Structure-Activity Relationships on the Molecular Descriptors Family Projects at the End

Lorentz JÄNTSCHI[1] and Sorana D. BOLBOACĂ[2]

[1]*Technical University of Cluj-Napoca, Romania*
[2]*"Iuliu Haţieganu" University of Medicine and Pharmacy Cluj-Napoca, Romania*
lori@academicdirect.org, sorana@j.academicdirect.ro

Abstract

Molecular Descriptors Family (MDF) on the Structure-Activity Relationships (SAR), a promising approach in investigation and quantification of the link between 2D and 3D structural information and the activity, and its potential in the analysis of the biological active compounds is summarized. The approach, attempts to correlate molecular descriptors family generated and calculated on a set of biological active compounds with their observed activity. The estimation as well as prediction abilities of the approach are presented. The obtained MDF SAR models can be used to predict the biological activity of unknown substrates in a series of compounds.

Keywords

Structure-Activity Relationship (SAR); Molecular Descriptors Family (MDF); Model Assessment.

Introduction

Structure-Activity Relationships (SARs), Structure-Property Relationships (SPRs) and Property-Activity Relationships (PARs) arise with the studies of Louis Plack HAMMETT in 1937 [1]. Since then, the Hammettôs equation found a lot of applications [2].

Quantitative relationships (QSAR, QSPR, QPAR) occurs when the property and/or activity are a quantitative one. Not all properties and activities of chemical compounds can be classified as being quantitative. Two interesting examples are LD_{50} (Median Lethal Dose, 50%)-dose necessary to kill half of the test population, and Sweetness (one of the five basic tastes, being almost universally related as a pleasure experience) of sugars, which can be appreciated only through comparison (relative scale), and we don't have two references and a scale (such as are boiling and freezing point and Celsius scale for temperature). Neither unanimous accepted as being quantitatively expressed properties does not have same accuracy degree expressed. From this reason in the last time are avoided to be used QSAR, QSPR, and QPAR, in their place being used (Q)SAR, (Q)SPR, and (Q)PAR, or more simple SAR, SPR, and PAR.

Moving the attention to the structure of compounds, the things are not so complicated. For example, an atom or a bound can exist and their existence can be identify through electronic transitions and/or molecular vibrations and/or rotations or can not (it is a problem of ñyesò or ñnoò). The things are a little bit complicated relative to the molecular geometry particularly in liquid or gas phases. Heisenberg principle presents the uncertainly rules at micro level (molecular and atomic level) [3]. Note that the molecular geometry depends on the environment on which molecule stays (vicinity of the molecule), temperature, pressure, and so on. From this point of view, dealing with molecular geometry is at least a matter of relativity if it is not a matter of uncertainty.

Thus, in Structure-Property-Activity Relationships (SPARs) approach we work with certainties (as molecular topology), uncertainties (as molecular geometry), relativities (as biological activities) and evidences (as physico-chemical properties).

The main goal of the researches was to develop an online system able to construct a family of structure based descriptors (called MDF-Molecular Descriptors Family), taking into consideration both geometrical and topological approaches without discrimination, in order to be used in a SAR procedure strengthened with a natural selection algorithm for obtaining best MDF-SAR (Molecular Descriptors Family (based) Structure Activity Relationship) model for given sets of compounds and given property or activity.

MDF Mathematical Model

A mathematical model composed from seven pieces has been developed. Each piece had a list of possibilities related with the physics approach. Every piece gives a letter in the descriptorôs name:

- Linearizing operator (1-st letter) make the link between micro, nano, and macro levels. Example: pH = -log[H+] itôs macro property (measure, effect) measured of micro environment (phenomena, cause), the presence and the number of H+ in a given solution. It takes six values.
- Molecular level superposing operator (2-nd letter) superposes fragmental contributions. Its existence is sustained by the variety of molecular property/activity causality, from specificity, regio-selectivity, and selectivity (most of biological activities) to structural formula independent (such as relative mass-same for all molecular formula isomers). It takes nineteen values.
- Pair-based fragmentation criteria (3-rd letter) implements different criterions. From first SAR studies of Hammet were observed that some parts of a molecule are more active and give the most of the activity/property of a molecule than others (substituentôs role). It takes four values.
- Interaction model (4-th letter) implements different levels of approximation (scalar and vector) for superposing of interaction descriptors at fragment level. Are well known that a series of field-type interactions (such as gravitational and electrostatic) are vectorial treated at low range and scalar treated at distance. It takes six values.
- Interaction descriptor (5-th letter) implements a series of interaction descriptors for physical entities (such as force, field, energy, potential), how are given in magnetism, electrostatics, gravity and quantum mechanics. It is a fact that different physical entities have different formulas. It takes twenty-four values.
- Atomic property (6-th letter) discriminates atoms one to each other through elemental properties. Every atom has a series of characteristics and/or properties making it similar and/or dissimilar to another. It takes six values.
- Distance operator (7-th letter) implements both 2D and 3D approaches (topology and geometry). It takes two values.

MDF Physical Model

Each characteristic of the mathematical model is a piece of the physical model. Table 1 presents all possibilities, associated significance and/or formula of MDF physical model. Constructing of descriptors family consists on calculation of 787968 (2 × 6 × 24 × 6 × 4 × 19 × 6) possibilities. Note that not all of them:

- Have a physical meaning (including here logarithm from a negative number, as example).
- Produce finite numbers (including here division by zero, as example).

Two types of degenerations can be observed in descriptors family: (1) a descriptors has the same values for all compounds from the set, and (2) two descriptors with different formula have the same value for all compounds from the set. When these kinds of descriptors are identified, a bias procedure is applied and the descriptors are discarding from the database. The average number of degenerated descriptors for a set of compounds is about 100000.

Table 1. Parameters values of MDF physical model

Nr	Encoding letter no	Parameter	Values
1	7-th (DO)	Distance operator:	Topological distance, `t` Geometrical distance, `g`
2	6-th (AP)	Atomic property:	Cardinality, `C` Count of directly bounded hydrogenôs, `H` Relative atomic mass, `M` Atomic electronegativity, `E` Group electronegativity, `G` Partial charge, `Q`
3	5-th (DIF)	Descriptor of interaction formula:	Distance, `D` = d Inverted distance, `d` = 1/d First atom's property, `O` = p1 Inverted O, `o` = 1/p1 Product of atomic properties, `P` = p1p2 Inverted P, `p` = 1/p1p2 Squared P, `Q` = ãp1p2 Inverted Q, `q` = 1/ãp1p2 First atom's Property multiplied by distance, `J` = p1d Inverted J, `j` = 1/p1d Product of atomic properties and distance, `K` = p1p2d Inverted K, `k` = 1/p1p2d Product of distance and squared atomic properties, `L` = dã(p1p2) Inverted L, `l` = 1/dãp1p2 First atom's property potential, `V` = p1/d First atom's property field, `E` = $p1/d^2$ First atom's property work, `W` = $p1^2/d$ Properties work, `w` = p1p2/d First atom's property force, `F` = p1^2/d^2 Properties force, `f` = p1p2/d^2 First atom's property weak nuclear force, `S` = $p1^2/d^3$ Properties weak nuclear force, `s` = p1p2/d^3 First atom's property strong nuclear force, `T` = $p1^2/d^4$ Properties strong nuclear force, `t` = p1p2/d^4

4	4-th (IM)	Interaction model:	SP(AP)= $\sum_{v \in Fragment} AP(v)$; CP(AP)=$\sum_{v \in Fragment} AP(v) \cdot DO(v,0)/SP(AP)$
		Rare model and resultant relative to fragment's head, `R`	DIF(SP(AP),AP(j),CP(AP))
		Rare model and resultant relative to conventional origin, `r`	DIF(SP(AP),AP(i),CP(AP))
		Medium model and resultant relative to fragment's head, `M`	$\sum_{v \in Fragment}$DIF(AP(v),AP(j),DO(v,j))
		Medium model and resultant relative to conventional origin, `m`	$\sum_{v \in Fragment}$DIF(AP(v),AP(j),DO(v,0))
		Dense model and resultant relative to fragment's head, `D`	$\sum_{v \in Fragment}$DIF(AP(v),AP(j),DO(v,j))×Versor(v,j)
		Dense model and resultant relative to conventional origin, `d`	$\sum_{v \in Fragment}$DIF(AP(v),AP(j),DO(v,0)) ×Versor(v,j)
5	3-th (FC)	Fragmentation criteria:	
		Minimal fragments, `m`	{i}
		Maximal fragments, `M`	{v| $d_{Gj}(v,i)<Ð$, Gj= G\{j}}
		Szeged distance based fragments, `D`	{v|d(v,i)<d(v,j)}
		Cluj path based fragments, `P`	{v| $d_{Gp}(v,i)<Ð$, Gp=G\p; p∈P(i,j)}
6	2-nd (MOSF)	Molecular overall superposing formula:	
		Conditional, smallest, `m`	Min(IM(f)| f-fragment, IM(f)<Ð)
		Conditional, highest, `M`	Max(IM(f)| f-fragment, IM(f)<Ð)
		Conditional, smallest absolute, `n`	Min(Abs(IM(f))| f-fragment, IM(f)<Ð)
		Conditional, highest absolute, `N`	Max(Abs(IM(f))| f-fragment, IM(f)<Ð)
		Averaged value, sum, `S`	$\sum_{f\|IM(f)<Ð}$IM(f)
		Averaged value, average, `A`	`S`/$\sum_{f|IM(f)<Ð}$1
		Averaged value, S/count(fragments), `a`	`S`/$\sum_{f}$1
		Aver. value, Avg.(Avg./atom)/count(atoms), `B`	`A`/$\sum_{v \in Mol}$1
		Averaged value, S/count(bonds), `b`	`S`/$\sum_{(u,v) \in Mol}$1
		Geometrical, product, `P`	$\prod_{f\|IM(f)<Ð}$IM(f)
		Geometrical, mean, `G`	$(`P`)^{1/\sum_{f\|IM(f)<Ð}1}$
		Geometrical, $P^{1/count(fragments)}$, `g`	$`S`^{1/\sum_{f}1}$
		Geometrical, Geom(Geom/atom)/count(atoms), `F`	`G`/$\sum_{v \in Mol}$1
		Geometrical, $P^{1/count(bonds)}$, `f`	$`S`^{1/\sum_{(u,v) \in Mol}1}$
		Harmonic, sum, `s`	$1/\sum_{0í f|IM(f)<Ð}$1/IM(f)
		Harmonic, mean, `H`	$\sum_{f|IM(f)<Ð}$1/`s`
		Harmonic, s/count(fragments), `h`	`s`/$\sum_{f}$1
		Harmonic, Harm.(Harm/atom)/count(atoms), `I`	`H`/$\sum_{v \in Mol}$1
		Harm., s/count(bonds), `i`	`H`/$\sum_{(u,v) \in Mol}$1
7	1-st (LO)	Linearization operator:	
		Identity (no change), `I`	f(x)=x
		Inversed I, `i`	f(x)=1/x
		Absolute I, `A`	f(x)=|x|
		Inversed A, `a`	f(x)=1/|x|
		Logarithm of A, `L`	f(x)=ln(x)
		Logarithm of I, `l`	f(x)=ln(abs(x))

MDF Methodology

MDF uses the data for a given set of molecules:

÷ Input:

- Molecular and/or structural formulas;
- Property/activity values;

÷ Output:
  - MDF of the set.

Following steps are applied:

÷ Draw (by hand) the topological model (2D) of every molecule from the set using HyperChem;
÷ Build (by software) the geometrical model (3D) of every molecule from the set using HyperChem;
÷ Apply (by software) a semiempirical model (for calculating the partial charge distribution on atoms) and (sometimes) a quantum mechanics model (going till most advanced ones such as Ab-initio and Time-Dependent Density Functional Theory) using specific modules of HyperChem (examples: HyperNewton, HyperGauss, HyperNDO) in order to obtain a optimized geometrical model in vitro or in vivo;
÷ Generate (using MDF Software) the MDF family;
÷ Apply the bias procedure;
÷ Obtain simple linear regression relationships between MDF members and given property/activity.

Multivariate MDF-SARs

Client-server applications for multivariate regressions using MDF members was build using Borland Delphi (v.6) and FreePascal (v.2). The applications use MySQL dynamic libraries to connect to MDF database. Following was subject of implementation:

÷ Systematic search (natural selection) in two independent variables (MDF members acting as independent variables);
÷ Systematic search in three independent variables (one being given by name as input data);
÷ Systematic search in four independent variables (two being given as input data);
÷ Systematic evolutionary search in N (N>2) variables (pair of two are natural selected based on input data from regression analysis in N-2 variables);
÷ Random search in N variables.

Note that a systematic search in three or more variables (with no input fixed variable) is too time and memory expensive (for three variables takes ~2Gb memory, ~120 days).

MDF-SAR Methodology

Followings act as input data in MDF-SAR approach:

÷ Topological (2D) and geometrical (3D) model of molecules from the set (HyperChem file);

÷ Values of the property/activity of a given set;

÷ Equation(s) with one or more MDF members;

÷ Estimated/predicted values of given property/activity with other SAR models (from specialty literature).

Following procedures were developed and used:

÷ Browse or Query MDF-SARs by sets. The application displays the obtained MDF-SARs models (including equation, determination coefficient, number of dependent variables, number of molecules in the set) for a selected set when the Browse mode are choused. When query mode are preferred, measured, estimated, and predicted (leave-one-out procedure) values are displayed, as well as cross-determinations between dependent variables are computed.

÷ Leave-one-out procedure (used as well in Query module) need independent variable values (measured property) and dependent variables values (structural descriptors) as input data for every molecule and produces (running inside Query module or independent) a column of predicted values (excluding one-by-one a molecule from the set, computing regression equation and using the regression equation for obtaining a prediction for the excluded molecule), and correlates the predicted values with measured property (cross-validated leave-one-out score).

÷ Training-versus-Test application has as input same measured and calculated values as leave-one-out procedure, and split the entire set in two sets (training and test) the number of molecules in training set being a user defined option. The split are made randomly. Using the molecules from training set, the SAR model is obtained. The SAR model is applied then on test set. Descriptive and inferential statistics are calculated on both training and test set.

÷ MDF-SAR Predictor is a featured application which allow to the user to select a learning set from the database (which contains a measured property on a molecules set). On the selected learning set, one or more MDF-SAR equations are proposed and the user must

chouse just one. Using the selected MDF-SAR equation, the user can submit a molecule in HIN format of which structural model were obtained using same level of approximation.

÷ Steigerôs Z test is used for comparison of two or more linear models, in order to see if one is significantly different from another. The procedure, known as correlated correlations, require the measured values, the estimated values by one model, and the estimated values by the another model, from which three correlation coefficients and sample size acts as input data for calculating Z distribution, from which the probability of identity are calculated.

MDF-SAR on Drug Design

This facility of MDF-SAR allows that having:

÷ A set of compounds of interest with known values of property/activity and an obtained, validated, and stored into the database MDF-SAR

÷ One of more similar/alike with selected set compound(s)

by made of:

÷ MDF-SAR equation

÷ Building of topological (2D) and geometrical (3D) through the same choices as were build on the selected set

to obtain

÷ Predicted value(s) for the property/activity of the new compounds, even if this (these) compound(s) were not yet synthesized, in order to see if the new structure (virtual compound at this time) comes or not with improvements in desired property/activity.

A summary of twenty-seven best performing models in terms of estimation and prediction are presented bellow. The information is summarized according with the investigated activity and compounds classes. The results are expressed as MDF-SAR equation accompanied by the sample size (n), correlation coefficient (r), associated 95% CI of correlation coefficient (95%CI_r), standard error of estimated (s_{est}), Fisher parameter (F_{est}) and its type I error of estimated (in round parentheses), prediction power expressed as cross-validation leave-one-out coefficient (r_{loo}) and its 95% CI (95%CI_{rloo}), standard error of predicted (s_{pred}), Fisher parameter (F_{pred}) and its type I error of predicted (in round

parentheses). The ř is the estimated activity by the MDF model, and *iMDRoQg* is for example the name of the molecular descriptors used by the model

1. Hydrophobic vs. hydrophilic character of standard amino acids

ř =-0.58 + 8.5·*iMDRoQg* [4]

n = 15 [5], r [95% CI_r] = 0.9514 [0.8565-0.9840], s_{est} = 0.44, F_{est} (p) = 124 ($5.05 \cdot 10^{-8}$), r_{loo} [95%CI_{rloo}] = 0.9351 [0.8028-0.9796], s_{pred} = 0.51, F_{pred} (p) = 90 ($3.26 \cdot 10^{-7}$).

2. Hydrophobic vs. hydrophilic character of standard amino acids

ř = 12-21·*IGDROQg* [4]

n = 15 [6], r [95% CIr] = 0.9759 [0.9270-0.9921], s_{est} = 0.71, F_{est} (p) = 260 ($5.66 \cdot 10^{-10}$), r_{loo} [95%CIrloo] = 0.9659 [0.8929-0.9894], s_{pred} = 0.80, F_{pred} (p) = 203 ($2.57 \cdot 10^{-9}$).

3. Hydrophobic vs. hydrophilic character of standard amino acids

ř = 81.72 + 817.95·*inMrpQg* [7]

n = 20 [8], r [95% CI_r] = 0.9232 [0.8126-0.9695], s_{est} = 20.73, F_{est} (p) = 104 ($6.69 \cdot 10^{-9}$), r_{loo} [95%CI_{rloo}] = 0.9082 [0.7727-0.9645], s_{pred} = 22.58, F_{pred} (p) = 85 ($3.16 \cdot 10^{-8}$).

4. Hydrophobic vs. hydrophilic character of standard amino acids

ř = 1.36-0.20·*iIPmLQt* [7]

n = 20 [9], r [95% CI_r] = 0.9252 [0.8172-0.9704], s_{est} = 0.36, F_{est} (p) = 107 ($5.30 \cdot 10^{-9}$), r_{loo} [95%CI_{rloo}] = 0.9003 [0.7546-0.9613], s_{pred} = 0.42, F_{pred} (p) = 75 ($8.02 \cdot 10^{-8}$).

5. Hydrophobic vs. hydrophilic character of standard amino acids

ř = -7.60 + 19.17·*IiDRLQt* [7]

n = 20 [6], r [95% CI_r] = 0.9328 [0.8348-0.9734], s_{est} = 1.11, F_{est} (p) = 120 ($2.10 \cdot 10^{-9}$), r_{loo} [95%CI_{rloo}] = 0.9226 [0.8062-0.9702], s_{pred} = 1.18, F_{pred} (p) = 103 ($7.25 \cdot 10^{-9}$).

6. Hydrophobic vs. hydrophilic character of standard amino acids

ř = 0.86-0.96·*lAmrLQg* [7]

n = 20 [10], r [95% CI_r] = 0.9376 [0.8461-0.9754], s_{est} = 0.12, F_{est} (p) = 131 ($1.09 \cdot 10^{-9}$), r_{loo} [95%CI_{rloo}] = 0.9263 [0.8149-0.9716], s_{pred} = 0.13, F_{pred} (p) = 109 ($4.73 \cdot 10^{-9}$).

7. Hydrophobic vs. hydrophilic character of standard amino acids

$\hat{r} = 86.05 + 843.88 \cdot inMrpQg$ [7]

n = 19 [11], r [95%CI_r] = 0.9504 [0.8794-0.9805], s_{est} = 16.49, F_{est} (p) = 159 ($4.77 \cdot 10^{-10}$), r_{loo} [95%CI_{rloo}] = 0.9380 [0.8428-0.9762], s_{pred} = 18.37, F_{pred} (p) = 125 ($3.00 \cdot 10^{-9}$).

8. Water activated carbon adsorption of organic compounds

$\hat{r} = 2.58 + 0.85 \cdot IiMMWHt + 0.003 \cdot lPMDVQg$ [12]

n = 16 [13], r [95%CI_r] = 0.9905 [0.9755-0.9963], s_{est} = 0.05, F_{est} (p) = 337 ($6.30 \cdot 10^{-12}$), r_{loo} [95%CI_{rloo}] = 0.9873 [0.9654-0.9953], s_{pred} = 0.06, F_{pred} (p) = 251 ($4.14 \cdot 10^{-11}$).

9. Toxicity of Polychlorinated Organic Compounds

$\hat{r} = 4.06 - 4.95 \cdot imDrkQt + 0.09 \cdot LHDROQg + 0.06 \cdot iSPRtQg$

n = 31 [14], r [95% CI_r] = 0.9692 [0.9364-0.9851], s_{est} = 0.15, F_{est} (p) = 140 ($1.11 \cdot 10^{-16}$), r_{loo} [95%CI_{rloo}] = 0.9613 [0.9194-0.9816], s_{pred} = 0.16, F_{pred} (p) = 109 ($3.22 \cdot 10^{-15}$).

10. Toxicity of mono-substituted nitrobenzene

$\hat{r} = 6.27 - 91.15 \cdot IBMrkGg$

n = 39 [15], r [95% CI_r] = 0.7717 [0.6029-0.8742], s_{est} = 0.35, F_{est} (p) = 54 ($8.87 \cdot 10^{-9}$), r_{loo} [95%CI_{rloo}] = 0.7474 [0.5619-0.8612], s_{pred} = 0.37, F_{pred} (p) = 48 ($4.71 \cdot 10^{-8}$).

11. Toxicity of benzene derivates

$\hat{r} = 3.25 - 9.66 \cdot ABmrsQg + 1.00 \cdot iGPrfHt$

n = 69 [16], r [95% CI_r] = 0.9331 [0.8937-0.9581], s_{est} = 0.28, F_{est} (p) = 222 ($1.48 \cdot 10^{-30}$), r_{loo} [95%CI_{rloo}] = 0.9267 [0.8834-0.9542], s_{pred} = 0.29, F_{pred} (p) = 201 ($2.97 \cdot 10^{-29}$).

12. Toxicity of alkyl metal compounds

$\hat{r} = 2.80 + 28.06 \cdot IbMmpMg + 0.08 \cdot LPPROQg$ [17]

n = 10 [18], r [95%CI_r] = 0.9988 [0.9947-0.9997], s_{est} = 0.06, F_{est} (p) = 1473 ($6.49 \cdot 10^{-10}$), r_{loo} [95%CI_{rloo}] = 0.9980 [0.9901-0.9995], s_{pred} = 0.07, F_{pred} (p) = 841 ($4.57 \cdot 10^{-9}$).

13. Toxicity of para-substituted phenols

$$\text{ŕ} = 0.09 + 5.56\cdot10^{-3}\cdot isDDkGg - 0.42\cdot IMmrKQg + 9.41\cdot10^{-3}\cdot lPMDKQg - 0.08\cdot lFMMKQg \quad [19]$$

n = 30 [20], r [95% CI_r] = 0.9890 [0.9767-0.9948], s_{est} = 0.17, F_{est} (p) = 279 ($1.10\cdot10^{-22}$), r_{loo} [95%CI_{rloo}] = 0.9839 [0.9655-0.9924], s_{pred} = 0.21, F_{pred} (p) = 189 ($2.58\cdot10^{-20}$).

14. Relative toxicity of para-substituted phenols

$$\text{ŕ} = -3.29 + 0.04\cdot ASMmVQt - 0.33\cdot lfDdOQg + 0.08\cdot InMrLQg - 0.35\cdot LsDMpQg \quad [21]$$

n = 30 [20], r [95% CI_r] = 0.9868 [0.9721-0.9937], s_{est} = 0.12, F_{est} (p) = $1.50\cdot10^{-21}$, r_{loo} [95%CI_{rloo}] = 0.9823 [0.9621-0.9917], s_{pred} = 0.14, F_{pred} (p) = $9.34\ 10^{-20}$.

15. Cytotoxicity of quinoline

$$\text{ŕ} = -4.49 + 8.35\cdot INDRLQt + 1.96\cdot lHPmTMt \quad [22]$$

n = 15 [23], r [95% CI_r] = 0.9882 [0.9638-0.9961], s_{est} = 0.17, F_{est} (p) = 250 ($1.65\cdot10^{-10}$), r_{loo} [95%CI_{rloo}] = 0.9805 [0.9377-0.9939], s_{pred} = 0.22, F_{pred} (p) = 149 ($3.34\cdot10^{-9}$).

16. Mutagenicity of quinoline

$$\text{ŕ} = -1.57 + 0.21\cdot lNMrSQg + 0.09\cdot ASPrVQg \quad [22]$$

n = 14 [23], r [95% CI_r] = 0.9782 [0.9306-0.9932], s_{est} = 0.18, F_{est} (p) = 122 ($3.12\cdot10^{-8}$), r_{loo} [95%CI_{rloo}] = 0.9666 [0.8891-0.9902], s_{pred} = 0.22, F_{pred} (p) = 78 ($3.18\cdot10^{-7}$).

17. Antioxidant efficacy of 3-indolyl derivates

$$\text{ŕ} = 7.18 - 1.10\cdot lbPMkHg - 33.24\cdot iAPrVGt \quad [24]$$

n = 8 [25], r [95%CI_r] = 0.9999 [0.9994-0.9999], s_{est} = 0.01, F_{est} (p) = 12591 ($5.55\cdot10^{-10}$), r_{loo} [95%CI_{rloo}] = 0.9997 [0.9978-0.9999], s_{pred} = 0.02, F_{pred} (p) = 3877 ($1.05\cdot10^{-8}$).

18. Antiallergic activity of substituted N 4-methoxyphenyl benzamides

$$\text{ŕ} = -0.15 + 9.02\cdot10^{-4}\cdot imMRkMg - 0.32\cdot imMDVQg - 5.24\cdot10^{-5}\cdot isDRtHg + 0.14\cdot iHMMtHg$$

n = 23 [26], r [95%CI_r] = 0.9986 [0.9966-0.9994], s_{est} = 0.07, F_{est} (p) = 1638 ($7.04\cdot10^{-27}$), r_{loo} [95%CI_{rloo}] = 0.9978 [0.9945-0.9991], s_{pred} = 0.08, F_{pred} (p) = 1007 ($1.45\cdot10^{-24}$).

19. Antituberculotic activity of polyhydroxyxanthones

$$\text{ř} = -19.11 + 2.32 \cdot lHPDOQg + 19.34 \cdot IsMRKGg \quad [27]$$

n = 10 [28], r [95%CI_r] = 0.9987 [0.9942-0.9997], s_{est} = 0.03, F_{est} (p) = 1327 ($9.33 \cdot 10^{-10}$), r_{loo} [95%CI_{rloo}] = 0.9974 [0.9871-0.9994], s_{pred} = 0.04, F_{pred} (p) = 663 ($1.05 \cdot 10^{-8}$).

20. Growth inhibition activity of taxoids

$$\text{ř} = -17.7 + 0.002 \cdot isMdTHg + 77.22 \cdot IiDrQHg \quad [29]$$

n = 34 [30], r [95% CI_r] = 0.9583 [0.9174-0.9791], s_{est} = 0.36, F_{est} (p) = 174 ($2.86 \cdot 10^{-18}$), r_{loo} [95%CI_{rloo}] = 0.9507 [0.9016-0.9755], s_{pred} = 0.39, F_{pred} (p) = 146 ($2.22 \cdot 10^{-16}$).

21. Anti-HIV-1 potencies of HEPTA and TIBO derivatives

$$\text{ř} = 17.72 - 7.11 \cdot InMdTHg - 1.23 \cdot lFDMwEt + 8.36 \cdot AiMrKQt + 6.59 \cdot 10^{5} \cdot ImDMtQt - 5.98 \cdot lIMdEMg \quad [31]$$

n = 57 [32], r [95% CI_r] = 0.9579 [0.9292-0.9750], s_{est} = 0.45, F_{est} (p) = 113 ($5.17 \cdot 10^{-28}$), r_{loo} [95%CI_{rloo}] = 0.9485 [0.9133-0.9696], s_{pred} = 0.49, F_{pred} (p) = 91 ($1.16 \cdot 10^{-25}$).

22. Inhibition activity on carbonic anhydrase I of substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides

$$\text{ř} = 1.14 + 8.79 \cdot 10^{-2} \cdot inPRlQg + 3.52 \cdot 10^{-3} \cdot lPDMoMg + 2.43 \cdot iAMRqQg + 1.04 \cdot inMRkQt \quad [33]$$

n = 40 [34], r [95% CI_r] = 0.9579 [0.9212-0.9776], s_{est} = 0.16, F_{est} (p) = 97 ($9.45 \cdot 10^{-20}$), r_{loo} [95%CI_{rloo}] = 0.9440 [0.8950-0.9704], s_{pred} = 0.19, F_{pred} (p) = 71 ($2.22 \cdot 10^{-16}$).

23. Inhibition activity on carbonic anhydrase II of substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides

$$\text{ř} = -9.99 + 4.56 \cdot imDdSCg + 2.94 \cdot 10\text{-}3 \cdot isDrqQg + 5.20 \cdot IIMDQQg + 1.48 \cdot lmMrsGg \quad [35]$$

n = 40 [34], r [95% CI_r] = 0.9506 [0.9079-0.9737], s_{est} = 0.17, F_{est} (p) = 82 ($1.85 \cdot 10^{-18}$), r_{loo} [95%CI_{rloo}] = 0.9383 [0.8846-0.9674], s_{pred} = 0.19, F_{pred} (p) = 64 ($1.22 \cdot 10^{-15}$).

24. Inhibition activity on carbonic anhydrase IV of substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides

$$\text{ř} = 0.62 + 0.10 \cdot inPRlQg + 9.92 \cdot 10^{-9} \cdot iHMMTQt - 9.25 \cdot IHMDTQg + 1.73 \cdot InPdJQg \quad [36]$$

n = 40 [34], r [95% CI_r] = 0.9593 [0.9238-0.9784], s_{est} = 0.16, F_{est} (p) = 101 ($5.03 \cdot 10^{-20}$),
r_{loo} [95%CI_{rloo}] = 0.9505 [0.9069-0.9739], s_{pred} = 0.18, F_{pred} (p) = 82 ($2.10 \cdot 10^{-18}$).

25. Inhibition activity of dipeptides

$$\text{ŕ} = -7.20 + 0.24 \cdot IbMmjHg + 0.02 \cdot IbPdPHg - 0.24 \cdot IBMRQCg + 2.08 \cdot ImDmEEt - 0.04 \cdot ImDrFEt$$

n = 58 [37], r [95% CI_r] = 0.9618 [0.9360-0.9772], s_{est} = 0.29, F_{est} (p) = 128 ($9.89 \cdot 10^{-30}$),
r_{loo} [95%CI_{rloo}] = 0.9539 [0.9226-0.9726], s_{pred} = 0.31, F_{pred} (p) = 145 ($1.87 \cdot 10^{-27}$).

26. Inhibition activity of 2,4-Diamino-5-(substituted-benzyl)-Pyrimidines

$$\text{ŕ} = 3.78 + 1.62 \cdot iImrKHt + 2.37 \cdot liMDWHg + 6.40 \cdot IsDrJQt - 0.09 \cdot LSPmEQg$$

n = 67 [38], r [95% CI_r] = 0.9517 [0.9223-0.9701], s_{est} = 0.19, F_{est} (p) = 149 ($2.78 \cdot 10^{-32}$),
r_{loo} [95%CI_{rloo}] = 0.9451 [0.9115-0.9661], s_{pred} = 0.20, F_{pred} (p) = 130 ($1.70 \cdot 10^{-30}$).

27. Inhibition activity of peptide analogues

$$\text{ŕ} = 0.81 - 5.21 \cdot 10^{-2} \cdot lmDRsQg + 1.84 \cdot 10^{-3} \cdot iAPrtQg + 240.89 \cdot IHMdpMg - 9.64 \cdot 10^{-2} \cdot IHMdOMg$$

n = 47 [39], r [95% CI_r] = 0.9697 [0.9459-0.9830], s_{est} = 0.16, F_{est} (p) = 165 ($1.12 \cdot 10^{-26}$),
r_{loo} [95%CI_{rloo}] = 0.9611 [0.9303-0.9784], s_{pred} = 0.18, F_{pred} (p) = 127 ($3.06 \cdot 10^{-24}$).

Conclusions and Final Remarks

Realized MDF method and their application MDF-SAR proved to be a very good tool for design of chemical compounds. A series of papers given on results section (over fifty) exposed their ability on investigated sets. The idea about realizing of MDF feigned close to finalizing of PhD studies of first author (Prof. Dr. Mircea V. DIUDEA being his PhD Advisor), but method were implemented just in 2004 (see [40], methodology being revised in 2005 [41]). Further studies will be done in this field, another project being started in 2007, having as main objective creating of a procedure for automatic generating of virtual compounds, based on concepts of combinatorial chemistry. A lesson learned: MDF and MDF-SAR shown miscarries of current methods of constructing/optimizing of molecular geometry (being not capable to provide verifiable and reproducible solutions at a reasonable confidence

level). Because MDF give too many weight on geometry, a new method will replace the MDF, a method called MDFV (being already online), a much conservative method regarding molecular topology relative to MDF. An online application compute statistics on physical models of best obtained MDF-SARs, being available at:

http://l.academicdirect.org/Chemistry/SARs/MDF_SARs/stats/.

Statistics are:

÷ Contribution of descriptors by sets for best models;
÷ Inclusion of descriptors by sets for best models;
÷ Classification of interactions by sets for best models;
÷ Contribution of descriptors by sets for all models;
÷ Inclusion of descriptors by sets for all models;
÷ Classification of interactions by sets for all models.

At the end, the best performing model obtained with MDF-SAR [42] as well as the developed methodology for assessing of structure-activity relationships [43] required to be mentioned here.

As further plans, the study [44] opens a new path in structure-activity relationships approach and will be further investigated.

Acknowledgements

Special acknowledgments from first author to Prof. Mircea V. DIUDEA, his PhD Advisor from 1997 to 2000. Knowledge basis in the field were obtained during this period.

The MDF project was granted from 2005 to 2007 (ET36). The MDF-SAR part of MDF is granted from 2006 to 2008 (ET108). First author (as principal investigator) and second author (as co-investigator) are gratefully to UEFISCSU Romania for this.

References

1. Hammett LP, The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives, J Am Chem Soc, 1937, 59(1), p. 96-103.

2. Hansch C, Leo A, Taft RW, A Survey of Hammett Substituent Constants and Resonance and Field Parameters, Chem Rev, 1991, 91, p. 165-195.

3. Heisenberg WK, Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik, Zeitschrift für Physik, 1927, 43, p. 172-198. English translation: Wheeler JA, Zurek H, Quantum Theory and Measurement Princeton Univ Press, 1983, p. 62-84.

4. BolboacŁ SD, Jäntschi L, A Structural Informatics Study on Collagen, Chemical Biology & Drug Design. In press

5. Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, Nilsson I, White SH, von Heijne G, Recognition of transmembrane helices by the endoplasmic reticulum translocon, Nature, 2005, 433, p. 377-381.

6. Kyte J, Doolittle RF, A Simple Method for Displaying the Hydropathic Character of a Protein, J Mol Biol 1982, 157, p. 105-132.

7. BolboacŁ SD, Jäntschi L, Is Amino Acids Hydrophobicity a Matter of Scale?, Recent Advances in Synthesys & Chemical Biology VI, Centre for Synthesis & Chemical Biology, University of Dublin, Symposium, December 14, Dublin, Ireland, 2007, Abstract in the Book of Abstracts at P2.

8. Sereda TJ, Mant CT, Sonnichsen FD, Hodges RS, Reversed-phase chromatography of synthetic amphipathic a-helical peptides as a model for ligand/receptor interactions effect of changing hydrophobic environment on the relative hydrophilicity/hydrophobicity of amino acid side- chains, J Chromatogr A, 1994, 676(1), p. 139-153.

9. Bull HB, Breese K, Surface tension of amino acid solutions: A hydrophobicity scale of the amino acid residues, Arch Biochem Biophys, 1974, 161, p. 665-670.

10. Black SD, Mould DR, Black SD, Mould DR, Development of Hydrophobicity Parameters to Analyze Proteins Which Bear Post- or Cotranslational Modifications, Anal Biochem, 1991, 193, p. 72-82.

11. Monera OD, Sereda TJ, Zhou NE, Kay CM, Hodges RS, Relationship of sidechain hydrophobicity and alpha-helical propensity on the stability of the single-stranded amphipathic alpha-helix, J Pept Sci, 1995, 1(5), 319-329.

12 Jäntschi L. Water Activated Carbon Organics Adsorption Structure - Property Relationships, Leonardo Journal of Sciences, 2004, 5, p. 63-73.

13. Brasquet C, Le Cloirec P, QSAR for Organics Adsorption onto Activated Carbon In Water: What About The Use Of Neural Networks? Water Research, 1999, 33(17), p. 3603-3608.
14. Wei D, Zhang A, Wu C, Han S, Wang L, Progressive study and robustness test of QSAR model based on quantum chemical parameters for predicting BCF of selected polychlorinated organic compounds (PCOCs), Chemosphere 2001, 44, p. 1421-1428.
15. Agrawala VK, Khadikarb PV, QSAR Prediction of Toxicity of Nitrobenzenes, Bioorganic & Medicinal Chemistry, 2001, 9, p. 3035-3040.
16. Toropov AA, Toropova AP, QSAR modeling of toxicity on optimization of correlation weights of Morgan extended connectivity, Journal of Molecular Structure (THEOCHEM), 2002, 578, p. 129-134.
17. BolboacĂ SD, Jäntschi L, Modeling of Structure-Toxicity Relationship of Alkyl Metal Compounds by Integration of Complex Structural Information, Terapeutics, Pharmacology and Clinical Toxicology, 2006, X(1), p. 110-114.
18. Ade T, Zaucke F, Krug HF, The structure of organometals determines cytotoxicity and alteration of calcium homeostasis in HL-60 cells, Fresenius Journal of Analytical Chemistry, 1996, 354, p. 609-614.
19. Jäntschi L, BolboacĂ SD, Modeling the octanol-water partition coefficient of substituted phenols by the use of structure information, International Journal of Quantum Chemistry, Wiley, 2007, 107(8), p. 1736-1744.
20. Ivanciuc O, Artificial neural networks applications, Part 4. Quantitative structure-activity relationships for the estimation of the relative toxicity of phenols for Tetrahymena. Revue Roumanian de Chimie, 1998, 43(3), p. 255-260.
21. Jäntschi L, Popescu V, BolboacĂ SD, Toxicity Caused by Para-Substituents of Phenole on Tetrahymena Pyriformis and Structure-Activity Relationships, Electronic Journal of Biotechnology, Accepted.
22. Jäntschi L, BolboacĂ S, Molecular Descriptors Family on QSAR Modeling of Quinoline-based Compounds Biological Activities, The 10th Electronic Computational Chemistry Conference, April 2005.
23. Smith CJ, Hansch C, Morton MJ, QSAR treatment of multiple toxicities: the mutagenicity and cytotoxicity of quinolines, Mutation Research, 1997, 379, p. 167-175.

24. BolboacŁ S, Filip C, ŝigan ķ, Jäntschi L, Antioxidant Efficacy of 3-Indolyl Derivates by Complex Information Integration, Clujul Medical 2006, LXXIX(2), p. 204-209.

25. Shertzer GH, Tabor MW, Hogan ITD, Brown JS, Sainsbury M, Molecular modeling parameters predict antioxidant efficacy of 3-indolyl compounds, Archives of Toxicology, 1996, 70, p. 830-834.

26. Zhou YX, Xu L, Wu YP, Liu BL, A QSAR study of the antiallergic activities of substituted benzamides and their structures, Chemometrics and Intelligent Laboratory Systems, 1999, 45, p. 95-100.

27. BolboacŁ SD, Jäntschi L, Molecular Descriptors Family on Structure Activity Relationships 3. Antituberculotic Activity of some Polyhydroxyxanthones, Leonardo Journal of Sciences, 2005, 4(7), p. 58-64.

28. Frahm AW, Chaudhuri R, ^{13}C-NMR-Spectroscopy of substituted xamthones-II 13C-NMR spectra study of polyhydroxyxanthone, Tetrahedon, 1979, 35, p. 2035-2038.

29. BolboacŁ SD, Jäntschi L, Structure Activity Relationships of Taxoids therein Molecular Descriptors Family Approach, Archives of Medical Sciences, Sent for publication.

30. Morita H, Gonda A, Wei L, Takeya K, Itokawa H, 3D QSAR Analysis Of Taxoids From Taxus Cuspidata Var. Nana by Comparative Molecular Field Approach, Bioorganic & Medicinal Chemistry Letters, 1997, 7(18), p. 2387-2392.

31. BolboacŁ S, ŝigan ķ, Jäntschi L. Molecular Descriptors Family on Structure-Activity Relationships on anti-HIV-1 Potencies of HEPTA and TIBO Derivatives. Proceedings of the European Federation for Medical Informatics Special Topic Conference, April 6-8, 2006, p. 222-226.

32. Castro EA, Torrens F, Toropov AA, Nesterov IV, Nabiev OM, QSAR Modeling ANTI-HIV-1 Activities by Optimization of Correlation Weights of Local Graph Invariants, Molecular Simulation, 2004, 30(10), p. 691-696.

33. BolboacŁ SD, Jäntschi L, Modelling the Inhibitory Activity on Carbonic Anhydrase I of Some Substituted Thiadiazoleand Thiadiazoline-Disulfonamides: Integration of Structure Information, Computer-Aided Chemical Engineering, Elsevier Netherlands & UK, 2007, 24, p. 965-970.

34. Supuran CT, Clare BW, Carbonic anhydrase inhibitors - Part 57: Quantum chemical QSAR of a group of 1,3,4-thiadiazole- and 1,3,4-thiadiazoline disulfonamides with carbonic anhydrase inhibitory properties, European Journal of Medical Chemistry, 1999, 34, p. 41-50.

35. Jäntschi L, Ungureşan ML, Bolboacă SD, Complex Structural Information Integration: Inhibitor Activity on Carbonic Anhydrase II of Substituted Disulfonamides, Applied Medical Informatics, 2005, 17, p. 12-21.

36. Jäntschi L, Bolboacă S, Modelling the Inhibitory Activity on Carbonic Anhydrase IV of Substituted Thiadiazole- and Thiadiazoline- Disulfonamides: Integration of Structure Information, Electronic Journal of Biomedicine, 2006, 2, p. 22-33.

37. Opris D, Diudea MV, Peptide Property Modeling by Cluj Indices, SAR and QSAR in Environmental Research, 2001, 12, p. 159-179.

38. Selassie CD, Li R-L, Poe M, Hansch C. On the Optimization of Hydrophobic and Hydrophilic Substituent Interactions of 2,4-Diamino-5-(substituted-benzyl)pyrimidines with Dihydrofolate Reductase. J Med Chem 1991, 34, p. 46-54.

39. Hellberg S, Eriksson L, Jonsson J, Lindgren F, Sjostrom M, Skagerberg B, Wold S, An drews, P Int J Pept Protein Res, 1991, 37, p. 414-424.

40. Jäntschi L, MDF - A New QSAR/QSPR Molecular Descriptors Family, Leonardo Journal of Sciences, 2004, 3(4), p. 68-85.

41. Jäntschi L, Molecular Descriptors Family on Structure Activity Relationships 1. Review of the Methodology, Leonardo Electronic Journal of Practices and Technologies, 2005, 4(6), p. 76-98.

42. Jäntschi L, Bolboacă SD, Diudea MV, Chromatographic Retention Times of Polychlorinated Biphenyls: from Structural Information to Property Characterization, International Journal of Molecular Sciences, 2007, 8(11), p. 1125-1157.

43. Bolboacă SD, Jäntschi L, Modelling the Property of Compounds from Structure: Statistical Methods for Models Validation, Environmental Chemistry Letters, DOI 10.1007/s10311-007-0119-9.

44. Bolboacă SD, Jäntschi L, How Good the Characteristic Polynomial Can Be for Correlations?, International Journal of Molecular Sciences, 2007, 8(4), p. 335-345.