STUDIA UNIVERSITATIS BABEŞ-BOLYAI, CHEMIA, XLV, 1-2, 2000

# CLUJ AND SZEGED INDICES IN PROPERTY MODELING

**M. ARDELEAN[A]; G. KATONA[B]; I. HOPARTEAN[A]; M. V. DIUDEA[A]**

*[a]Faculty of Chemistry and Chemical Engineering,*
*"Babeş-Bolyai" University, 3400 Cluj, Romania*
*[b]"BIOS" Research and Production Centre, 3400 Cluj Napoca, Romania*

***ABSTRACT.*** The novel Cluj and Szeged property indices are used for modeling biological and physico-chemical properties of a set of pirazolidin-diones (synthesised in our laboratory). These 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. They offer good description for the antimicrobial and antifungal activity, surface tension ε and TLC Rf index of the compounds belonging to this class and some insight in the nature of intra- and intermolecular interactions governing the investigated molecular properties.

## INTRODUCTION

A recent trend in structure-activity relationship (QSAR) and structure - property relationship (QSPR) is the use of topological and geometric parameters in evaluating and predicting the biological and physico-chemical properties of organic molecules. Topological indices (TIs) are single number descriptors of molecular topology and encode information regarding the size, shape, branching or centricity of molecular graphs. Geometric descriptors, such as the Euclidean distance between chemical substructures, total surface area and volume will complement the information supplied by TIs.[1-3]

Several molecular properties, of which numerical value vary with changes in the molecular structure, such as the normal boiling point, critical temperature and pressure, viscosity, solubility, retention chromatographic index, are often used for characterizing chemicals in databases. It happens that a certain property is not available in tables or other reference sources, or it is dangereus to determine (e.g. because the substance is too toxic or explosive). Some novel compounds, with an

assumed structure, are rather instable in given experimental conditions, or simply, a compound is not available. In such cases, methods of evaluating physico-chemical properties from the structural features of organic molecules are welcome.[2-6]

Another interesting field of research in biochemistry and pharmacology is the rationalization of the action of classes of chemicals with specialized modes of action. Specificity in enzymology, imunology and toxicity arises out of specific structural features which lead to particular types of interaction, within the complex effector-receptor. Topological and geometric descriptors were used in many QSAR studies in the above domain.[7,8]

The present paper reports an attempt to model the biological (antibiotic and antimycotic) and physico-chemical (chromatographic retention) properties of a group of 15 pirazolidin-3,5-diones (synthetised in our laboratoties) by Cluj and Szeged property indices.

## CLUJ AND SZEGED THEORETICAL DESCRIPTORS

The graph-theoretical descriptors[9-14] *CJ*, *CF* and *SZ* represent the theoretical ground for counting the fragmental property indices. They are derived from the cardinality of the vertex sets defined by:

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

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

$$SZ_{i,j} = \{v \mid v \in V(G);\ di(G)_{v,i} < di(G)_{v,j} \quad (3)$$

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*.

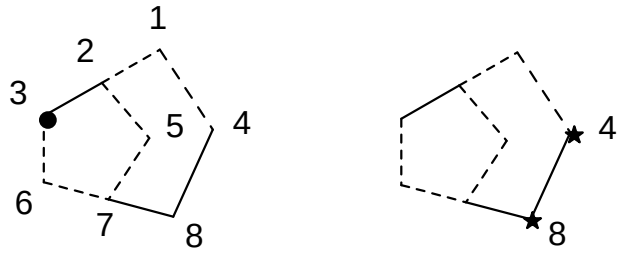
In defining *Cluj indices*, the *path p* plays the central role in selecting the subgraphs (eqs 1 and 2), particularly in cycle-containing graphs, where more than one path could join the pair (*i,j*). In such graphs, more than one subgraph (i.e. fragment), referred to *i*, can be counted. By this reason,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

$$[\mathbf{UM}]_{ij} = \max_{p} \left| \mathbf{V}_{i,j,p} \right| \quad (4)$$

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**). W**hen $p \in De(G)$, (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 **UCJDe**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 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 entries in the *unsymmetric Szeged distance matrix*, **USZDi,** are supplied by the cardinality of the sets in eq 3. Note that in defining the Szeged fragments, the path joining the vertices *i* and *j* is irrelevant. Thus, for each pair (*i,j*) it results one and only one fragment.

When the distance criterion $di(G)_{v,i} < di(G)_{v,j}$ (eq 3) is changed by the *detour criterion* $de(G)_{v,i} < de(G)_{v,j}$, the cardinality of the sets thus supplied represent entries in the unsymmetric Szeged detour matrix, **USZDe**.

The above definitions hold for any connected graph.

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

$$\mathbf{SM_p = UM \bullet (UM)^T} \quad (5)$$

$$\mathbf{SM_e = SM_p \bullet A} \quad (6)$$

The symbol • 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 6, 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.

## FRAGMENTAL PROPERTY INDICES

### Model Parameters

In physical phenomena, the macroscopic interactions are often interactions of field-type. The field is produced by a scalar function of potential. Let *f*(*x, y, z*) be such a scalar function. The field induced by this function can be written as:

$$\vec{\nabla} \cdot f = \left( \frac{\partial}{\partial x}\vec{i} + \frac{\partial}{\partial y}\vec{j} + \frac{\partial}{\partial z}\vec{k} \right) \cdot f(x,y,z) = \frac{\partial f}{\partial x}\vec{i} + \frac{\partial f}{\partial y}\vec{j} + \frac{\partial f}{\partial z}\vec{k} \qquad (7)$$

For the potential of type

$$f(x, y, z) = pz \qquad (8)$$

the associated field can be derived as

$$\vec{\nabla} \cdot f = \frac{\partial f}{\partial x}\vec{i} + \frac{\partial f}{\partial y}\vec{j} + \frac{\partial f}{\partial z}\vec{k} = \frac{\partial(pz)}{\partial x}\vec{i} + \frac{\partial(pz)}{\partial y}\vec{j} + \frac{\partial(pz)}{\partial z}\vec{k} = 0\vec{i} + 0\vec{j} + p\vec{k} = p\vec{k} = \vec{p} \qquad (9)$$

This is the case of the well-known uniform gravitational field:

$$\vec{G} = m\vec{g} \qquad (10)$$

with the corresponding potential given by

$$E_p = E_p(z) = mgz \qquad (11)$$

where *m* is the mass of the probe and *z* is the reference coordinate.

Note that eq 9 is applicable both to the Newtonian (gravitational) interactions and the Coulombian (electrostatic) interactions. In both cases the relation is valid if the mass *m* (or the charge *q*) that generates the potential *f* and associated field $\vec{\nabla} \cdot f$ is far enough (*r* >> *z*) so that the approximation $(r + z)^2/r^2 = (r^2 + 2rz + z^2)/r^2 = 1 + 2z/r + (z/r)^2 \cong 1$ is valid.

For the potential of type:

$$f(x, y, z) = p/z \qquad (12)$$

eq 7 leads to the associated field:

$$\vec{\nabla}\cdot f=\frac{\partial f}{\partial x}\vec{i}+\frac{\partial f}{\partial y}\vec{j}+\frac{\partial f}{\partial z}\vec{k}=\frac{\partial(p/z)}{\partial x}\vec{i}+\frac{\partial(p/z)}{\partial y}\vec{j}+\frac{\partial(p/z)}{\partial z}\vec{k}=$$
$$=0\vec{i}+0\vec{j}+\frac{-p}{z^2}\vec{k}=-\frac{p}{z^2}\vec{k}=-\frac{p}{z^3}\vec{z}=-\frac{\vec{p}}{z^2} \quad (13)$$

This is the case of well-known (non-uniform) gravitational field:

$$\vec{G}=\vec{G}(m,r)=-k\frac{m}{r^3}\vec{r} \quad (14)$$

and the associated potential of the form:

$$U=U(m,r)=k\frac{m}{r} \quad (15)$$

where *m* is the mass of the probe and *r* is the position relative to the location of the point producing the field.

For the Coulombian field eq 13 becomes:

$$\vec{F}_C=\vec{F}_C(r)=-k\frac{q}{r^3}\vec{r} \quad (16)$$

and the potential associated to the Coulombian field:

$$U=U(q,r)=k\frac{q}{r} \quad (17)$$

Four models were implemented in the view of building the *fragmental property indices*: two of them *topological* (dense topological and rare topological) and two others *geometric* (dense geometric and rare geometric). In these models a *weak dependence on distance* for the potential of the type (8) generating a uniform field (9), and a *strong dependence on distance* for the potential of the type (12) that generates a non-uniform field (13) were considered.

The variables in the models are: *property* $\Phi$ (mass *M*, electronegativity *E*, cardinality *C*, partial charge or any other atomic property *P*), *property descriptor* $\Omega$ (*p, d, pd, 1/p, 1/d, p/d, p/d2, p2/d2*) and *superposition* $\Psi$ (*S, P, A, G, H*).
The expressions for the *property descriptors* are:

$$\Omega: p=\mathrm{p};\ d=\mathrm{d};\ pd=p\cdot d;\ 1/p=\frac{1}{p};\ 1/d=\frac{1}{d};\ p/d=\frac{p}{d};\ p/d2=\frac{p}{d^2};\ p2/d2=\frac{p^2}{d^2} \quad (18)$$

where *p* is any property ($p\in\Phi$) and *d* is any metric of distance.

The (mathematical) superposition $\Psi$, given by

$$\Psi:\ S=\sum_{i=1}^{n}x_i;\ P=\prod_{i=1}^{n}x_i;\ \mathrm{A}=S/n;\ G=\left(\mathrm{sgn}(\mathrm{P})\right)^n\cdot\sqrt[n]{abs(P)};\ H=\left(\sum_{i=1}^{n}\frac{1}{x_i}\right)^{-1} \quad (19)$$

is applied upon a string of vertex descriptors to give a fragment descriptor. The used symbols are: *S = sum*; *P = product*; *A = arithmetic mean*; *G = geometric mean* and *H = harmonic sum*. The summation is suitable in case of any additive property (mass, volume, partial charges, electric capacities, etc.). The other operators find appropriate justification.

## 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*:

$$PD(Fr_{i,j}) = \underset{v \in Fr_{i,j}}{\Psi} ( \Omega ( d_{T\,v,j}, p_v ) ) \qquad (20)$$

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*:

$$PD(Fr_{i,j}) = \Omega ( d_{T\,i,j}, \underset{v \in Fr_{i,j}}{\Psi} ( p_v )) \qquad (21)$$

**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):

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

$$d_{E\,CP,j} = \Omega_p^{-1} ( DG(Fr_{i,j}), P(Fr_{i,j}) ), \qquad (22)$$

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.):

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

$$CP_i(x_{CP_i,j}, y_{CP_i,j}, z_{CP_i,j});\ x_{CP_i,j} = \sum_{v \in Fr_{i,j}} x_v \cdot p_v \Big/ \sum_{v \in Fr_{i,j}} p_v \qquad (23)$$

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

Some Particular Fragmental Property Models were discussed elsewhere.[14]

**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 $P_{i,j} = \{ p^1_{i,j}, p^2_{i,j}, ..., p^k_{i,j}\}$ paths joining *i* and *j*, then cf. *CJ* or *CF* definition (eqs 1-3), the fragments $Fr^1_{i,j}, Fr^2_{i,j}, ..., Fr^k_{i,j}$ 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.:

$PD(Fr^{\sigma_1}_{i,j})$, $PD(Fr^{\sigma_2}_{i,j})$, ..., $PD(Fr^{\sigma_m}_{i,j})$

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

$$PD_{i,j} = \frac{\sum_{t=1}^{m} PD(Fr^{\sigma_t}_{i,j})}{m} \qquad (24)$$

The resulting matrices are in general *unsymmetric* but they can be symmetrized (see eqs 5, 6). The symbols for the fragmental property matrices will be detailed below.

**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:

$$P_{-}(M) = \tfrac{1}{2}\Sigma\Sigma[\mathbf{M}]_{i,j} \quad ; \quad P2(M) = \tfrac{1}{2}\Sigma\Sigma[\mathbf{M}]_{i,j}[\mathbf{M}]_{j,i};$$

$$E_{-}(M) = \tfrac{1}{2}\Sigma\Sigma[\mathbf{M}]_{i,j}[\mathbf{A}]_{i,j}\ ;\ E2(M) = \tfrac{1}{2}\Sigma\Sigma[\mathbf{M}]_{i,j}[\mathbf{M}]_{j,i}[\mathbf{A}]_{i,j} \qquad (25)$$

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** (26)

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** ∈ $\Phi$ (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** ∈ $\Omega$ (i.e. **fffff**∈{**__p__**, **_1/p_**, **__d__**, **_1/d_**, **_p.d_**, **_p/d_**, **_p/d2**, **p2/d2**}

**G** ∈ $\Psi$ (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* (27)

where:

*ii* ∈ {*P_*, *P2*, *E_*, *E2*} with the known meaning (eq 25).

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

$$\begin{aligned} &\textit{lnABcDdEfffffGii} := \ln(\textit{ABcDdEfffffGii}); \\ &\textit{1/ABcDdEfffffGii} := \frac{1}{\textit{ABcDdEfffffGii}} \end{aligned} \quad (28)$$

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*** original 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

$$\mathbf{Y_i = b_0 + \sum_j^m b_{ij} X_{ij}} \quad (29)$$

The regression coefficients $b_{ij}$ can be determined by the least-squares method. Eq (29) 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.[15]

Within this work, biological activities as well as some physical properties of a set of 15 pirazolidin-3,5-diones were modeled by using ***FPIF*** descriptors.

**The set of pirazolidin-3,5-diones**

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

1. *Modeling Biological Activity*

Pirazolidin-3,5-diones are known having antiinflamatory activity.[16] They also show some antimicrobial and antifungal activity (on Staphylococus aureus, Bacilus subtilis, Escherichia coli, Pseudomonas aeruginosa, Candida albicans, etc.). Table 1. shows the biological activity, in mm inhibition zone.

***Table 1.*** 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 |

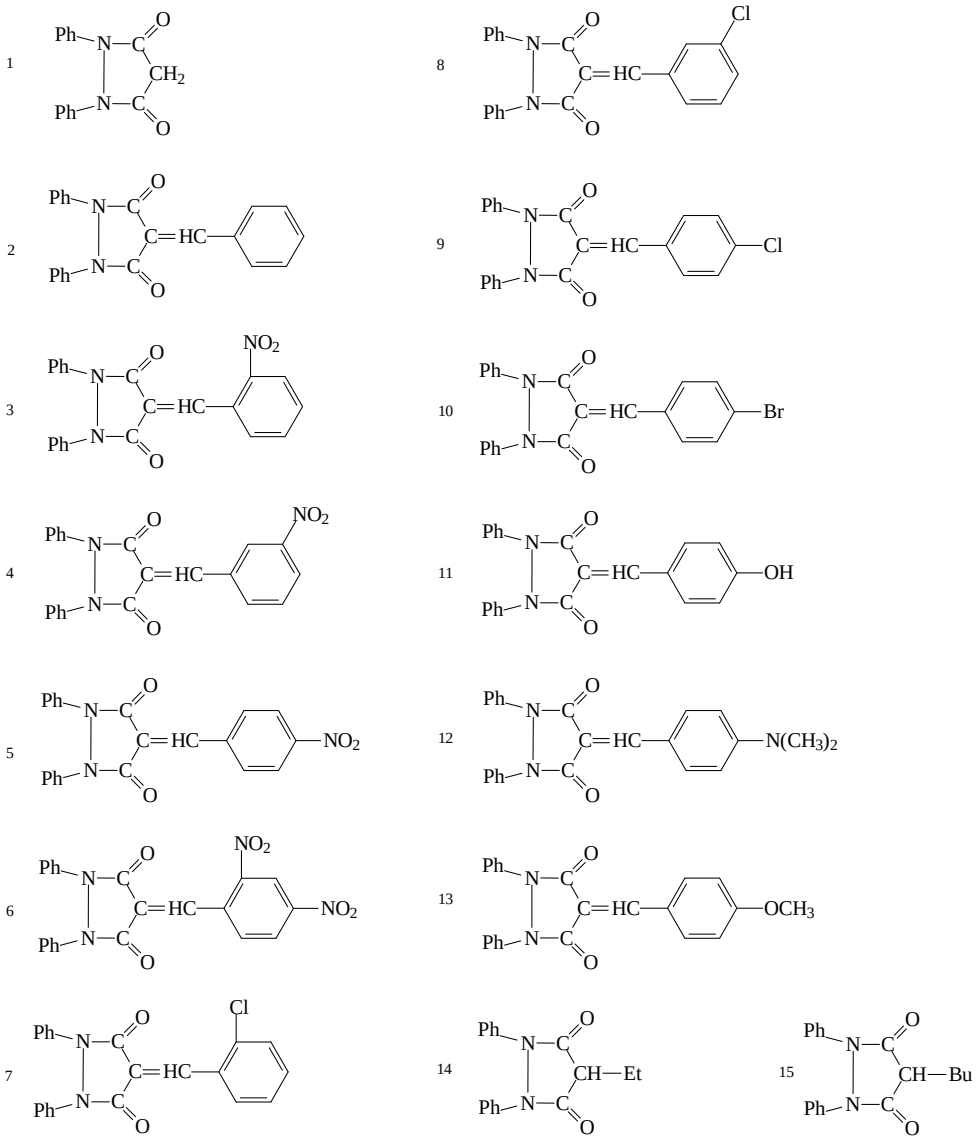



***Figure 2.*** The set of Pirazolidin-3,5-diones

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 minovariate regression. The best three regression equations are given below:

$BA_{calc}$ = – 147,1950 + 35,4337* lnRGjDeE_p/d_HP2 (30)

n = 15; r = 0.8320

$BA_{calc}$ = – 147,3113 + 35,4626* lnRGfDeE_p/d_HP2 (31)

n = 15; r = 0.8319

$BA_{calc}$ = – 75,9515 + 35,6724* lnRGjDeC_p/d_HP2 (32)

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_ (33)

n = 15; r = 0.9857

$BA_{calc}$ = 6.2856 - 578.8831*1/DGjDeP_p/d2PE_ +
4.8740*lnDGjDeP_p*d_GE_ (34)

n = 15; r = 0.9883

$BA_{calc}$ = 4.0166 - 191.7906*1/RGjDeC_p/d2HE_+
4.9157*lnDGjDeP_p*d_GE_ (35)

n = 15; r = 0.9885

Figure 3 shows the plots given by eqs 33-35.

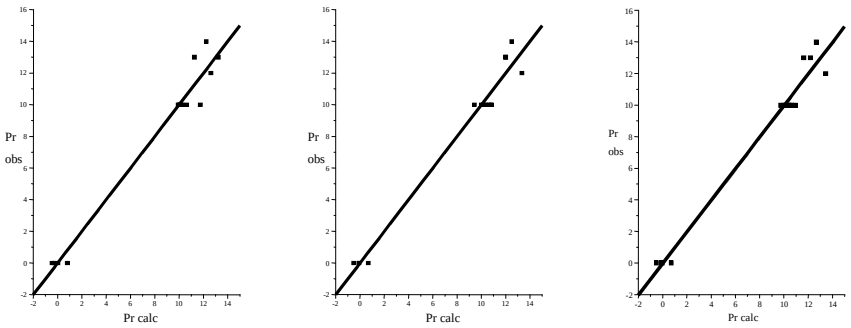



***Figure 3.*** Bivariate regression: Plots $BA_{obs}$ vs. $BA_{calc}$ cf. eqs 33-35.

Table 2 includes the observed inhibitory activity vs. *Bacillus subtilis* and calculated BA by the above equations.

***Table 2.*** Biological Activity $BA_{obs}$. and $BA_{calc}$ by eqs 33-35.

| Comp. No. | $BA_{obs}$ | BA (eq 33) | BA (eq 34) | BA (eq 35) |
|---|---|---|---|---|
| 1 | | 9.9993 | 9.99972 | 9.99975 |
| 2 | | 10.22805 | 10.60008 | 10.60684 |
| 3 | | 9.9127 | 10.19882 | 10.1479 |
| 4 | | 10.58298 | 10.36314 | 10.41585 |
| 5 | | 10.07425 | 9.42429 | 9.72681 |
| 6 | | 12.21854 | 12.49914 | 12.65926 |

| | | | | |
|---|---|---|---|---|
| 7 | | 13.21094 | 12.01476 | 12.15526 |

***Table 2*** (continued)

| | | | | |
|---|---|---|---|---|
| 8 | | 11.72823 | 10.77013 | 10.90124 |
| 9 | | 0.01434 | 0.68207 | 0.69076 |
| 10 | | 0.79884 | -0.07073 | -0.04351 |
| 11 | | -0.46738 | -0.53454 | -0.5304 |
| 12 | | 12.60716 | 13.32932 | 13.42582 |
| 13 | | 11.2599 | 11.98897 | 11.61886 |
| 14 | | 10.13089 | 10.84914 | 10.37846 |
| 15 | | -0.29874 | -0.1143 | -0.15289 |

As can be seen from eqs 30-35, 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 (36)
n = 15; r = 0,9252

$BA_{calc}$ = - 4.1732 + 1.5461* lnDTjDeP_p*d_PP2 (37)
n = 15; r = 0,9235

$BA_{calc}$ = - 2.3616 + 1.4733* lnDTfDiP_p*d_PP2 (38)
n = 15; r = 0,8777

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

$BA_{calc}$ = 58.0019 + 1.9258 lnDTfDiP_p*d_PP2 – (39)
14.1524*lnRGsDiEp2/d2GP2
n = 15 ; r = 0.9415

$BA_{calc}$ = 39.1986 + 1.9336*lnDTfDiP_p*d_PP2 – (40)
18.7211*lnRGsDiE_p/d2AP2
n = 15; r = 0.9429

$BA_{calc}$ = 7.0326 + 2.3522 lnDTjDiP_p*d_PP2 – (41)
42.8766 RGfDeP_p/d_AP2
n = 15; r = 0.9523

Figure 4 shows the plots given by eqs 39-41.

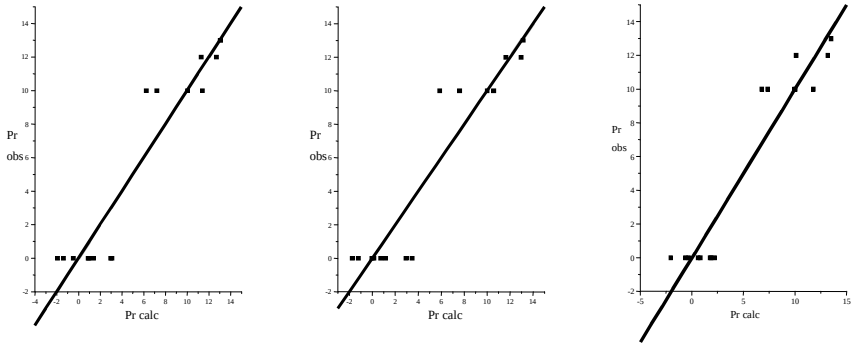



***Figure 4.*** Bivariate regression: plots $BA_{obs}$ vs. $BA_{calc}$ cf. eqs 39-41.

***Table 3.*** Biological Activity $BA_{obs}$. and $BA_{calc}$ by eqs 39-41.

| Comp. No. | $BA_{obs}$ | BA (eq 39) | BA (eq 40) | BA (eq 41) |
|---|---|---|---|---|
| 1 | | 9.99997 | 10.0001 | 9.99971 |
| 2 | | 3.04322 | 3.45234 | 2.19974 |
| 3 | | -0.50416 | -0.04724 | -2.04358 |
| 4 | | -1.94599 | -1.75724 | -0.63512 |
| 5 | | 11.38851 | 10.55436 | 11.786 |
| 6 | | 6.21494 | 5.85683 | 6.79134 |
| 7 | | 1.34103 | 0.11115 | 0.59396 |
| 8 | | -1.41583 | -1.24372 | -0.31398 |
| 9 | | 2.93834 | 2.92499 | 1.88119 |
| 10 | | 0.91587 | 0.71163 | 0.8347 |
| 11 | | 12.66418 | 12.95567 | 13.16947 |
| 12 | | 13.03083 | 13.13947 | 13.49032 |
| 13 | | 11.26882 | 11.61634 | 10.11394 |
| 14 | | 7.1924 | 7.60017 | 7.34929 |
| 15 | | 0.86788 | 1.12515 | 1.78304 |

From eqs 36-41, 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.

2. *Modeling Physico-Chemical Properties*

Two physico-chemical properties were considered: the surface tension ε, (in Dyn/cm$^2$) and the chromatographic Rf values.

The ***surface tension*** ε,[16] was calculated by eq

$$\varepsilon=(P_i/M_v)^4 \qquad (42)$$

where $P_i$ is the parachor and $M_v$ represents the molar volume. The values for ε are included in Table 4, along with FPIF descriptors showing the best scores in monovariate regression. The best mono- and bivariate QSPRs are listed below:

ε = - 65.7844 + 1.9019* DTjDeE_p/d2PE2 (43)
n = 15; r =0.8500

ε= 2597 -0.2155*DTfDeE_p/d2PP2 - 39,3342*1/RTjDeP_p/d_GP2 (44)
n = 15; r = 0,9299

The models are basically topological (see eqs 43-44). However, the results are far from those required in predicting studies.

The ***TLC Rf values*** were studied in the following. These values could give information about the global charge distribution as well as the steric feature of molecules.

Four solvent systems were considered for modeling Rf. The best scored regression equations are listed and plotted below:

***Monovariate regression***:

***Solvent system*** 1

$Rf_{calc}$ = -22.0579 – 220.8479*1/RGfDeP_p/d_HP_ (45)
n = 15; r = 0.9683

***Solvent system*** 2

$Rf_{calc}$ = 69.6434 - 231.0150*1/RTsDiPp2/d2HP2 (46)
n = 15; r = 0.8251

***Solvent system*** 3

$Rf_{calc}$ = 6.6979 - 151.3253*1/RGfDeP_p/d_HP_ (47)
n = 15; r = 0.8896;

***Solvent system*** 4

$Rf_{calc}$ = 9.5721 - 181.2313*1/RGfDeP_p/d_HP_ (48)
n = 15; r = 0.8856

The plot supplied by eq 45 is shown in Figure 5.

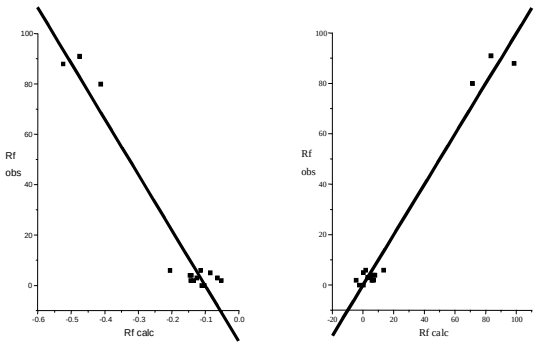



***Figure 5.*** Mono- and bivariate regressions: solvent system 1, plots $Rf_{obs}$ vs. $Rf_{calc}$ cf. eqs 45 and 49.

***Bivariate regression*:** the best scored regression equations are presented.

***Solvent system*** 1

$Rf_{calc}$ = -26.7898 - 261.2937*1/RGfDeP_p/d_HP_ (49)
+ 0.01356*DGjDeP__p__SP_
n = 15; r = 0.9866

***Solvent system*** 2

$Rf_{calc}$ = 66.0139 + 0.0005*1/RTsDiPp2/d2HP2 – (50)
0.0213*RTjDeP_1/p_AE2
n = 15; r = 0.9361

***Solvent system*** 3

$Rf_{calc}$ = 63.5664 + 14909.1133*lnRGjDeP_p/d_HP_ - (51)
38.3984*RTfDePp2/d2PP2
n = 15; r = 0.9575

***Solvent system*** 4

$Rf_{calc}$ = 15.9934 + 2.8441*1/RGfDeP_p/d_HP_ - (52)
162.3196*1/RTjDiP_p/d_GP_
n = 15; r = 0.9566

The plot supplied by eq 49 is shown in Figure 5.

As can be seen from eqs 45-52, only the solvent system 1 Rf values are acceptably modeled by FPIF descriptors. The model is geometric (G) and the local property is the partial charge (P), as just expected for this property. Recall that Rf is strongly influenced by the steric and electronic properties of molecules in the mobile phase.

## Conclusions

Some biological and physico-chemical properties of a set of pirazolidin-diones (synthesised in our laboratory) were modeled by the aid of FPIF descriptors. These 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.

***FPIF*** offer good description for various molecular properties of this class of compounds: the antimicrobial and antifungal activity, surface tension ε and TLC Rf index.

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.

***Acknowledgement***. This work is under financial support of GRANT CNSIS, T 34, 2000.

## REFERENCES

1. Balaban, A. T., editor, From Chemical Topology to Three-Dimensional *Geometry*, Plenum Publishing Corporation, New York. (a) Chapter 1 by Balaban, A.T. *From Chemical Graphs to 3D Molecular Modeling*", pp. 1-24. (b) Chapter 2 by Mezey, P. G., *Descriptors of Molecular Shape in 3D*", pp. 25-42. (c) Chapter 3 by Mekenyan, O.; Veith, G. D., *3D Molecular Design: Search for Active Conformers in QSAR*, pp. 43-72. (d) Chapter 4 by Basak, S. C.; Grunwald, G. D.; Niemi, G. J., *Use of Graph-Theoretic and Geometric Molecular Descriptors in Structure-Activity Relationships*, pp. 73-116. (e) Chapter 6 by Randic, M.; Razinger, M., *On Characterization of 3D Molecular Structure*, 1997, pp. 159-236.
2. Devillers, J.; Balaban, A. T., editors, *Topological Indices and Related Descriptors in SAR and QSPR*, Gordon and Breach, Reading, UK, 1999.
3. Diudea, M. V., editor, *QSAR/QSPR Studies by Molecular Descriptors*, Nova Science, Huntington, N.Y. (in press).
4. Katritzky, A. R.; Gordeeva, E. V., *Traditional Topological Indices vs. Electronic, Geometrical and Combined Molecular Descriptors in QSAR/QSPR Research*, J. Chem. Inf. Comput. Sci., 1993, 33, 835-857.
5. Suzuki, T. Ohtaguchi, K.; Koide, K. *Computer-assisted approach to develop a new prediction method of liquid viscosity of organic compounds*, Computers Chem. Engng. 1996, 20, 161-173.
6. Viswanadhan, V. N.; Ghose, A. K.; Singh. U. C.; Wendoloski, J. J. *Prediction of* solvation free energies of small organic molecules: additive-constitutive models based on molecular fingerprints and atomic constants, J. Chem. Inf. Comput. Sci. 1999, 39, 405-412.
7. Basak, S. C.; Gute, B. D.; Ghatak, S. *Prediction of complement-inhibitory activity of benzamidines using topological and geometric parameters*, J. Chem. Inf. Comput. Sci. 1999, 39, 255-260.
8. Raychaudhury, C.; Banerjee, A.; Bag, P.; Roy, S. *Topological Shape and Size of Peptides: identification of potential allele specific helper T cell antigen sites*, J. Chem. Inf. Comput. Sci. 1999, 39, 248-254.
9. Diudea, M. V. *Cluj matrix invariants*, J.Chem.Inf.Comput.Sci. 1997, 37, 300-305.
10. 10. Diudea, M. V. *Cluj Matrix $CJ_u$ : source of various graph descriptors*, Commun. Math. Comput. Chem. (MATCH), 1997, 35, 169-183.
11. Diudea, M., Minailiuc V. O., Katona G. and Gutman I. *Szeged Matrices and Related Numbers*, Commun. Math. Comput. Chem. (MATCH), 1997, 35, 129-143.

12. Diudea, M. V., Pârv, B. and Topan, M. I. *Derived Szeged and Cluj Indices*, J. Serb.Chem.Soc., 1997, 62, 267-276.

13. Minailiuc O., Katona G., Diudea, M. V., Strunje, M., Graovac, A. and Gutman, I. *Szeged Fragmental Indices*, Croat. Chem. Acta., 1998, 71, 473-488.

14. 14. Jäntschi, L., Katona, G. and Diudea, M. V. *Modeling Molecular Properties by Cluj Indices*, Commun. Math. Comput. Chem. (MATCH), 2000, 41, 151-188.

15. Topliss, J. G.; Edwards, R. P. *Chance factors in in studies of Quantitative Structure-Activity Relationships*, J. Med. Chem. 1979, 22, 1238.

16. 16. Ardelean, M. Master Dissertation, Babeş-Bolyai Univ., Faculty of Chemistry and Chemical Engineering, 2000.