rouvray,[46] *W* (*r* = 0.9632/VP25; *r* = -0.8863/VP100) and Khadikar,[47] SZ (0.9843/VP25 - in error with the reported data!; corrected result: r = 0.9647, see Table 6, column 4; *r* = - 0.8921/VP100). The fragmental property indices take into account the chemical nature of atoms (mass, electronegativity and partial charge), various kinds of interactions between the fragments of molecules as generated by Cluj and Szeged criteria and the 3D geometry of molecular structures as well.

Biological activity

Pirazolidin-3,5-diones

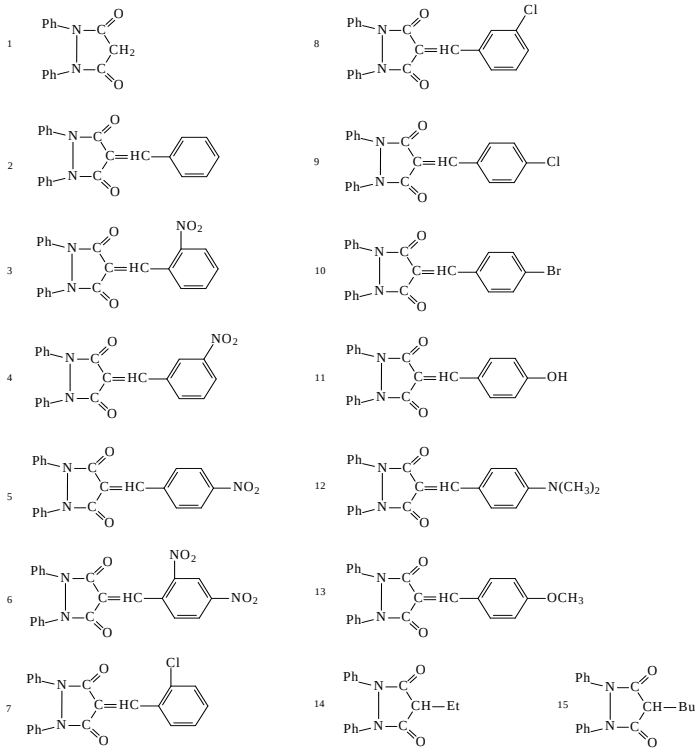


Figure 3. The set of Pirazolidin-3,5-diones.

The molecules presented in Figure 2 were synthesized in our laboratory[50]. The molecular structures were input and optimized by HyperChem (HyperCube Inc.) package. Partial charges were calculated by AM1 semiempirical approach.

Modeling Biological Activity

Pirazolidin-3,5-diones are known having antiinflamatory activity[50]. They also show some antimicrobial and antifungal activity (on *Staphylococus aureus, Bacilus subtilis, Escherichia coli, Pseudomonas aeruginosa, Candida albicans*, etc.). Table 7 shows the biological activity, in mm inhibition zone.

Table 7.

Antimicrobial Activity; Inhibition Zone (mm).

Compound	Gram-positives			Gram-negatives			Fungi
	Staphyl. aureus	Staphyl epider	Bacill subtilis	Esc coli	Prot. vulg.	Pseu. aerug.	Cand. albicans
1	0	0	0	0	0	0	0
2	10	10	10	0	0	0	0
3	14	10	13	0	0	10	0
4	12	17	14	0	0	10	10
5	0	0	10	0	0	10	10
6	0	0	10	0	0	10	10
7	12	0	10	0	0	10	0
8	12	0	10	0	0	10	0
9	12	0	10	0	0	10	0
10	12	10	10	0	0	0	10
11	12	13	13	0	0	10	12
12	0	0	12	0	0	10	13
13	0	0	0	0	0	10	12
14	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0

In the following, only two activities are considered for modeling: BA vs *Bacillus subtilis* and BA vs *Candida albicans*.

The activity vs. *Bacillus subtilis*, was estimated, in monovariate regression. The best three regression equations are given below:

$$BA_{calc} = -147{,}1950 + 35{,}4337 * \ln RGjDeE\_p/d\_HP2 \quad (22)$$
$$n = 15;\ r = 0.8320$$

$$BA_{calc} = -147{,}3113 + 35{,}4626 * \ln RGfDeE\_p/d\_HP2 \quad (23)$$
$$n = 15;\ r = 0.8319$$

$$BA_{calc} = -75{,}9515 + 35{,}6724 * \ln RGjDeC\_p/d\_HP2 \quad (24)$$
$$n = 15;\ r = 0.8316$$

In *bivariate* regression the model is still improved:

$$BA_{calc} = 21.0809 - 306.3117*1/RGfDeC\_p/d\_HP2 + 4.4915*lnDGjDeP\_\_p\_\_GE\_ \quad n = 15;\ r = 0.9857 \quad (25)$$

$$BA_{calc} = 6.2856 - 578.8831*1/DGjDeP\_p/d2PE\_ + 4.8740*lnDGjDeP\_p*d\_GE\_ \quad n = 15;\ r = 0.9883 \quad (26)$$

$$BA_{calc} = 4.0166 - 191.7906*1/RGjDeC\_p/d2HE\_+ 4.9157*lnDGjDeP\_p*d\_GE\_ \quad n = 15;\ r = 0.9885 \quad (27)$$

Table 8 includes the observed inhibitory activity vs. *Bacillus subtilis* and calculated BA by the above equations.

Table 8.

Biological Activity BA_{obs}. and BA_{calc} by eqs 25-27.

Comp. No.	BA (eq 25)	BA (eq 26)	BA (eq 27)
1	9.9993	9.9997	9.9998
2	10.2281	10.6001	10.6068
3	9.9127	10.1988	10.1479
4	10.5830	10.3631	10.4159
5	10.0743	9.4243	9.7268
6	12.2185	12.4991	12.6593
7	13.2109	12.0148	12.1553
8	11.7282	10.7701	10.9012
9	0.0143	0.6821	0.6908
10	0.7988	-0.0707	-0.0435
11	-0.4674	-0.5345	-0.5304
12	12.6072	13.3293	13.4258
13	11.2599	11.9890	11.6189
14	10.1309	10.8491	10.3785
15	-0.2987	-0.1143	-0.1529

As can be seen from eqs 25-27, the inhibiting activity of phtalazines vs *Bacillus subtilis* is controlled by the geometry (G in the symbol of indices) and electronic features of these molecules (E - electronegativity and P - partial charges).

The activity vs. *Candida albicans*, was estimated, in *monovariate* regression, as shown below:

$$BA_{calc} = - 4.3416 + 1.5663*\ lnDTfDeP\_p*d\_PP2 \quad n = 15;\ r = 0{,}9252 \quad (28)$$

$$BA_{calc} = - 4.1732 + 1.5461*\ lnDTjDeP\_p*d\_PP2 \quad n = 15;\ r = 0{,}9235 \quad (29)$$

$$BA_{calc} = - 2.3616 + 1.4733*\ lnDTfDiP\_p*d\_PP2 \quad n = 15;\ r = 0{,}8777 \quad (30)$$

In *bivariate* regression the improvement of correlation was not so sound as in case of *Bacillus subtilis:*

$$BA_{calc} = 58.0019 + 1.9258\ \ln DTfDiP\_p*d\_PP2 - 14.1524\ \ln RGsDiEp2/d2GP2 \quad n = 15\ ;\ r = 0.9415 \quad (31)$$

$$BA_{calc} = 39.1986 + 1.9336*\ln DTfDiP\_p*d\_PP2 - 18.7211*\ln RGsDiE\_p/d2AP2 \quad n = 15;\ r = 0.9429 \quad (32)$$

$$BA_{calc} = 7.0326 + 2.3522\ \ln DTjDiP\_p*d\_PP2 - 42.8766\ RGfDeP\_p/d\_AP2 \quad n = 15;\ r = 0.9523 \quad (33)$$

From eqs 31-33, it is suggesting that the antimycotic activity of phtalazines is controlled basically by the topology (T) and geometry (G), on one hand and electronic features (P - partial charges and E - electronegativity) of molecules.

Dipeptide ACE Inhibitors

The set consists of 58 dipeptides and was taken from Cocchi's report[16]. The molecular structure of these peptides was input and optimized by using the MM+ and then by semiempirical AM_1 procedure of the HyperChem Program (HyperCube Inc.). Table 4 includes the dipeptide names by using the one-letter code for aminoacids, the observed ACE inhibitory activity (biological activity, ***BA***, as $\log(1/IC_{50})$), the calculated BA according to the best model and the corresponding residuals. As above mentioned, ***FPIF*** descriptors take explicitly into account 3D-structural features of the whole molecule of dipeptides[17].

Table 9 collects the statistics of monovariate and bivariate regression in modeling the ACE inhibiting potency of dipeptides by ***FPIF***. Cross-validation tests (Leave-20%-out ***L20%o*** or Leave-one-out ***Loo*** procedures) are given here only for bivariate regressions.

Table 9.

Statistics for ACE inhibitors set.

Index	DTfDiM_p /d2GP_	lnDGsDiE_1 /p_GE_	DTjDeM_p /d2GP_	lnDTjDeEp2 /d2AE_	DTsDeP_1 /d_GP2	lnRGsDeMp2 /d2AE_
		DTfDiM_p /d2GP_		DTjDeM_p /d2GP_		DTsDeP_1 /d_GP2
r	0.7819	**0.8870**	0.7884	0.8754	0.7923	0.8717
r^2	0.6114	**0.7867**	0.6216	0.7663	0.6277	0.7599
s	0.630	0.471	0.622	0.493	0.616	0.500
F	88.106	101.426	91.999	90.147	94.420	87.029
b_0	-0.759	35.992	0.776	21.816	0.479	3.325
b_1	0.286	-11.802	0.143	-6.681	0.268	-3.194
b_2		0.822		0.529		0.571

Table 9. (continued)

	Cross-validated		
	L20%o [a]	**L20%o**	**L20%o**
r	**0.8715**	0.8592	0.8552
r^2	**0.7595**	0.7382	0.7314
s	0.495	0.517	0.524

[a] average of twenty five 20% sets of randomly chosen objects.

The best-found model was:

$$\boldsymbol{BA}_{calc} = 35.992 + 0.822 * \boldsymbol{DTfDiM\_p/d2GP\_} - 11.802 * \ln\boldsymbol{DGsDiE\_1/p\_GE\_}$$
$$n = 58;\ r = 0.88696;\ s = 0.471;\ F = 101.426 \qquad (34)$$

Table 10.

Comparative statistics of QSAR models of 58 ACE inhibitors and 48 sweeteners dipeptides.

	Peptide Set (Reference)	Descriptors per Residue	No. Comp.	r^2 (fitting)	r^2 (cross-validated)
1	ACE (Cocchi et al.)[16]	7	1	0.744	nd[a]
2	ACE (Collantes et al.) [19]	2	nd	0.700	nd
3	ACE (Zaliani et al. -extended)[18]	3	2	0.708	0.637
4	ACE (Zaliani et al. -rotameric)[18]	3	6	0.657	0.541
5	ACE (***FPIF***) [this work]	2	2	**0.787**	**0.759**[b]
6	Sweeteners (Jonsson et al.)[20]	3	1	nd	0.780
7	Sweeteners (Collantes et al.)[16]	2	2	0.847	nd
8	Sweeteners (Zalini et al. -extended)[18]	3	3	0.754	0.710
9	Sweeteners (Zalini et al. -rotameric)[18]	3	3	0.704	0.633
10	Sweeteners (***FPIF***) [this work]	2	2	**0.851**	**0.833**[b]

[a] Not determined; [b] Leave-20%-out, 25 times, of randomly chosen objects.

Both topology (***T*** - in the index symbol) and geometry (***G***) contribute to the best model. As local property, the atomic mass (***M***) and electronegativity (***E***) modulate the structure-activity relationship. For the best model (see also column 3, Table 10) the L20%o cross-validation was averaged on 25 randomly chosen 20% objects. The drop in ***r*** is around 1.6 % that proves a good predicting ability of the models. The plot of observed ***BA*** vs calculated ***BA*** is presented in Figure 1.

The model given by ***BA*** equation is superior, both in estimation and prediction, to those reported in literature (see Table 11). Note that the Zaliani's results refer both to a single conformation (i.e., extended) of amino acids and to a library conformation family (i.e., rotameric).

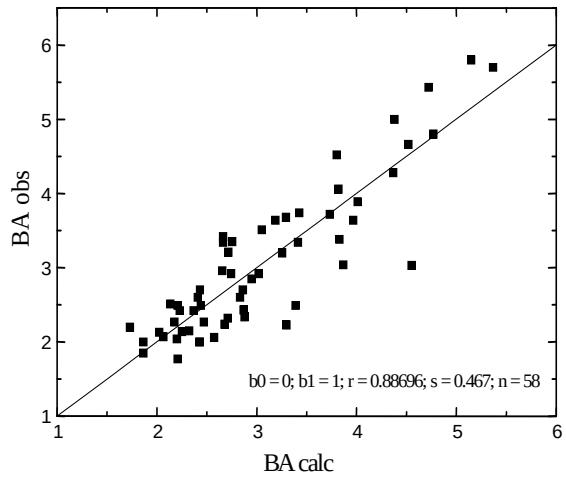


Figure 4. The plot of observed vs calculated BA.

Table 11.

The best ten bivariate regressions in ACE inhibitors test.

No.	Score 1	Score 2	Index 1	Index 2	***r***
1	89	5831	DTfDiM_p/d2GP_	lnDGsDiE_1/p_GE_	0.8870
2	54	1771	DTjDeM_p/d2GP_	lnDTjDeEp2/d2AE_	0.8754
3	29	7894	DTsDeP_1/d_GP2	lnRGsDeMp2/d2AE_	0.8717
4	29	2644	DTsDeP_1/d_GP2	lnRTjDeEp2/d2HE_	0.8686
5	18	8213	DTsDeM_p/d_GP2	lnRGsDeEp2/d2AE_	0.8681
6	18	7725	DTsDeM_p/d_GP2	lnRGsDeE_p/d2AE_	0.8624
7	15	6476	DTfDeM_p/d_PP_	lnRTsDiEp2/d2AE2	0.8618
8	1	15876	RTfDeE_1/p_PP2	lnRGsDeCp2/d2HP2	0.8614
9	1	8719	RTfDeE_1/p_PP2	DGfDiP_p/d_GP_	0.8518
10	1	6485	RTfDeE_1/p_PP2	lnRTsDiEp2/d2GE2	0.8465

Table 11 shows the occurrence of descriptors in the best 10 regression equations. All indices of the first variable in bivariate regression are topological (***T*** in index symbol) while only six of ten of the second variable are geometric (***G*** in index symbol). In general, a model is built up by using a training set of structures (that provides a calibration equation) and further it is validated by a cross-validation procedure and also by using an external prediction set. Due to the fixed mode of selection, the ***Loo*** procedure strongly requires an external set for prediction. It is not the case of *averaged* ***L20%o*** procedure, when the predicting sets (and implicitly the 80% training sets) can be randomly selected, thus getting enough statistical meaning for the model. A similar procedure was used in Zaliani's report[18]. This result correlates with the Zaliani's best result when used extended conformations (see Table 10).

As local property, the atomic mass (***M***) occurs five times in the first variable while the electronegativity (***E***) seven times in the second variable. Other occurring properties are the partial charge (***P***) and cardinality (***C***). Clearly, the chemical features play an important role in discriminating vertices (i.e., atoms or atom groups), fragments and whole molecules of dipeptides. They are strongly involved in modeling the biological activity of dipeptide ACE inhibitors.

Dipeptide Sweeteners

The set including 48 dipeptides was taken from Jonsson's paper[20]. The molecular structures were input and optimized by using MM+ and then by semiempirical AM_1 procedure of the HyperChem Program (HyperCube Inc.).

Table 12 collects the statistics of monovariate and bivariate regression in modeling BA of dipeptide sweeteners by ***FPIF***. The same remark holds for the cross-validation tests.

The best-found model was:

$$\boldsymbol{BA}_{calc} = 1.142 + 0.474^{*}\ \boldsymbol{RTsDiM\_1/p\_SP\_} - 0.043^{*}\ \boldsymbol{DGsDiE\_1/p\_AP\_}$$
$$n = 48;\ r = 0.92272;\ s = 0.248;\ F = 128.922 \qquad (35)$$

As in the previous test, both topology and geometry contribute to the best model and again the local property, was the atomic mass (***M***) and electronegativity (***E***).

In predicting tests, (see Table 12, columns 3, 5 and 7) the drop in ***r*** was around 1 %, proving a good stability of the models. The plot of observed BA vs calculated BA (eq 35) is presented in Figure 5. The model given by eq 35 surpasses those reported in literature (see Table 10).

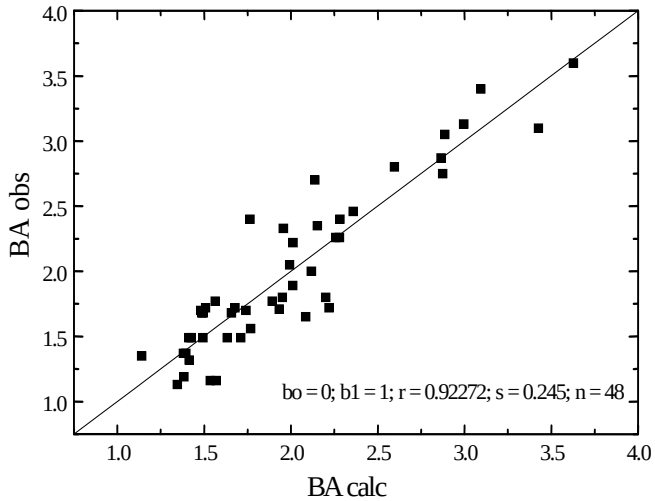


Figure 5. The plot of observed vs calculated BA (eq 35).

Table 12.

Statistics for sweeteners dipeptides.

1	2	3	4	5	6	7
Index	RTsDiE_1/ p_AE2	RTsDiEp2/ d2AE_	RTsDeE_1 /p_AE_	lnRTsDiEp2/ d2GE2	RTsDiM_1 /p_SP_	DGsDiE_1/ p_AP_
		RTsDiE_1/ p_AE2		RTsDeE_1/ p_AE_		RTsDiM_1/ p_SP_
r	0.81448	0.9169	0.8008	0.91525	0.7719	**0.9227**
r^2	0.66337	0.8407	0.6413	0.8377	0.5959	**0.8514**
s	0.369	0.257	0.381	0.259	0.404	0.248
F	90.650	118.714	82.231	116.116	67.821	128.922
b_0	0.106	0.482	0.115	0.471	-0.058	1.142
b_1	0.898	-0.058	0.315	-0.006	0.085	-0.043
b_2		4.479		1.364		0.474
	Cross-validated					
		L20%o		**L20%o**		***L20%o*** (aver.)[a]
r		0.9067		0.9047		**0.9129**
r^2		0.8221		0.8185		**0.8333**
s		0.268		0.271		0.259

[a] average of twenty five 20% sets of randomly chosen objects.

Table 13.

The best ten bivariate regressions in sweeteners dipeptides test.

No.	Score 1	Score 2	Index 1	Index 2	***r***
1	1219	7437	RTsDiM_1/p_SP_	DGsDiE_1/p_AP_	0.9227
2	6076	132	RTsDiEp2/d2AP2	DGsDeEp2/d2GE_	0.9209
3	33	6051	RTsDeE_1/p_AE_	RTsDiEp2/d2GE2	0.9153
4	1	3180	RTsDiE_1/p_AE2	RTsDiEp2/d2AE_	0.9169
5	1	3154	RTsDiE_1/p_AE2	RTfDiE_p/d2AP2	0.9076
6	1	3093	RTsDiE_1/p_AE2	RTsDiEp2/d2AP_	0.9027
7	1	3074	RTsDiE_1/p_AE2	RTjDiE_p/d_GP2	0.8846
8	1	3012	RTsDiE_1/p_AE2	RTsDeE_p/d_GP2	0.8846
9	1	2769	RTsDiE_1/p_AE2	lnDGsDeE_1/p_PP2	0.8768
10	1	2076	RTsDiE_1/p_AE2	DTsDiEp2/d2SE	0.8755

Table 13 shows the occurrence of descriptors in the best 10 regression equations. Seventeen indices in bivariate regression are topological while only three geometric. This result proves that the topology is the main feature in describing this dipeptide activity. In fact, topological indices are descriptors invariant to rototranslation, so that it is not surprising that Zaliani obtained the best correlation when used extended conformations of aminoacids.

As local property, the electronegativity (***E***) occurs nineteen times while the atomic mass (***M***) only once, in bivariate regression. It appears that the bitter tasting activity is controlled by electronic factors. The fragmental property indices take into account the chemical nature of atoms (mass, electronegativity and partial charge), various kinds of interactions between the fragments of molecules as generated by Cluj and Szeged criteria and the 3D geometry of molecular structures as well. For other ***FPIF*** modeling examples the reader can consult[9].

Substituted 3-(Phthalimidoalkyl)-pyrazolin-5-ones

We tested the correlating ability of ***FPIF*** on a set of 17 molecular structures from the class of substituted 3-(Phthalimidoalkyl)-pyrazolin-5-ones[9] with the sum of one-electron energy calculated at single point semi-empirical extended-Huckel and the inhibitory activity on *Lepidium sativum L*. (Cresson).

The molecular structure of the selected chemicals is given in Figures 6. It was performed by using the MM+ (for 3D-geometries) and semiempirical AM1 (for partial charge calculation) procedures of the HyperChem Program (HyperCube Inc.). The modeled properties were the sum of one-electron energy calculated at the Extended-Huckel level and the inhibitory activity (in %) of a solution of 0.05 mg/ml pyrazolin-5-one on *Lepidium sativum L*. (Cresson). The data are listed in Table 14.

Table 14.

The Sum of One-Electron Energy Calculated at Single Point Semi-Empirical Extended-Huckel and the Inhibitory Activity on *Lepidium sativum L*. (Cresson) for 17 Substituted 3-(Pthalimidoalkyl)-Pyrazolin-5-Ones*

Molecule no.	Energy (kcal/mol)	Inhibition (%)
1	50978.19	28.4
8	64751.09	65.2
7	64752.65	49.4
6	62330.33	68.3
5	38604.68	14.3
4	53416.95	27.7
3	53441.43	30.4
2	51000.36	28
17	41057.46	15.1
16	67104.64	50.6
15	64701.39	71.7
14	43473.37	18.2
13	41020.54	12.2
12	55832.12	32.6
11	55729.99	28.9
10	53424.19	29.3
9	50012.42	46.9

• Values of inhibition are taken from ref. [21]

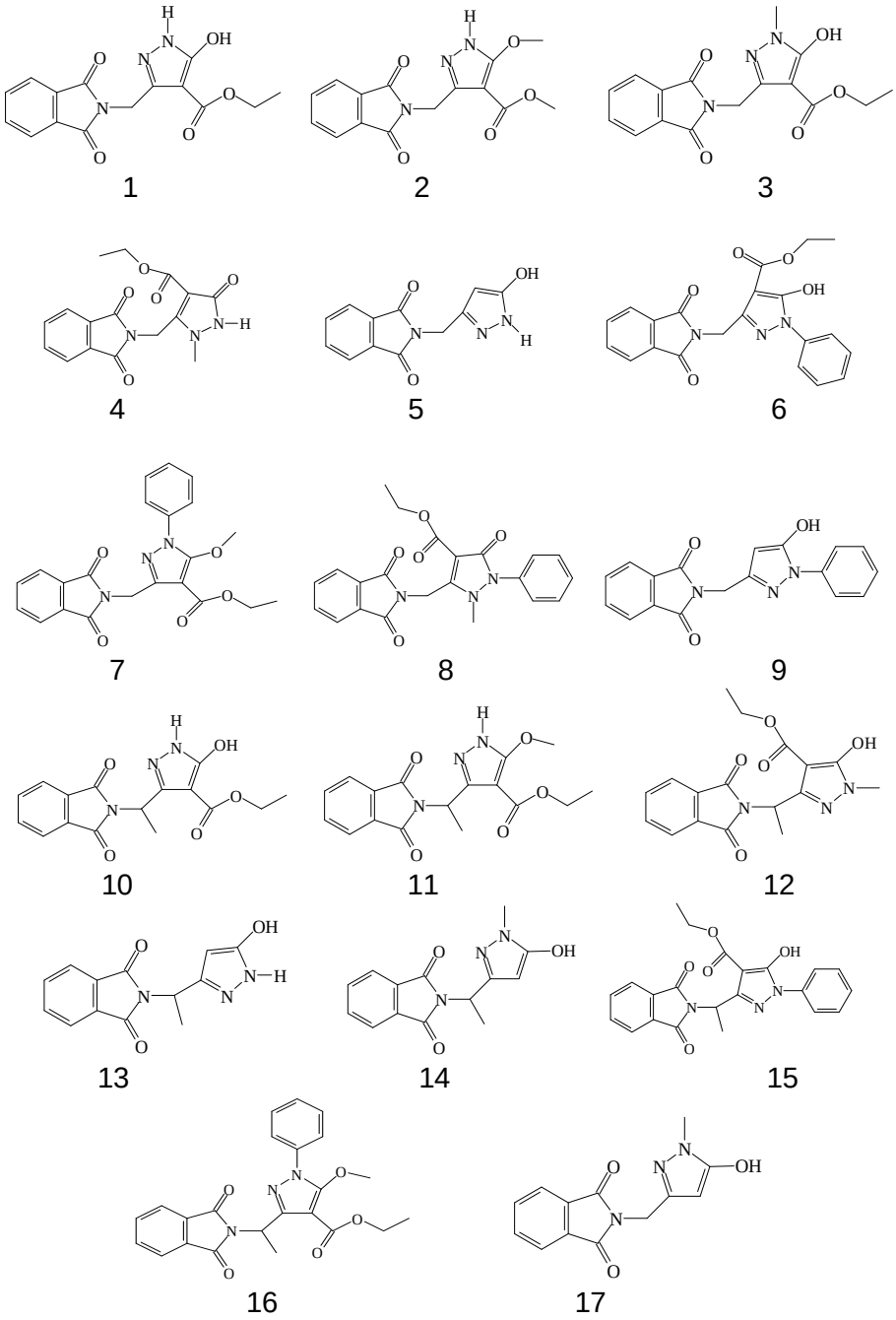


Figure 6. Structure of 17 substituted 3-(Phthalimidoalkyl)-pyrazolin-5-ones.

Monovariate Regression for Energy

The best results, in monovariable and divariable regression for energy are listed in Table 15.

Table 15.

The Bests Correlations of Energy in Monovariate and Divariate Regression.

Index No.	Index Name	R	b_0	b_1
1	lnDGjDeE_p/d2PE_	0.9997	5370	3760
2	DTjDeEp2/d2SE_	0.9996	8802.3	138
3	lnDGjDeE_p/d2PE2	0.9996	1289.5	3671.1
4	DTjDeMp2/d2SE_	0.9994	9188.7	922.34
4	DTjDeMp2/d2SE_	0.9999	13056	1108.8
4315	DTfDiE_p/d_AP2			-95.598
34	RTsDiM_p/d2GP2	0.99997	-1193	1674.3
5947	DTjDeEp2/d2AP_			-41.168
492	RTjDeM_p/d2SP_	0.99997	58267	46.095
1698	1/RTsDeM_p/d2AE2			-686800
492	RTjDeM_p/d2SP_	0.99998	56222	47.864
1737	1/RTsDeM_p/d2AP2			-711240

The best single variable QSPR (boldface in Table 15) was

$$\text{Predicted energy} = 5370 - 3760 * \ln DGjDeE\_p/d^2PE\_ \quad (36)$$

$$R = 0.99973;\ n = 17$$

This correlation could be satisfactory but usually a molecular property shows more than one dimension dependency. For this reason, we performed the bivariate regression.

Bivariate Regression for Energy

The first 16383 indices, labeled in decreasing order of their score in monovariate regression, are submitted for bivariate correlation. A procedure for finding subsets of optimal even number descriptors was developed. It is a simple, iterative technique that eludes the investigation of all possible descriptor combinations and reduces the time for drawing the best property model. More details will be presented in a future paper.

Here, the bivariate correlation for six pairs of indices is exemplified. The pairs are: (1, 2); (1, 10175); (4, 4315); (34, 5947); (492, 1698) and (492, 1737). The first two pairs are taken to show that the first scored index in monovariate regression does not provide the best bivariate correlation. Selection of the pairs of indices for bivariate correlation must be done by traversing the whole pool (1...16383). For additional descriptors, our procedure for optimum descriptor selection avoid the mining of all possible index combinations.

The best bivariate score was provided by the pair (492, 1737):

Predicted energy = 56221.885 + 47.864*RTjDeM_p/d2SP_ 711240.703*1/RTsDeM_p/d2AP2 (37)
R = 0.99998; s = 57.40; n = 17

An insight in Table 15 reveals that the best models (i.e., those showing R > 0.9999) show a dependency of this energy by the molecular topology (topological models) and the nature of atoms (mass and electronegativity).

Monovariate Regression for Inhibition

For the first six best indices in monovariate regression the indices and statistics are given in Table 16.

The best monovariate QSAR was:

Predicted inhibition = -336.760 + 96.378* lnDGsDeC_1/p_SE_ (38)
R = 0.9539; n = 17

which is, of course, not satisfactory, despite in ref.[21] a value of R = 0.92 was reported. Thus, we performed the bivariate regression.

Table 16.

The Best Correlations of Inhibition in Monovariate and Divariate Regression.

Index No.	Index Name	R	b_0	b_1
1	lnDGsDeC_1/p_SE_	0.9539	-336.76	96.378
2	1/DGsDeC_1/p_SE_	-0.9523	137.01	-4754.8
3	lnDTjDeE_p*d_HE_	0.9517	135.80	-493.02
18	DTjDeEp2/d2AE_	0.9883	121.21	1.076
16842	RGjDeP_p/d_GP_			-1.5194
37	DTjDeE_p/d_AE_	0.9906	-73.183	2.1644
11362	lnDGjDeP_p/d_PE2			-4.1769
4304	DTsDiM_p*d_HP_	0.9927	-26.846	1.5619
7649	DGjDeE_p/d2SE2			-1.7043

Bivariate Regression for Inhibition

Six pairs of indices are considered here for bivariate correlation: (1, 2); (1, 1369); (2, 13227); (18, 16842); (37, 11362) and (4304, 7649).

As in the case of energy, the best scored index in monovariate correlation is not present in the pair of best bivariate correlation.

The best bivariate score was done by the pair (4304, 7649):

Predicted inhibition = -26.846 + 1.562*DTsDiM_p*d_HP_ - 1.704*DGjDeE_p/d2SE2
R = 0.9927; s = 2.374; n = 17 (39)

Figure 7 illustrates the plot of inhibition vs predicted inhibition cf eq 39.

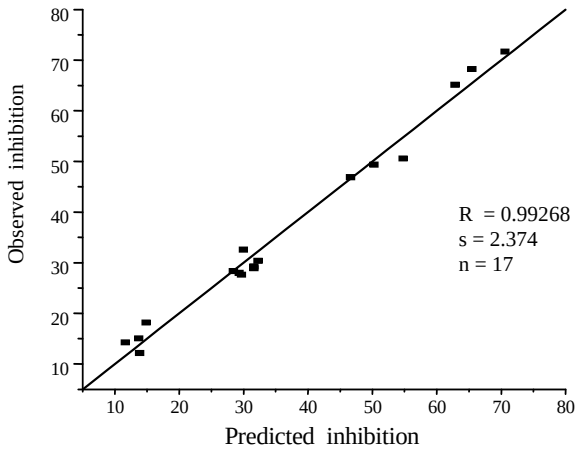


Figure 7. The plot: inhibition vs predicted inhibition cf eq 32.

The constant high correlation (see Table 16) between the best indices and the mitodepressive activity on *Lepidium Savitium L. (Cresson)* demonstrate ability of this family of indices to estimate the biological activity of the considered set of chemical structures. The models with R > 0.983 suggest that the mito-depressive activity on *Lepidium Savitium L. (Cresson)* is dependent both on the geometric and topological features of molecules, the nature of atoms (mass and electronegativity) and the electrostatic field of atoms induced by their partial charges.

Aromatase Inhibitors

A set of substituted dichlorodiphenyls (4, 4'-dichlorodiphenyl-methanes) inhibitors of aromatase[22] were considered. Enzymatic aromatization of androgens is involved in the biosynthesis of estrogens, and consequently in the estrogen-dependent diseases.

For modeling the inhibition, the authors[42] used two dipole moment related descriptors. We modeled the inhibition in monovariate regression but no satisfactory correlation (R^2 around 0.828) was found. In divariate regression, the correlation improved.

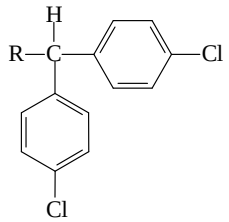
Predicted inhibition = 6.177 + 0.513*lnRTjDiP__p__HP2 – 0.071*1/DGjDeP_1/p_SP2

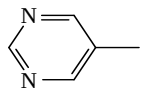
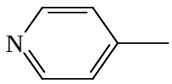
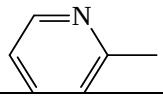
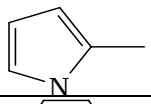
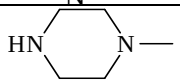
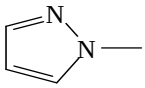
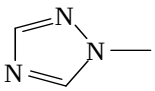
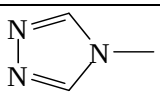
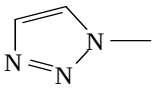
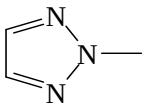
R^2 = 0.9716; s = 0.205; n = 10 (40)

The best reported[22] correlation for this subset was: R^2 = 0,89; s = 0.44. In our model, both the topology and geometry (see the indices in eq 40) are important in modeling the aromatase inhibition by dichlorodiphenyl methanes.

Table 17.

Dichlorodiphenyl Methanes Aromatase Inhibitors.



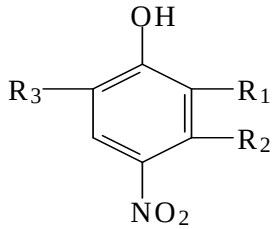
No.	R	-log EC_{50} obs
1		7.43
2		8.03
3		8.06
4		5.70
5		5.71
6		5.30
7		5.30
8		6.80
9		5.30
10		7.26

Nitrophenols

A set of 25 nitrophenols[27] showing herbicidal activity (Table 18) was considered for correlation with the Cluj Property indices. Nitrophenols are known to inhibit the electronic flux of photosynthesis.

Table 18.

Nitrophenols and Their Herbicidal Activity



No	R_1	R_2	R_3	pI_{50}
1	H	methyl	methyl	3.3
2	H	methyl	isopropyl	4.1
3	H	H	t-butyl	5.7
4	H	H	phenyl	4.35
5	H	H	cyclohexyl	4.85
6	Cl	methyl	methyl	4.89
7	Cl	methyl	isopropyl	6.07
8	Cl	H	t-butyl	6.88
9	Cl	H	phenyl	6.45
10	Cl	H	cyclohexyl	6.52
11	Br	methyl	methyl	5.25
12	Br	methyl	isopropyl	6.70
13	Br	H	t-butyl	6.15
14	Br	H	phenyl	6.52
15	Br	H	cyclohexyl	6.75
16	I	methyl	methyl	6.24
17	I	methyl	isopropyl	6.70
18	I	H	t-butyl	7.03
19	I	H	phenyl	6.86
20	I	H	cyclohexyl	6.65
21	NO_2	H	H	3.00
22	NO_2	H	methyl	3.70
23	NO_2	H	s-butyl	5.10
24	NO_2	H	t-butyl	5.79
25	NO_2	H	cyclohexyl	6.05

Table 19 lists the best scores of correlation in decreasing order. From this table it can be seen that the monovariate and divariate regression are not satisfactory. Additional variables are needed for good statistics (entries 6-13).

Table 19.

Mono- and Multivariate Regression for Nitrophenols

No	X_1	X_2	X_3	X_4	X_5	X_6	R
1	1						0.9499
2	6	155					0.9709
3	7	174					0.9697
4	11	10620					0.9665
5	5	260					0.9662
6	13028	15806	91	15636			0.9908
7	13028	15806	12	15398			0.9901
8	12	15806	13891	15749			0.9907
9	7	13028	382	14214			0.9893
10	13028	15806	12	15398	15228	15865	0.9985
11	12	15806	13891	15749	15648	16378	0.9967
12	7	13028	382	14214	13186	16282	0.9965
13	13028	15806	91	15636	14064	14943	0.9956

The best model is given in eq 41 (see also entry 10, Table 19):

Predicted activity = 8.062 - 0.003*RGsDeM_p/d2PE2 +
+ 0.395*1/DTsDiP_p*d_HE_ - 0.000008*1/RTsDiPp2/d2HE2
- 229.564*1/DGjDeMp2/d2PE2 + 0.003*RGjDiPp2/d2HP_
+ 0.004*DTsDeP_p*d_HP2

R = 0.9985; s = 0.067; n = 25 (41)

The plot of the predicted vs observed herbicidal activity, cf. eq 41, is shown in Figure 8.

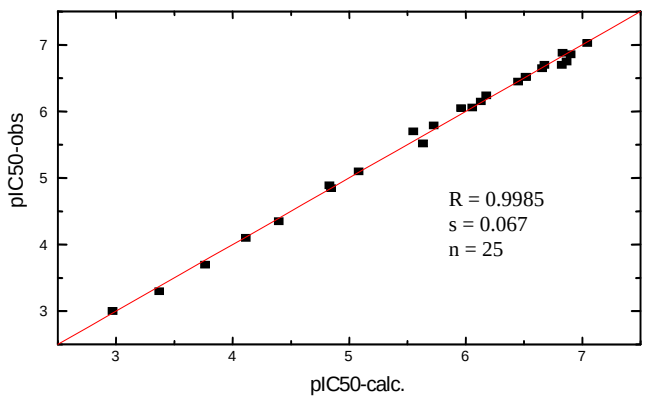


Figure 8. The plot of predicted vs. observed herbicidal activity.

The descriptors involved in eq 41 show a rather low inter-correlation (Table 20). The average absolute value of the pairwise correlation coefficients was 0.2200.

Table 20.

Intercorrellation of the indices in entry 10, Table 19.

	X_2	X_3	X_4	X_5	X_6
X_1	0.216489	0.6970	0.0674	0.1217	0.1710
X_2		0.1371	0.0150	0.0609	0.2646
X_3			0.2018	0.0397	0.2698
X_4				0.6824	0.2219
X_5					0.1337

The fragmental property indices take into account the chemical nature of atoms (mass and electronegativity), various kinds of interactions between the fragments of molecules and the 3D geometry of molecular structures.

Bivariate correlation with indices belonging to ***FPIF*** offer good quality models for quite diverse molecular properties such as the inhibition of mito-depressive activity on *Lepidium Savitium L.* (R > 0.99) and the aromatase inhibition as well. The same is true for the sum of one-electron energy calculated at the Extended-Hückel level (R > 0.9999).

Multivariate regression provided good models for the boiling points of a very diverse set of N-containing organic molecules or for the herbicidal activity.

Benzimidazole Derivatives

The correlating ability of *FPIF* was tested on a set of 15 molecular structures belonging to the class of benzimidazole[28]. Derivatives of benzimidazole are known to show various biological activities.[29,30]

The antiviral activity of a set of sexteen alkyl-benzimidazoles was proved by Tamm et al.[31] This set was studied by Kier and Hall[32] by using connectivity-type descriptors and recently by Estrada and Rodriguez,[33] by the aid of some sub-structural distance-based descriptors.

The first group of authors[32] founs the best model of the biological activity ***BA*** (as log(1/C)) of benzimidazoles after excluding one (of sixteen structures) N-methyl- derivative:

$$BA = 1.11 + 1.40\ {}^{6}\chi_P \qquad (42)$$

n = 15; r = 0.950; s = 0.166; F = 120.3

where ${}^{6}\chi_P$ is the molecular connectivity index of the sixth path order.[33] The authors supposed it may act by a different mechanism.

The second group,[33] foud another outlier in the remaining 15 members set: the compound no. 13 (see Table 18). They eliminated this compound by the following statistical tests: residuals, standardized residuals, studentized residuals and Cooks distance.[35,36]

The new equation, in terms of Kier and Hall,[32] is:

$$BA = 0.92 + 1.36\ {}^{6}\chi_{P} \tag{43}$$

n = 14; r = 0.971; s = 0.125; F = 195.1

Estrada and Rodriquez[33] have modeled the antiviral activity of these benzimidazoles by using the number of pairs of homodistant vertices of different length in the graph. Their bivariate regression equation is:

$$BA = 0.26 + 0.0884\eta_{12} + 0.0599\eta_{6} \tag{44}$$

n = 14; r = 0.976; s = 0.118; F = 110.7

where the first variable describes global molecular features and the second one is related to some specific paths in benzimidazoles.

Within this paper, we tried the modeling ability of the novel molecular descriptors *FPIF* on structures calculated by using MM+ (for 3D-geometries) and semiempirical AM_1 (for partial charge calculation) procedures of the HyperChem Program (HyperCube Inc.). The optimized geometries and partial charges thus obtained were submitted to the Cluj programes. Topological indices and several properties are presented in ref 28. Table 21 shows the statistics of the regression analysis.

In *monovariate regression*, only the index DTjDeE_p/d2PE2 (read: Dense Topological, CJ- Detour, Electronegativity, property per squared distance, Product, Edge-calculated index) succeeded in giving a better model (r = 0.9685; s = 0.132, n = 15, entry 11). The *leave-one-out* ***Loo*** procedure indicated the *structure 13 as an outlier*, confirming the finding of Estrada. Table 21 clearly shows improved results for all the used indices in the set of 14 structures. Again, the index DTjDeE_p/d2PE2 was the best.

Table 21.

Statistics of Benzimidazole Regressions.

	Index	*n*	*a*	b_i	*r*	*s*	*cv %*	*F*
1	IP(Di)	15	1.8872	0.0069	0.9102	0.2197	7.020	62.7504
2	lnIP(Di)	15	-3.2524	1.2456	0.9428	0.1767	5.646	104.0777
		14	-3.1411	1.2174	0.9729	0.1205	3.905	212.2853
3	IP(RDe)	15	1.6298	0.1308	0.9292	0.1960	6.263	82.1537
4	ln(IP(RDe))	15	-0.2685	1.4228	0.9420	0.1780	5.689	102.3426
		14	-0.2293	1.3920	0.9735	0.1191	3.859	217.5559
5	IP2(CJDi)	15	2.0610	0.0011	0.8964	0.2351	7.511	53.1575
6	ln(IP2(CJDi))	15	-4.3584	1.1072	0.9443	0.1745	5.577	106.9970
		14	-4.2002	1.0792	0.9703	0.1260	4.084	192.9996
7	IP2(CFDi)	15	2.0286	0.0010	0.9018	0.2291	7.322	56.6254
8	ln(IP2(CFDi))	15	-4.5475	1.1155	0.9433	0.1761	5.627	104.8861
		14	-4.3974	1.0890	0.9716	0.1232	3.992	202.5371
9	IP2(SZDi)	15	2.2051	0.0006	0.8946	0.2370	7.572	52.0955
10	ln(IP2(SZDi))	15	-3.4295	0.9105	0.9390	0.1823	5.826	96.9773
		14	-3.3325	0.8922	0.9725	0.1213	3.932	209.1839

Table 21. (continued)

	Index	*n*	*a*	b_i	*r*	*s*	*cv %*	*F*
11	DTjDeE_p /	15	7.3955	-0.0700	0.9685	0.1319	4.216	196.9635
	d2PE2	14	7.2489	-0.0680	0.9771	0.1109	3.595	252.6036
12	DTjDeE_p /	15	9.7711	-0.0950	0.9822	0.1036	3.309	164.4465
	d2PE2			0.0885				
	lnDTsDiM_p /	14	9.2202	-0.0888	0.9864	0.0893	2.896	198.4100
	d2PE2			0.0721				

Figure 9 shows the plot of DTjDeE_p/d2PE2 vs. ***BA***, in the set of 14 benzimidazoles. This result surpasses the Estrada's results both in mono and bivariate regression.

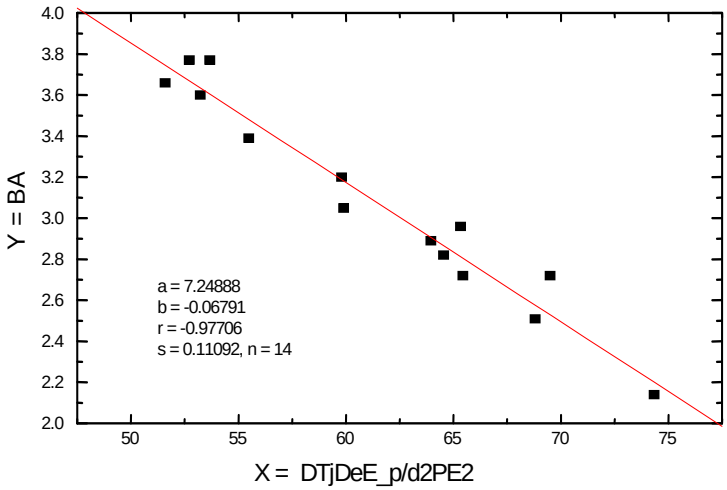


Figure 9. The best nomovariate regression model of antiviral activity of 14 Benzimidazoles

In bivariate regression, the best scored property index (above discussed) and lnDTsDiM_p/d2PE2 offered the best model (the covariance *cv* less than 3% - entry 12, Table 21) reported in literature.

BA = 9.2202 - 0.0888* DTjDeE_p /d2PE2 + 0.0721* lnDTsDiM_p/ d2PE2
n = 14; r = 0.9864; s = 0.0893; *cv %* = 2.896; F = 198.41 (45)

Note that, in monovariate regression, the best four *FPIF* descriptors are of topological model (T - in the index symbol) and among the first ten descriptors only three are of geometric model (G). Similarly, in bivariate regression, the best three couples of indices are of topological model and only the fourth pair is mixed topological and geometric. It appears that the topology, reflecting the basic structure of imidazoles, is the dominant feature involved in their antiviral activity. The geometry comes as a fine tuning, that cannot, alone, decide the basic activity.

This result suggests that: the more suitable molecular description, the less outliers in modeling a chosen property of a set of compounds. The question: how large should be the drop in standard deviation (or variation in other statistical parameter) for assuming the *outlier* status for a given structure (or better, for its *measured* property) - is a matter of choice. Recall that the topological descriptors (even the Cluj *property* indices) are only mathematical properties of the molecular graphs representing chemical compounds and, therefore, no direct causal relationship can be addressed to QSAR equations.

Urea Derivatives

It is known that hydroxyureas inhibit the enzymatic conversion of ribonucleotides to deoxyribonucleotides.[37] The molecular mechanism of this inhibition is not known but QSAR studies[38] suggested that position and identity of substituents may control the ability of hydroxyureas to complex some metallo-enzymes possibly involved in.

In *monovariate regression*, the best model shows the statistics: r = 0.96878; s = 9.096; cv% = 10.07 (column 4, Table 22). The best ten monovariate regressions show a variance cv% between 10 and 11.

In *bivariate regression*, the model is still improved:

$$\boldsymbol{BA} = 84.905 - 13.909*\text{lnDGjDiP\_\_p\_\_GP2} + 2.035*\text{DGsDeE\_p*d\_PP2} \quad (46)$$

n = 9; r = 0.99391; s = 4.367; cv % = 4.834; F = 244.081

The best ten monovariate regressions show a variance cv% between 5 and 6. Some best models, both in mono- and bivariate description, show a strong dependence of bioactivity by the molecular geometry (G in the symbol of indices) and electronic properties (partial charge P and electronegativity E).

Table 22.

Topological Data and Statistics for the Set of Hydroxyureas

No	Compound	*BA*	lnDGjDiP __p__GP2	lnRTsDeE_p /d2HP2	lnDGfDiP __p__GP2	DGsDeE_p *d_PP2	DGsDiE_p *d_PP2	lnDGsDeE _p*d_PP2
1	Hydroxyurea HU	100	1.4641	0.1038	1.4641	20.0153	20.0153	2.9965
2	N-Methyl-HU	133	0.0465	0.1020	0.0465	23.8930	23.8930	3.1736
3	N-Ethyl-HU	91	3.2254	0.1148	3.2254	25.2829	25.2829	3.2301
4	N-Acetyl-HU	111	2.0073	0.1059	2.0073	25.5269	25.5269	3.2397
5	3-Phenyl-1-HU	38	4.9244	0.1729	4.6653	8.7658	14.8164	2.1709
6	Di-HU	108	1.5595	0.1087	1.5595	21.8853	21.8853	3.0858
7	N-Hydroxyurethane	92	2.2266	0.1282	2.2266	20.6392	20.6392	3.0272
8	N-Hydroxyguanidine	110	1.7146	0.1089	1.7146	20.8342	20.8342	3.0366
9	3-Phenyl-1-hydroxy -2-thiourea	30	5.2621	0.1783	4.9578	10.5093	16.7358	2.3523
			(1)[a]	(2)	(18)	(2716)	(7478)	(2021)

Table 22. (continued)

No	Compound	*BA*	lnDGjDiP __p__GP2	lnRTsDeE_p /d2HP2	lnDGfDiP __p__GP2	DGsDeE_p *d_PP2	DGsDiE_p *d_PP2	lnDGsDeE _p*d_PP2
			monovar.			**bivar.**		
						lnDGjDiP _p__GP2 (1)	lnDGjDiP _p__GP2 (1)	lnDGfDiP _p__GP2 (18)
		r	**0.9688**	0.9658	0.9611	**0.9939**	0.9932	0.9913
		s	9.096	9.520	10.125	4.367	4.611	5.231
		cv%	10.070	10.539	11.208	4.834	5.105	5.791
		F	106.875	96.961	84.910	244.081	218.552	169.178
		b_0	139.260	229.056	140.865	84.905	71.330	20.173
		b_1	-19.632	-1111.106	-20.798	2.035	4.609	35.522
		b_2				-13.909	-12.234	-13.868

[a] score in monovariate regression

CONCLUSIONS

Despite a correlational model does not involve a causal relationship between descriptors and a molecular property. However, a look upon the nature of the best scored fragmental property indices can give insight of the type of intra- and/or intermolecular interactions. The results are encouraging in case of modeling the activity vs *Bacillus subtilis* and *Candida albicans* as well as for the Rf index. They demonstrate the usefulness of our descriptors in modeling biological and physical properties of organic compounds.

FPIF offer good description for various molecular properties of this class of compounds: the antimicrobial and antifungal activity, surface tension ε, the antiviral activity of benzimidazoles, enzyme inhibiting activity of hydroxyureas, vapor pressure of PCBs and TLC Rf index. The fragmental property indices take into account the chemical nature of atoms (mass, electronegativity and partial charge), various kinds of interactions between the fragments of molecules as generated by Cluj criteria and the 3D geometry of molecular structures as well. As it is known, a correlational model does not involve a causal relationship between descriptors and a molecular property. However, a look upon the nature of most occuring ***FIPF*** indices with the best scores (and implicitly best structure description) can give insight of the type of intra- and/or intermolecular interactions. The above results demonstrate the usefulness of our descriptors in modeling biological and physical properties of organic compounds.

The original Cluj type indices demonstrated a good ability in modeling some important physico-chemical properties. In some the particular case, the recorded results surpass that reported in literature and can be used in predicting studies. It represents a promise for further **QSPR/QSAR** studies.

REFERENCES

1. S. M. Free and J. W. Wilson, *A mathematical contribution to structure-activity studies*, J. Med. Chem. **1964**, 7, 395.
2. M.V. Diudea, *Cluj matrix* $\boldsymbol{CJ_u}$ *: source of various graph descriptors,* Commun. Math. Comput. Chem. (MATCH), **1997**, 35, 169-183.
3. M.V. Diudea, O.M. Minailiuc, G. Katona, I. Gutman, *Szeged matrices and related numbers*, Commun. Math. Comput. Chem. (MATCH), **1997**, 35, 129-143.
4. M.V. Diudea, *Cluj matrix invariants*, J.Chem.Inf. Comput. Sci. **1997**, 37, 300-305.
5. M.V. Diudea, B. Pârv, M.I. Topan, *Derived Szeged and Cluj indices,* J. Serb. Chem. Soc. **1997**, 62, 267-276.
6. A.A. Kiss, G. Katona, M.V. Diudea, *Szeged and Cluj matrices within the matrix operator* $W_{(M1,M2,M3)}$ Coll. Sci. Papers Fac. Sci. Kragujevac **1997,** 19, 95-107.
7. I. Gutman, M.V. Diudea, *Defining Cluj matrices and Cluj matrix invariants,* J. Serb. Chem. Soc. **1998**, 63, 497-504.
8. M.V. Diudea, G. Katona, I. Lukovits, N. Trinajstić, *Detour and Cluj-Detour Indices*, Croat. Chem. Acta, **1998**, 71, 459-471.
9. L. Jäntschi, G. Katona, M.V. Diudea, *Modeling molecular properties by Cluj indices*, Commun. Math. Comput. Chem. (MATCH), **2000**, 41, 151-188.
10. J.G. Topliss, R.P. Edwards, *Chance factors in in studies of Quantitative Structure- Activity Relationships*, J. Med. Chem., **1979**, 22, 1238.
11. M. Randić*, On Molecular Identification Numbers*, J.Chem.Inf.Comput.Sci. **1984**, 24, 164-175 .
12. F.R. Burden, *Molecular Identification Number for Substructure Search*, J. Chem. Inf. Comput. Sci. **1989**, 29, 225-227.
13. S.B. Elk, I. Gutman, *Further Properties Derivable from the Matula Numbers of an Alkane*, J.Chem.Inf.Comput.Sci. **1994**, 34, 54-57.
14. M.V. Diudea, A. Graovac, *Cyclic Graphs with Degenerate Sequences: A Study of Similarity*, Commun. Math. Comput. Chem. (MATCH), (submitted)
15. V. Pachaiyappan, S.H. Ibrahim, N.R. Kuloor, Chem. Eng. **1967**, 74, 193-196.
16. M. Cocchi, E. Johansson, *Amino Acids Characterization by GRID and Multivariate Data Analysis*. Quant. Struct.-Act. Relat., **1999**, 12, 1-8.
17. D. Opris, M.V. Diudea, *Peptide Property Modeling by Cluj Indices*, SAR/QSAR Environ.Res. **2001**, 12, 159-179
18. A. Zaliani, E.J. Gancia, *MS-WHIM Scores for amino-acids: a new 3D-description for peptide QSAR and QSPR Studies*. Chem. Inf. Comput. Sci., **1999**, 39, 525-533.
19. E.R. Collantes, W.J. Dunn III, *Amino acids side chain descriptors for Quantitative Structure-Activity relationship studies of peptide analogues*. J. Med. Chem., **1995**, 38, 2705-2713.
20. J. Jonsson, L. Eriksson, S. Hellberg, M. Sjostroem, S. Wold, *Multivariate parametrization of 55 coded and non-coded amino acids.* Quant. Struct.-Act. Relat., **1988**, 8, 203-209.
21. S. Nikolić, M. Medić-Sarić, J. Matijević-Sosa, *A QSAR study of 3-(Phtalimidoalkyl)-pyrazolin-5-ones*, Croat. Chem. Acta, **1993**, 66, 151-160.
22. P.I. Nagy, J. Tokarski, A.J. Hopfinger, *Molecular shape and QSAR analysis of a family of substituted dichlorodiphenyl aromatase inhibitors*, J. Chem. Inf. Comput. Chem. **1994**, 34, 1190-1197.
23. M.D. Wessel, P.C. Jurs, *Prediction of normal boiling points for a diverse set of industrially important organic compounds from molecular structure*, J. Chem. Inf. Comput. Sci. **1995**, 35, 841-850.

24. D.T. Stanton, P.C. Jurs, *Development and use of charged partial surface area structural descriptors for Quantitative Structure-Property Relationships studies.* Anal. Chem. **1990**, 62, 2323.
25. E.S. Goll, P.C. Jurs, *Prediction of the Normal Boiling Points of Organic Compounds from Molecular Structures with a Computational Neural Network Model*, J. Chem. Inf. Comput. Chem. **1999**, 39, 974-983.
26. A.J. Stuper, W.E. Brugger, P.C. Jurs, *Computer-assisted studies of chemical structure and biological function*, Wiley-Interscience: New York, **1979**.
27. A. Trebst, W. Draber, Advan. Pest. Sci., Symp. Papers IV-th Int. Congress Pest. Chem., Zurich, Swiss, **1978**, pp. 223.
28. G. Katona, G. Turcu, T. Kiss, O.M. Minailiuc, M.V. Diudea, *QSAR/QSPR Studies by Cluj and Szeged Descriptors*, Rev. Roumaine Chim. **2001**, 46, 137-151.
29. A. Korolkovas, *Essentials in Molecular Pharmacology*, Wiley, New York, **1970**.
30. C. D. Nenitescu, *Organic Chemistry*, E. D. P. Bucuresti, **1986**.
31. I. Tamm; K. Folkers; C. H. Shunk; D. Hely; F. L. Horofall J. Exp. Med. **1953**, 98, 245.
32. L. H. Hall; L. B. Kier J. Pharm. Sci. **1978**, 67, 1743.
33. E. Estrada; L. Rodriguez, *Decomposition of the Wiener number into contributions comming from homodistant pairs of vertices. Definition and a QSAR application*, J. Serb. Chem. Soc. **1997**, 62, 199-205.
34. L. B. Kier; L. H. Hall *Molecular Connectivity in Chemistry and Drug Research* Acad. Press, **1976**.
35. D. A. Besley; E. Kuh; R. E. Welsch *Regression Diagnostics*, Wiley, New York, **1980**.
36. M. D. Wessel; P. C. Jurs, *Predicting of normal boiling points of hydrocarbons from molecular structure.* J. Chem. Inf. Comput. Sci. **1995**, 35, 68-76.
37. C. W. Young; H. Hodas, Science, **1964**, 146, 1172.
38. P.V. Khadikar; S. Karmarkar; S. Sharma; A. D. Seerwani; S. Joshi, *Estimation of the inhibition of DNA synthesis by the Wiener index*, J. Serb. Chem. Soc. **1997**, 62, 219-226.
39. O. Hutzinger; S. Safe; V. Zitko, *The Chemistry of PCBs*, CRC Press, Cleveland, **1974**.
40. M. D. Erickson, *Analytical Chemistry of PCBs*, Butterworth, Boston, **1968**.
41. Report of the Environmental Directorate of the Organisation for Economic Co-operation and Development, Paris, **1973**, 44.
42. R. L. Dufree; G. Contos; F. C. Whitmore; J. D. Barden; E. E. Wackman; R. A. Westin Report of the U. S. Environmental Protection Agency, Office of Toxic Substances, No. EPC 560/6-76-005 (NITS No. PB-25012), **1976**, 488.
43. S. Jensen, New Sci. **1966**, 32, 612.
44. D. J. Findly; F. H. Siff; V. J. Declaro, Report of the U. S. Environmental Protection Agency, No. EPA 5607-76-001 (N.T.I. S. No. PB - 253735) **1976**, 143.
45. J. Jatsukawa in *PCB Poisoning and pollution*, K. Higueti, Ed., Acad. press, London, **1976**, 147.
46. D. H. Rouvray; W. Tatong, Int. J. Environ. Studies, **1989**, 33, 247.
47. S. Karmarkar; S. Karmarkar; S. Joshi; A. Das; P. V. Khadikar, *Novel application of Wiener vis-a-vis Szeged indices in predicting polychlorinated biphenils in the environment*, J. Serb. Chem. Soc. **1997**, 62, 227-234.
48. A.A. Kiss; G. Turcu; M. V. Diudea, *Correlating Studies by Cluj and Szeged Indices*, Studia Univ. Babes-Bolyai, **2001**, 45, 99-106.
49. M. Ardelean, G. Katona, I. Hopartean, M.V. Diudea, *Cluj Property Indices in Property Modeling*, Studia Univ. “Babes-Bolyai”, **2001**, 45, 81-95.
50. M. Ardelean, *Master Dissertation*, Babes-Bolyai Univ., Faculty of Chemistry and Chemical Engineering, **2000**.