Peptide Property Modeling by Cluj Indices

Authors: D. Opris [a]; M. V. Diudea [a]

Affiliation: [a] Department of Organic Chemistry, Faculty of Chemistry and Chemical Engineering, "Babes-Bolyai" University. Cluj-Napoca, 3400. Romania

DOI: 10.1080/10629360108035377

Publication Frequency: 8 issues per year

Published in: **SAR and QSAR in Environmental Research**, Volume **12**, Issue **1 & 2** April 2001 , pages 159 - 179

Subjects: **Applied & Industrial Chemistry**; **Chemistry**; **Environmental & Ecological Toxicology**; **Environmental Sciences**; **History & Philosophy of Mathematics**;

Formats available: PDF (English)

Abstract

The novel Cluj property indices are used for modeling the biological properties of dipeptides: the ACE inhibition activity of a set of 58 dipeptides and the bitter tasting activity of a set of 48 dipeptides, taken from the literature. The results are compared to those reported in some previous works.

Keywords: Molecular graph; Topological index; Peptide; Cluj indices

PEPTIDE PROPERTY MODELING BY CLUJ INDICES

D. OPRIS and M. V. DIUDEA*

Department of Organic Chemistry, Faculty of Chemistry and Chemical Engineering, "Babeş-Bolyai" University, 3400 Cluj-Napoca, Romania

SAR/QSAR Environ.Res. 2001**, *12*, 159-179.**

Dedicated to Professor Alexandru T. Balaban, for his bright contribution to Chemical Graph Theory and development of this science branch in Romania

* Author to whom all correspondence should be addressed.

e-mail address: diudea@chem.ubbcluj.ro

Abstract

The novel Cluj property indices are used for modeling the biological properties of dipeptides: the ACE inhibition activity of a set of 58 dipeptides and the bitter tasting activity of a set of 48 dipeptides, taken from the literature. The results are compared to those reported in some previous works.

Keywords: molecular graph; topological index; peptide

Quantitative Structure-Activity Relationships (QSARs) establish a mathematical relation between the biological activity of chemical compounds and their molecular structure. They provide quantitative models aimed to accurately predict a certain activity from the structural attributes. This topic has become a well-delimited branch in chemistry and was favored by the progress in computer science.

Many biological and physico-chemical properties can be correlated with the topostructural and topochemical features. Topological indices, TIs, (encoding information regarding the size, shape, branching or centricity) are among the simplest and efficient descriptors of molecular structure. Biological properties, more than physical or chemical properties, depend on the three-dimensional (3D) arrangements of the atoms in a molecule.

Coding of 3D structural information can be achieved in many different ways but all procedures make use of the Euclidean distances, irrespective they are considered explicitly or are involved in some more cryptic (3D sensitive) molecular descriptors. Since the actual conformation of a biological receptor is known only in fortunate cases, the researcher has to choose between the use of a single conformation (e.g. the minimum energy isomer) of the actives and of rotamer-library conformations. Excepting cases of very specific receptors (where more elaborated 3D grid descriptors are strongly needed for drawing virtual receptor maps - see the CoMFA-like procedures) [1-4], biological activity can be satisfactorily described by using a single conformation of actives and, moreover, by using 2D topological descriptors. Indeed, topological indices encode some information regarding the spatiality of molecules (e.g. as topological distance). Such a description (that is invariant to rotation and torsion) can sometimes be assimilated to the so-called "extended" conformation [5].

As the computer technology developed, novel descriptors, with enhanced ability in coding the 3D structural information and modeling a certain molecular property /activity could be designed. Large pools of descriptors were thus created.

In this paper a new approach, leading to a *fragmental property index family, FPIF*, is presented. These indices are calculated as local descriptors of some fragments of the molecule and, a global index is then obtained by summing the fragmental contributions. The modeling ability of *FPIF* is demonstrated on two sets of dipeptides, taken from literature [5-8].

CLUJ AND SZEGED THEORETICAL DESCRIPTORS

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

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

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

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

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

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

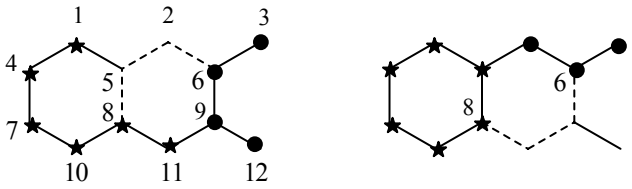
In defining *Cluj indices*, the *path* ***p*** plays the central role in selecting the subgraphs (eqs 1 and 2), particularly in cycle-containing graphs, where more than one path could join the pair (***i,j***). In such graphs, more than one subgraph (i.e. fragment), referred to ***i***, can be counted. 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

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

where $\boldsymbol{V_{i,j,p}}$ is either $\boldsymbol{CJ_{i,j,p}}$ or $\boldsymbol{CF_{i,j,p}}$ and consists of vertices ***v*** lying *closer* to the vertex ***i*** than to the vertex ***j***. When $\boldsymbol{p} \in \boldsymbol{D}(\boldsymbol{G})$, (i.e. the set of all topological distances, or geodesics i n ***G***) then **M** = **CJD** (Cluj-Distance) or **CFD** (Cluj-Fragmental-Distance**). W**hen $\boldsymbol{p} \in \boldsymbol{\Delta}(\boldsymbol{G})$, (i.e. the set of all topological detours, or the longest distances in ***G***) **M = CJΔ** (Cluj-Detour) or **CFΔ** (Cluj-Fragmental-Detour). The diagonal entries are zero. The Cluj matrices are square arrays, of

dimension NxN**,** usually *unsymmetric* (excepting some symmetric regular graphs). Figures 1 and 2 illustrate the construction of matrices **CJD** and **CJΔ** respectively.

Figure 1. Construction of Cluj Distance Matrix, UCJD



a b

Cluj Distance Sets $CJD_{i,j,p}$; pair (6,8):

(a) (6, 8) [6, 2, 5, 8] { 3, 6, 9, 12 }
(b) (6, 8) [6, 9, 11, 8] { 2, 3, 6 }

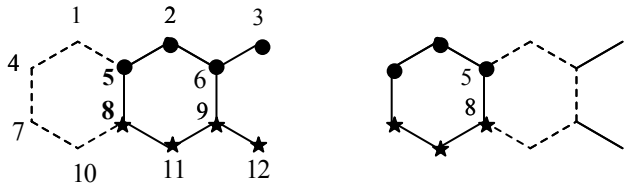
(a) (8, 6) [8, 5, 2, 6] { 1, 4, 7, 8, 10, 11 }
(b) (8, 6) [8, 11, 9, 6] { 1, 4, 5, 7, 8, 10 }

Cluj Distance Matrix **UCJD**

0	3	5	9	3	5	5	2	3	5	3	5
5	0	7	7	5	7	5	3	4	5	4	7
1	1	0	1	1	1	1	1	1	1	1	1
3	3	4	0	2	3	6	2	3	2	2	3
9	7	7	8	0	6	8	6	6	5	4	6
5	5	11	5	4	0	5	***4***	6	4	3	6
2	2	3	6	2	3	0	2	3	3	3	4
5	4	6	8	6	***6***	8	0	6	9	7	7
4	3	6	5	4	6	5	4	0	5	5	11
5	3	5	5	2	3	9	3	5	0	3	5
5	4	7	5	3	4	7	5	7	5	0	7
1	1	1	1	1	1	1	1	1	1	1	0

$IP2(CJD) = 1024$
$IE2(CJD) = 378$

Figure 2. Construction of Cluj Detour matrix, UCJΔ



a b

Cluj Detour Sets $CJ\Delta_{i,j,p}$; pair (5,8):

(a) (5, 8) [5, 1, 4, 7, 10, 8] { 2, 3, 5, 6 }

(b) (5, 8) [5, 2, 6, 9, 11, 8] { 1, 4, 5 }

(a) (8, 5) [8, 10, 7, 4, 1, 5] { 8, 9, 11, 12 }

(b) (8, 5) [8, 11, 9, 6, 2, 5] { 7, 8, 10 }

Cluj Detour Matrix **UCJΔ**

0	1	2	1	1	2	1	2	1	2	1	1
1	0	1	2	1	1	1	3	1	1	3	3
1	1	0	1	1	1	1	1	1	1	1	1
1	3	3	0	1	3	1	2	2	1	2	2
1	1	2	1	0	1	2	***4***	2	2	2	2
5	2	11	5	2	0	3	4	2	3	2	2
1	2	2	1	2	2	0	1	3	1	3	3
2	2	2	2	***4***	2	1	0	1	1	1	2
3	2	2	3	4	2	5	2	0	5	2	11
2	1	1	1	2	1	1	1	2	0	1	2
1	3	3	1	3	1	2	1	1	1	0	1
1	1	1	1	1	1	1	1	1	1	1	0

$IP2(CJ\Delta) = 247$
$IE2(CJ\Delta) = 53$

The entries in the unsymmetric Szeged distance matrix, **USZD,** 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 $\boldsymbol{d}(\boldsymbol{G})_{v,i} < \boldsymbol{d}(\boldsymbol{G})_{v,j}$ (eq 3) is changed by the *detour criterion* $\boldsymbol{\delta}(\mathbf{G})_{v,i} < \boldsymbol{\delta}(\mathbf{G})_{v,j}$, the cardinality of the sets thus supplied represent entries in the unsymmetric Szeged detour matrix, **USZ**$\boldsymbol{\Delta}$.

The above definitions hold for any connected graph

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

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

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

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, **CJD**, **CJ**$\boldsymbol{\Delta}$, **CF**$\boldsymbol{\Delta}$ and **CF**$\boldsymbol{\Delta}$, are identical, due to the uniqueness of the path joining a pair of vertices (***i***, ***j***). Some special properties of Cluj matrices in trees were exposed elsewhere [9, 10, 15-17].

In any graph, $\mathbf{CJD}_e = \mathbf{CFD}_e = \mathbf{SZD}_e$. In cyclic graphs, $\mathbf{CJD}_p \neq \mathbf{CFD}_p \neq \mathbf{SZD}_p$, $\mathbf{CJ}\Delta_p \neq \mathbf{CF}\Delta_p \neq \mathbf{SZ}\Delta_p$.

The above-discussed matrices allow the calculation of integer value indices by relations given for the fragmental property indices (see below) [14].

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:

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

For the potential of type

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \boldsymbol{pz} \tag{8}$$

the associated field can be derived as

$$\begin{aligned} \vec{\nabla} \cdot \boldsymbol{f} &= \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} \vec{\boldsymbol{i}} + \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}} \vec{\boldsymbol{j}} + \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{z}} \vec{\boldsymbol{k}} = \frac{\partial (\boldsymbol{pz})}{\partial \boldsymbol{x}} \vec{\boldsymbol{i}} + \frac{\partial (\boldsymbol{pz})}{\partial \boldsymbol{y}} \vec{\boldsymbol{j}} + \frac{\partial (\boldsymbol{pz})}{\partial \boldsymbol{z}} \vec{\boldsymbol{k}} = \\ &= 0\vec{\boldsymbol{i}} + 0\vec{\boldsymbol{j}} + \boldsymbol{p}\vec{\boldsymbol{k}} = \boldsymbol{p}\vec{\boldsymbol{k}} = \vec{\boldsymbol{p}} \end{aligned} \tag{9}$$

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

$$\vec{\boldsymbol{G}} = \boldsymbol{m}\vec{\boldsymbol{g}} \tag{10}$$

with the corresponding potential given by

$$\boldsymbol{E}_{\boldsymbol{p}} = \boldsymbol{E}_{\boldsymbol{p}}(z) = \boldsymbol{mgz} \tag{11}$$

where $\boldsymbol{m}$ is the mass of the probe and $\boldsymbol{z}$ 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 $\boldsymbol{m}$ (or the charge $\boldsymbol{q}$) that generates the potential $\boldsymbol{f}$ and associated field $\vec{\nabla} \cdot \boldsymbol{f}$ is far enough ($\boldsymbol{r} >> \boldsymbol{z}$) so that the approximation $(\boldsymbol{r} + \boldsymbol{z})^2/\boldsymbol{r}^2 = (\boldsymbol{r}^2 + 2\boldsymbol{rz} + \boldsymbol{z}^2)/\boldsymbol{r}^2 = 1 + 2\boldsymbol{z}/\boldsymbol{r} + (\boldsymbol{z}/\boldsymbol{r})^2 \cong 1$ is valid.

For the potential of type:

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \boldsymbol{p}/\boldsymbol{z} \tag{12}$$

eq 7 leads to the associated field:

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

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

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

and the associated potential of the form:

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

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:

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

and the potential associated to the Coulombian field:

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

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* Φ (mass M, electronegativity E, cardinality C, partial charge or any other atomic property P), *property descriptor* Ω (p, d, pd, $1/p$, $1/d$, p/d, $p/d2$, $p2/d2$) and *superposition* Ψ (S, P, A, G, H).

The expressions for the *property descriptors* are:

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

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

The (mathematical) superposition Ψ, given by

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

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.) [18]. The multiplication occurs in concurrent phenomena (probabilistically governed) [19-21]. Other operators find appropriate justification [22-26].

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* (see the *CoMFA* approach) [1-4]. However, the chemical identity of ***j*** is not considered.

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):

$$\boldsymbol{PD}(\boldsymbol{Fr}_{i,j}) = \left\| \sum_{v \in \boldsymbol{Fr}_{i,j}} \vec{\Omega}\left(\boldsymbol{d}_{E\,v,j}, \boldsymbol{p}_v\right) \right\|; \quad \vec{\Omega} = \Omega \cdot \frac{\vec{\boldsymbol{d}}_{E\,v,j}}{\boldsymbol{d}_{E\,v,j}}; \quad \boldsymbol{P}(\boldsymbol{Fr}_{i,j}) = \Psi_{v \in \boldsymbol{Fr}_{i,j}}(\boldsymbol{p}_v);$$

$$\boldsymbol{d}_{E\,CP,j} = \Omega_{\boldsymbol{p}}^{-1}(\boldsymbol{DG}(\boldsymbol{Fr}_{i,j}), \boldsymbol{P}(\boldsymbol{Fr}_{i,j})), \tag{22}$$

where $\boldsymbol{d}_{E\,CP,j}$ is the distance that satisfies: $\Omega(\boldsymbol{d}_{E\,CP,j}, \boldsymbol{P}(\boldsymbol{Fr}_{i,j}) = \boldsymbol{PD}(\boldsymbol{Fr}_{i,j})$

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.):

$$\boldsymbol{PD}(\boldsymbol{Fr}_{i,j}) = \Omega(\boldsymbol{d}_{E\,CP,j}, \Psi_{v \in \boldsymbol{Fr}_{i,j}}(\boldsymbol{p}_v)); \quad \boldsymbol{CP}_i(\boldsymbol{x}_{\boldsymbol{CP}_i,j}, \boldsymbol{y}_{\boldsymbol{CP}_i,j}, \boldsymbol{z}_{\boldsymbol{CP}_i,j},); \tag{23}$$

$$\boldsymbol{x}_{\boldsymbol{CP}_i,j} = \sum_{v \in \boldsymbol{Fr}_{i,j}} \boldsymbol{x}_v \cdot \boldsymbol{p}_v \Big/ \sum_{v \in \boldsymbol{Fr}_{i,j}} \boldsymbol{p}_v; \quad \boldsymbol{y}_{\boldsymbol{CP}_i,j} = \sum_{v \in \boldsymbol{Fr}_{i,j}} \boldsymbol{y}_v \cdot \boldsymbol{p}_v \Big/ \sum_{v \in \boldsymbol{Fr}_{i,j}} \boldsymbol{p}_v;$$

$$\boldsymbol{z}_{\boldsymbol{CP}_i,j} = \sum_{v \in \boldsymbol{Fr}_{i,j}} \boldsymbol{z}_v \cdot \boldsymbol{p}_v \Big/ \sum_{v \in \boldsymbol{Fr}_{i,j}} \boldsymbol{p}_v$$

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

Fragmental Property Matrices

The fragmental property matrices are square matrices of the 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 $\boldsymbol{i}, \boldsymbol{j} \in \boldsymbol{V}(\boldsymbol{G})$, $\boldsymbol{i} \neq \boldsymbol{j}$ and $\boldsymbol{P}_{i,j} = \{ \boldsymbol{p}^1_{i,j}, \boldsymbol{p}^2_{i,j}, ..., \boldsymbol{p}^k_{i,j} \}$ paths joining ***i*** and ***j***, then cf. ***CJ*** or ***CF*** definition (eqs 1-3), the fragments $\boldsymbol{Fr}^1_{i,j}, \boldsymbol{Fr}^2_{i,j}, ..., \boldsymbol{Fr}^k_{i,j}$ are generated. Let ***m*** be the number of

maximal fragments among all the k fragments, $1 \le m \le k$, 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 and, consequently, the matrix entry associated to the pair (i,j) is the mean value, e.g.

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

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:

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

$$E\_(M) = \tfrac{1}{2}\Sigma\Sigma\,[\mathbf{M}]_{i,j}[\mathbf{A}]_{i,j} \quad ; \quad E2(M) = \tfrac{1}{2}\Sigma\Sigma\,[\mathbf{M}]_{i,j}[\mathbf{M}]_{j,i}[\mathbf{A}]_{i,j} \tag{25}$$

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:

$$ABcDdEfffffG \tag{26}$$

where:

$A \in \{D, R\}$; D = Dense; R = Rare;

$B \in \{T, G\}$; T = Topological; G = Geometric;

$c \in \{f, j, s\}$; f = CF-type; j = CJ-type; s = Sz-type;

$Dd \in \{Di, De\}$; Di = Distance; De = Detour;

$E \in \Phi$ (i.e. $\mathbf{E} \in \{\mathbf{M}, \mathbf{E}, \mathbf{C}, \mathbf{P}\}$ where M = *mass*; E = *electronegativity*; C = *cardinality*; P = other atomic property - implicitly, *partial charge*; explicitly, a property given by manual input);

$fffff \in \Omega$ (i.e. $fffff \in \{\_\_p\_\_, \_1/p\_, \_\_d\_\_, \_1/d\_, \_p.d\_, \_p/d\_, \_p/d2, p2/d2\}$

$G \in \Psi$ (i.e. $G \in \{S, P, A, G, H\}$ with the known meaning (see above).

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

$$ABcDdEfffffGii \tag{27}$$

where:

$ii \in \{P\_, P2, E\_, E2\}$ 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:

$$lnABcDdEfffffGii := \ln(ABcDdEfffffGii);$$

$$1/ABcDdEfffffGii := \frac{1}{ABcDdEfffffGii} \tag{28}$$

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*** 16-bit windows computer programs.

CORRELATING STUDIES

The mathematical models of a certain property can be achieved by MLR (Multiple Linear Regression), CNN (Computational Neural Networks) [27-31], or other mathematical procedures. In the following, the MLR procedure is presented.

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

$$Y_i = b_0 + \sum_{j=1}^{m} b_{ij} X_{ij} \tag{29}$$

or in matrix form as

$$\mathbf{Y} = \mathbf{bX} \tag{30}$$

where **Y** is the ***n*** x 1 vector of responses, **X** is an ***n*** x (***m*** + 1) matrix of independent variables and **b** is the (***m*** + 1) x 1 vector of regression coefficients. The regression coefficients can be determined by the least-squares solution of (30)

$$\mathbf{b} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} \qquad (31)$$

With **b** calculated, eq 30 can be used for estimating the chosen property for other chemical structures.

We tested the correlating ability of ***FPIF*** on two sets:

*Dipeptide ACE Inhibitors***.** The set consists of 58 dipeptides and was taken from Cocchi's report [6]. 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 I. 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 (eq 32) and the corresponding residuals. As above mentioned, ***FPIF*** descriptors take explicitly into account 3D-structural features of the whole molecule of dipeptides.

Table I.

No.	Peptide	***BA***obs	***BA***calc(eq 32)	Residuals	Peptide	***BA***obs	***BA***calc(eq 33)	Residuals
1	YG	2.7	2.859794	-0.15979	YL	2.4	2.282458	0.117542
2	YA	3.34	3.411811	-0.07181	WW	3.6	3.627912	-0.02791
3	WG	2.23	3.297515	-1.06751	WE	1.56	1.768665	-0.20867
4	VY	4.66	4.519753	0.140247	VV	1.71	1.93277	-0.22277
5	VW	5.8	5.149804	0.650196	VL	2	2.117623	-0.11762
6	VP	3.38	3.828656	-0.44866	VG	1.19	1.384127	-0.19413
7	VG	2.96	2.652658	0.307342	VA	1.16	1.534544	-0.37454
8	VF	4.28	4.366133	-0.08613	SL	1.49	1.412394	0.077606
9	TG	2	2.426741	-0.42674	PY	1.8	2.199467	-0.39947
10	SG	2.07	2.06538	0.00462	PL	2.22	2.011651	0.208349
11	RW	4.8	4.77005	0.02995	PI	2.33	1.956997	0.373003
12	RP	3.74	3.427529	0.312471	PF	2.8	2.595868	0.204132
13	RF	3.64	3.966233	-0.32623	PA	1.32	1.413292	-0.09329
14	RA	3.34	2.663339	0.676661	LY	2.46	2.358832	0.101168
15	QG	2.13	2.023636	0.106364	LW	3.4	3.094146	0.305854
16	PG	1.77	2.208404	-0.4384	LL	2.35	2.151635	0.198365
17	MG	2.32	2.708227	-0.38823	LG	1.72	1.507413	0.212587
18	LG	2.06	2.572545	-0.51254	LF	2.75	2.874128	-0.12413
19	LA	3.51	3.052476	0.457524	LA	1.72	1.677078	0.042922
20	KG	2.49	2.208121	0.281879	IW	3.05	2.885932	0.164068
21	KA	3.42	2.663597	0.756403	IV	2.05	1.992664	0.057336
22	IY	5.43	4.720289	0.709711	IT	1.49	1.49369	-0.00369
23	IW	5.7	5.366469	0.333531	IS	1.49	1.425665	0.064335

24	IP	3.89	4.009775	-0.11977	IQ	1.49	1.711316	-0.22132
25	IG	2.92	2.74235	0.17765	IP	2.4	1.764216	0.635784
26	IF	3.03	4.551403	-1.5214	IN	1.49	1.633757	-0.14376
27	HL	2.49	3.389366	-0.89937	IL	2.26	2.279924	-0.01992
28	HG	2.2	1.729607	0.470393	IK	1.65	2.084282	-0.43428
29	GY	3.68	3.292568	0.387432	II	2.26	2.258983	0.001017
30	GW	4.52	3.801399	0.718601	IG	1.68	1.486461	0.193539
31	GV	2.34	2.875864	-0.53586	IE	1.37	1.394035	-0.02403
32	GT	2.24	2.675964	-0.43596	ID	1.37	1.380127	-0.01013
33	GS	2.42	2.227733	0.192267	IA	1.68	1.656997	0.023003
34	GR	2.49	2.439863	0.050137	GY	1.77	1.563785	0.206215
35	GQ	2.15	2.321493	-0.17149	GW	1.89	2.010001	-0.12
36	GP	3.35	2.755225	0.594775	GV	1.13	1.345441	-0.21544
37	GM	2.85	2.947157	-0.09716	GP	1.35	1.14037	0.20963
38	GL	2.6	2.832248	-0.23225	GL	1.68	1.493838	0.186162
39	GK	2.27	2.468511	-0.19851	GI	1.7	1.479963	0.220037
40	GI	2.92	3.019004	-0.099	GF	1.8	1.950026	-0.15003
41	GH	2.51	2.134515	0.375485	FY	3.13	2.995819	0.134181
42	GG	2.14	2.250075	-0.11007	FP	2.7	2.137678	0.562322
43	GF	3.2	3.253063	-0.05306	FL	2.87	2.865392	0.004608
44	GE	2.27	2.174955	0.095045	FG	1.77	1.892504	-0.1225
45	GD	2.04	2.200932	-0.16093	FF	3.1	3.426157	-0.32616
46	GA	2.7	2.430725	0.269275	AV	1.16	1.569456	-0.40946
47	FR	3.04	3.868559	-0.82856	AL	1.7	1.741811	-0.04181
48	FG	2.43	2.867713	-0.43771	AF	1.72	2.219069	-0.49907
49	EG	2	1.865655	0.134345				
50	EA	2	2.427307	-0.42731				
51	DG	1.85	1.863917	-0.01392				
52	DA	2.42	2.367776	0.052224				
53	AY	4.06	3.819573	0.240427				
54	AW	5	4.381854	0.618146				
55	AP	3.64	3.186382	0.453618				
56	AG	2.6	2.410909	0.189091				
57	AF	3.72	3.732496	-0.0125				
58	AA	3.21	2.716249	0.493751				

TABLE II. Statistics for ACE inhibitors set.

1	2	3	4	5	6	7
Index	DTfDiM_p/d2GP_	lnDGsDiE_1/p_GE_ DTfDiM_p/d2GP_	DTjDeM_p/d2GP_	lnDTjDeEp2/d2AE_ DTjDeM_p/d2GP_	DTsDeP_1/d_GP2	lnRGsDeMp2/d2AE_ DTsDeP_1/d_GP2
r	0.78192	**0.88696**	0.78843	0.87536	0.79228	0.87171
r^2	0.61140	**0.78670**	0.62162	0.76626	0.62770	0.75988
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
	Cross-validated					
		L20%o (aver.)[a]		L20%o		L20%o
r		**0.87148**		0.85920		0.85522
r^2		**0.75948**		0.73822		0.73140
s		0.495		0.517		0.524

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

Table II 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.

The best-found model was:

$$\boldsymbol{BA}_{calc} = 35.992 + 0.822 * \boldsymbol{DTfDiM\_p/d2GP\_} - 11.802 * \ln\boldsymbol{DGsDiE\_1/p\_GE\_}$$
$$\boldsymbol{n} = 58;\ \boldsymbol{r} = 0.88696;\ \boldsymbol{s} = 0.471;\ \boldsymbol{F} = 101.426 \qquad (32)$$

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 II) 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*** (eq 33) is presented in Figure 3.

The model given by eq 32 is superior, both in estimation and prediction, to those reported in literature (see Table III). 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). The correlation recorded in case of extended conformation surpasses the correlation obtained in rotameric case.

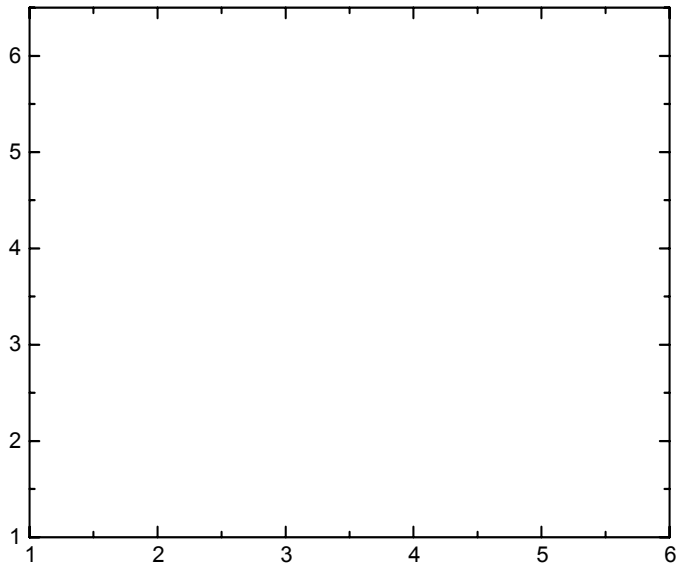


Figure 3. The plot of observed vs calculated BA (eq 32).

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

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

9	Sweeteners (Zalini et al. - rotameric) [5]	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.

TABLE IV. 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_	0.88696
2	54	1771	DTjDeM_p/d2GP_	lnDTjDeEp2/d2AE_	0.87536
3	29	7894	DTsDeP_1/d_GP2	lnRGsDeMp2/d2AE_	0.87171
4	29	2644	DTsDeP_1/d_GP2	lnRTjDeEp2/d2HE_	0.86856
5	18	8213	DTsDeM_p/d_GP2	lnRGsDeEp2/d2AE_	0.86812
6	18	7725	DTsDeM_p/d_GP2	lnRGsDeE_p/d2AE_	0.86237
7	15	6476	DTfDeM_p/d_PP_	lnRTsDiEp2/d2AE2	0.86181
8	1	15876	RTfDeE_1/p_PP2	lnRGsDeCp2/d2HP2	0.86136
9	1	8719	RTfDeE_1/p_PP2	DGfDiP_p/d_GP_	0.85178
10	1	6485	RTfDeE_1/p_PP2	lnRTsDiEp2/d2GE2	0.84654

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 [5].

Table IV 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). This result correlates with the Zaliani's best result when used extended conformations (see Table III).

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.

2. Dipeptide Sweeteners. The set including 48 dipeptides was taken from Jonsson's paper [7]. The molecular structures were input and optimized by using MM+ and then by semiempirical AM_1 procedure of the HyperChem Program (HyperCube Inc.). Table I includes the dipeptide names by using the one-letter code for aminoacids, the observed bitter tasting activity (biological activity, BA, as log(1/T)), the calculated BA (according to eq 33) and the corresponding residuals.

Table V 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:

$$\boldsymbol{BA}_{\text{calc}} = 1.142 + 0.474^{*}\ \boldsymbol{RTsDiM}\_1/\boldsymbol{p\_SP}\_ - 0.043\ ^{*}\ \boldsymbol{DGsDiE}\_1/\boldsymbol{p\_AP}\_$$
$$\boldsymbol{n} = 48;\ \boldsymbol{r} = 0.92272;\ \boldsymbol{s} = 0.248;\ \boldsymbol{F} = 128.922 \qquad (33)$$

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 V, 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 33) is presented in Figure 4.

The model given by eq 33 surpasses those reported in literature (see Table III).

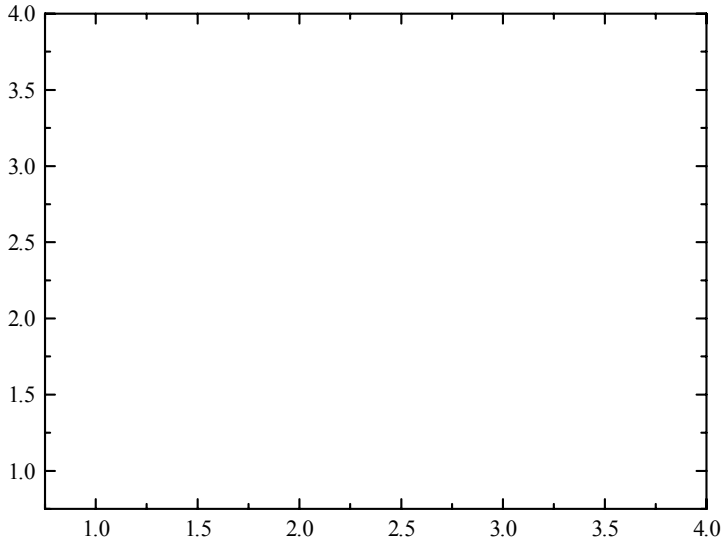


Figure 4. The plot of observed vs calculated BA (eq 33).

TABLE V. Statistics for sweeteners dipeptides.

1	2	3	4	5	6	7
Index	RTsDiE_1/p_AE2	RTsDiEp2/d2AE_ RTsDiE_1/p_AE2	RTsDeE_1/p_AE_	lnRTsDiEp2/d2GE2 RTsDeE_1/p_AE_	RTsDiM_1/p_SP_	DGsDiE_1/p_AP_ RTsDiM_1/p_SP_
r	0.81448	0.91688	0.80079	0.91525	0.77192	**0.92272**
r^2	0.66337	0.84067	0.64126	0.83768	0.59586	**0.85141**
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		0.90670		0.90473		**0.91286**
r^2		0.82210		0.81854		**0.83331**
s		0.268		0.271		0.259

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

TABLE VI. 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_	0.92272
2	6076	132	RTsDiEp2/d2AP2	DGsDeEp2/d2GE_	0.92094
3	33	6051	RTsDeE_1/p_AE_	RTsDiEp2/d2GE2	0.91525
4	1	3180	RTsDiE_1/p_AE2	RTsDiEp2/d2AE_	0.91688
5	1	3154	RTsDiE_1/p_AE2	RTfDiE_p/d2AP2	0.90759
6	1	3093	RTsDiE_1/p_AE2	RTsDiEp2/d2AP_	0.90269
7	1	3074	RTsDiE_1/p_AE2	RTjDiE_p/d_GP2	0.88456
8	1	3012	RTsDiE_1/p_AE2	RTsDeE_p/d_GP2	0.88456
9	1	2769	RTsDiE_1/p_AE2	lnDGsDeE_1/p_PP2	0.87676
10	1	2076	RTsDiE_1/p_AE2	DTsDiEp2/d2SE	0.87551

Table VI 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 (see Table III).

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.

Concluding Remarks

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.

There exist an analogy between ***CoMFA*** and ***FPIF***: both of them calculate the interaction of a chemical structure (or substructure) with a *probe atom* in the 3D space. The fragmental property $\boldsymbol{Fr_{i,j}}$ is viewed as an interaction of atoms forming the fragment $\boldsymbol{Fr_{i,j}}$ with the atom $\boldsymbol{j}$. The major difference is that in ***CoMFA*** the probe atoms (with well defined chemical identity) is

external whereas ***FPIF*** considers internal probe atoms with no chemical identity. Only the fragments (i.e. substructures) are chemically well defined.

FPIF offer good description and modeling of dipeptides activity, such as the ACE inhibition or bitter tasting. As it is known, a correlational model does not involve a causal relationship between descriptors and a molecular property. However, a look upon the occurrence of indices with some best scores (and implicitly best structure description) can highlight some aspects of intra- and/or intermolecular interactions.

The above results demonstrate the usefulness of our descriptors in modeling peptide structures and properties. For other ***FPIF*** modeling examples the reader can consult [14].

Acknowledgement

This work in under financial support of GRANT CNSIS, T 34, 2000. The authors are grateful to the referees for their valuable comments.

References

[1] Cramer, R. D., Patterson, D. E. and Bunce, J. D. (1988). Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc*., **110**, 5959-5967.

[2] Clark, M., Cramer III, R. D., Jones, D. M., Patterson, D. E. and Simeroth, P.E. (1990). Comparative Molecular Field Analysis (CoMFA). 2. Toward Its Use with 3D-Structural Databases. *Tetrahedron Comput. Methodol*., **3**, 47-59.

[3] Kubinyi. (1997). HQSAR and 3B QSAR in Drug Design. Part 2: Application and Problems. *Drug Design Today*., **2**, 538-546.

[4] Ivanciuc, I. (2000). 3D QSAR Models, in *QSAR/QSPR Studies by Molecular Descriptors*, M. V. Diudea , Ed., NOVA SCIENCE Huntington, New York, U.S.A., (to appear).

[5] Zaliani, A. and Gancia, E. J. (1999). MS-WHIM Scores for amino-acids: a new 3D-description for peptide QSAR and QSPR Studies. *Chem. Inf. Comput. Sci*., **39**, 525-533.

[6] Cocchi, M. and Johansson, E. (1993). Amino Acids Characterization by GRID and Multivariate Data Analysis. *Quant. Struct.-Act. Relat*., **12**, 1-8.

[7] Jonsson, J., Eriksson, L., Hellberg, S., Sjostroem, M. and Wold, S. (1989). Multivariate parametrization of 55 coded and non-coded amino acids. *Quant. Struct.-Act. Relat*., **8**, 203-209.

[8] Collantes, E. R. and Dunn III, W. J. (1995). Amino acids side chain descriptors for Quantitative Structure-Activity relationship studies of peptide analogues. *J. Med. Chem*., **38**, 2705-2713.

[9] Diudea, M. V. (1997). Cluj matrix invariants. *J.Chem.Inf.Comput.Sci.*, **37**, 300-305.

[10] Diudea, M. V. (1997). Cluj Matrix $\mathbf{CJ_u}$: source of various graph descriptors, *Commun. Math. Comput. Chem. (MATCH),* **35**, 169-183.

[11] Diudea, M., Minailiuc V. O., Katona G. and Gutman I. (1997). Szeged Matrices and Related Numbers. *Commun. Math. Comput. Chem. (MATCH),* **35**, 129-143.

[12] Diudea, M. V., Pârv, B. and Topan, M. I. (1997). Derived Szeged and Cluj Indices,

J. Serb.Chem.Soc., **62**, 267-276.
[13] Minailiuc O., Katona G., Diudea, M. V., Strunje, M., Graovac, A. and Gutman, I. (1998). Szeged Fragmental Indices. *Croat. Chem. Acta.,* **71**, 473-488.
[14] Jäntschi, L., Katona, G. and Diudea, M. V. (2000). Modeling Molecular Properties by Cluj Indices. *Commun. Math. Comput. Chem. (MATCH),* **41**, 151-188.
[15] Diudea, M. V. (1996). Wiener and Hyper-Wiener Numbers in a Single Matrix, *J.Chem.Inf.Comput.Sci.*, **36**, 833-836.
[16] Diudea, M.V., Parv, B. and Gutman, I. (1997). Detour-Cluj Matrix and Derived Invariants, *J. Chem. Inf. Comput. Sci.,* **37**, 1101-1108.
[17] Diudea, M. V. (1999). Valencies of Property, *Croat. Chem. Acta.,* **72**, 835-851.
[18] Golender, V., Vesterman, B. and Vorpagel, E. (1996). APEX-3D Expert System for Drug Design. *Network Science*, *Jan*, http:\\www.netsci.org/Science/Compchem/feature09.html.
[19] Landau, L. D. and Lifshitz, E. M. (1978). Statistical Physics, 3-rd edition, revised by Lifshitz, E. M and Pitaevkim, L. P., Nauka, Moscow, Chap. Phase Transition of Rank Two and Critical Phenomena.
[20] Rose, V. S. and Wood, J. (1998). Generalized Cluster Significance Analysis and Stepwise Cluster Significance Analysis with Conditional Probabilities. *Quant. Struct.-Act. Relat.*, **17**, 348-356.
[21] Labute, P. (1998). QuaSAR-Binary: A New Method for the Analysis of High Thoughput Screening Data. *Network Science*, *May*, http://www.netsci.org/Science/Compchem/feature21.html.
[22] Young, H. D. (1962). *Statistical Treatment of Experimental Data*. McGraw-Hill, New York.
[23] Reif, F. (1965). *Fundamentals of Statisticals and Thermal Physics*, McGraw-Hill, Chap.1 New York.
[24] Crawford, F. S. Jr. (1968). *Waves*, Berkeley Physics Course, Newton, Massachusetts, vol. 3.
[25] Purcell, E. M. (1965). *Waves*, Berkeley Physics Course, Newton, Massachusetts, vol. 3.
[26] Atkins, P. W. (1994). *Physical Chemistry*, fifth edition, Oxford University Press (Oxford, Melbourne, Tokyo).
[27] Zupan, J. and Gasteiger, J. (1993). *Neural Networks for Chemists*; VCH: Weinheim.
[28] Bulsari, A.B. (1995). *Neural Networks for Chemical Engineers*; Elsevier: Amsterdam.
[29] Devillers, J. (1996). *Neural Networks in QSAR and Drug Design*. Academic Press, London, p. 279.
[30] Ivanciuc, O. (1998). Artificial Neural Networks Applications. Part 5. Prediction of the Solubility of C_{60} in Organic Solvents. *Rev. Roum. Chim*., **43**, 775-783.
[31] Ivanciuc, O. (1995). Artificial Neural Networks Applications. Part 6. Use of Non-Bonded van der Waals and Electrostatic Intramolecular Energies in the Estimation of ^{13}C-NMR Chemical Shifts in Saturated Hydrocarbons. *Rev. Roum. Chim*., **40**, 1093-1101.