

Results from the Use of Molecular Descriptors Family on Structure Property/Activity Relationships

Lorentz Jäntschi ¹ and Sorana-Daniela Bolboacă ^{2,*}

¹ Technical University, Cluj-Napoca, 15 Constantin Daicoviciu, 400020 Cluj-Napoca, Romania

E-mail: lori@academicdirect.org

² "Iuliu Hațieganu" University of Medicine and Pharmacy, 13 Emil Isac, 40023 Cluj-Napoca, Romania

E-mail: sorana@sorana.academicdirect.ro

* Author to whom correspondence should be addressed. E-mail: sorana@sorana.academicdirect.ro

Received: 14 January 2007 / Accepted: 2 March 2007 / Published: 9 March 2007

Abstract: The aim of the paper is to present the results obtained by utilization of an original approach called Molecular Descriptors Family on Structure-Property (MDF-SPR) and Structure-Activity Relationships (MDF-SAR) applied on classes of chemical compounds and its usefulness as precursors of models elaboration of new compounds with better properties and/or activities and low production costs. The MDF-SPR/MDF-SAR methodology integrates the complex information obtained from compound's structure in unitary efficient models in order to explain properties/activities. The methodology has been applied on a number of thirty sets of chemical compounds. The best subsets of molecular descriptors family members able to estimate and predict property/activity of interest were identified and were statistically and visually analyzed. The MDF-SPR/MDF-SAR models were validated through internal and/or external validation methods. The estimation and prediction abilities of the MDF-SPR/MDF-SAR models were compared with previous reported models by applying of correlated correlation analysis, which revealed that the MDF-SPR/MDF-SAR methodology is reliable. The MDF-SPR/MDF-SAR methodology opens a new pathway in understanding the relationships between compound's structure and property/activity, in property/activity prediction, and in discovery, investigation and characterization of new chemical compounds, more competitive as costs and property/activity, being a method less expensive comparative with experimental methods.

Keywords: Molecular Descriptors Family (MDF), Structure Property Relationships (SPR), Structure Activity Relationships (SAR), Activity/Property Modeling; Chemical Compounds, Open System

1. Introduction

Beginning with nineteenth century, characterization of activities and/or properties of chemical compound was done by applying of structure-activity relationships (QSAR) or quantitative structure-property (QSPR) methodologies, mathematical approaches of linking chemical structure and activity/property of chemical compounds in a quantitative manner [1].

Observations related to the relationships between activity/property and compounds structure has been actually reported before the apparition of the QSAR/QSPR concepts. In 1868, Crum-Brown and Fraser stipulate the idea that the activity of a compound is a function of its chemical composition and structure [2]. In 1893, Richet and Seancs shown for a set of organic molecules that the cytotoxicity was inverse related with water solubility [3]. Mayer suggests in 1899 that the narcotic action of a group of organic compounds is related with solubility into olive oil [4]. Ferguson introduced in 1939 a thermodynamic generalization to the correlation of depressant action with the relative saturation of volatile compounds in the vehicle in which they were administered [5]. Hammett [6] and Taft [7] put together the mechanistic basis of QSAR/QSPR development.

Ten years after defining of the QSAR/QSPR methods, these paradigms found their applicability in practice of agro-chemistry, pharmaceutical chemistry, toxicology and other chemistry related fields [8].

In QSAR/QSPR analysis, the electronic [9,10], hydrophobic [11,12], steric [13,14] and topologic [15,16] descriptors are most frequently used. Pure topological indices used in QSAR/QSPR analysis are: Wiener index [17], Szeged index [18], and Cluj index [19,20]. More, the QSAR is used nowadays in drug investigations being seen as a useful tool in design of new compounds [21,22], in characterization of activity by the use of gene expression programming [23], and in analysis of the relationships between compounds structure and associated biological activity [24,25].

An original approach called Molecular Descriptors Family on Structure-Property/Activity Relationships (MDF-SPR/MDF-SPR) has been developed [26]. The MDF-SAR/SPR methodology, a unitary approach based on minimal complex knowledge obtained from the compound's structure, was applied on different classes of compounds. Obtaining models proved to have estimation and prediction abilities and are presented here. Starting with the MDF-SAR/MDF-SPR models, an opens system has been developed in order to provide a virtual experimental environment with applicability in analysis and characterization of properties/activities of chemical compounds.

2. Materials and Methods

A number of thirty sets of chemically compounds were investigated with MDF-SPR/MDF-SAR methodology. Twenty out of thirty sets (66.66%, 95%CI [46.78 – 83.22]) has been sets with an activity of interest while the others (33.33%, 95%CI [16.78 – 53.22]) had a property of interest. The abbreviation of the set, the type of the observed or measured property/activity, and the class of the compounds are presented in Table 1.

Table 1. Set abbreviation, observed/measured property/ activity, and compounds classes.

No	Set name	Observed/Measured Activity/ Property	Compounds
1	IChr	Retention chromatography index	Organophosphorus herbicides
2	PCB_rrf	Relative response factor	Polychlorinated biphenyls
3	PCB_lkow	Octanol/water partition coefficient	Polychlorinated biphenyls
4	PCB_rrt	Relative retention time	Polychlorinated biphenyls
5	23159	Octanol/water partition coefficients	Polychlorinated biphenyls
6	23159e	Octanol/water partition coefficients	Polychlorinated biphenyls
7	RRC_lko w	Octanol/water partition coefficient	Octanol/water partition coefficient
8	36638	Water activated carbon adsorption	Organic compounds
9	MR10	Molar refraction	Cyclic Organophosphorus
10	33504	Boiling point	Alkanes
11	Ta395	Cytotoxicity	Quinolines
12	Tox395	Mutagenicity	Quinolines
13	RRC_lbr	Toxicity	Para substituted phenols
14	RRC_pka	Relative toxicity	Para substituted phenols
15	52730	Toxicity	Alkyl metal compounds
16	23110	Toxicity	Benzene derivates
17	23167	Toxicity	Polychlorinated organic compounds
18	23158	Toxicity	Mono-substituted nitrobenzenes
19	41521	Insecticidal activity	Neonicotinoids
20	Triaz	Herbicidal activity	Substituted triazines
21	Dipep	Inhibition activity	Dipeptides
22	52344	Antioxidant efficacy	3-indolyl derivates
23	26449	Antituberculosic activity	Polyhydroxyxanthenes
24	23151	Antimalarial activity	2,4-diamino-6-quinazoline sulfonamides
25	22583	Anti-HIV-1 potencies	HEPTA and TIBO derivatives
26	19654	Antiallergic activity	Substituted N 4-methoxyphenyl benzamides
27	3300	Growth inhibition activity	Taxoids
28	40846_1	Inhibitory activity on carbonic anhydrase I	Substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides
29	40846_2	Inhibitory activity on carbonic anhydrase II	Substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides
30	40846_4	Inhibitory activity on carbonic anhydrase IV	Substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides

The property or activity of each sample of compound was modeled by the use of the MDF methodology [26]. The steps followed in the modeling process, described in details in [26], were:

- Compounds preparation: the three-dimensional representation of each compound was built up by using HyperChem software [27]. The property or activity of each compound for the sample of interest was store in *.txt file.
- Molecular descriptor family generation and computing: all compounds belonging to the sample of interest were used in the construction and generation of the molecular descriptors family. The algorithm used is strictly based on the complex information obtained from the compounds structure. Each calculated descriptor has an individual name of seven letters, which express the modality of construction:
 - Compound characteristic relative to its geometry (g) or topology (t) - *the 7th letter*;
 - Atomic property (C = cardinality, H = number of directly bonded hydrogen's, M = atomic relative mass, E = atomic electronegativity, G = group electronegativity, and Q = partial charge, semi-empirical Extended Hückel model, Single Point approach) - *the 6th letter*;
 - Atomic interaction descriptor - *the 5th letter*;
 - Overlapping interaction model - *the 4th letter*;
 - Fragmentation criterion (m = minimal fragments, M = maximal fragments, D = Szeged fragments criterion, and P = Cluj fragments criterion (P) [28,29] - *the 3rd letter*;
 - Cumulative method of fragmentation properties (*conditional group*: m = smallest fragmental descriptor value from the array, M = highest value, n = smallest absolute value, and N = highest absolute value; *average group*: S = sum of descriptor values, A = average mean for valid fragments, a = average mean for all fragments, B = average mean by atom, b= average mean by bond; *geometric group*: P = multiplication of descriptor values, G = geometric mean for valid fragments, g = geometric mean for all fragments, F = geometric mean by atom, and f = geometric mean by bond; *harmonic group*: s = harmonic sum of values, H = harmonic mean for valid fragments, h = harmonic mean for all fragments, I = harmonic mean by atom, and i = harmonic mean by bond) - *the 2nd letter*;
 - Linearization procedure applied in molecular descriptor generation (I = identity, i = inverse, A = absolute, a = an inverse of absolute, L = natural logarithm of absolute value, and l = simple natural logarithm) – *1st letter*.
- Search and identification of MDF-SPR/MDF-SAR models: the criteria imposed into searching and identification of the model were: the model significance, the values for the correlation and squared correlation coefficients, the standard error, and the significances of the coefficients associated to the molecular descriptors.
- MDF-SPR/MDF-SAR models validation: two methods were applied for internal validation of the obtained models. The methods were: the leave-one-out procedure (the property or activity of each compound was predicted by the regression equation calculated based on all the other compounds), and the leave-many-out procedure (the property or activity of a number of compounds discard from the sample were predicted by the regression equation calculated based on all the other compounds).
- MDF-SPR/MDF-SAR analysis and comparison with previous reported models (where is applicable): the chosen MDF-SPR/MDF-SAR models were analyzed through computing and interpreting of a number of seven statistical parameters and visually by model plotting.

Starting from the previous experience in development of online systems [30,31], PHP (Hypertext Preprocessor) and MySQL (My Structure Query Language) has been used in development of the open system.

The characteristics of the previous reported models and of the MDF-SPR/MDF-SAR models were summarized by using Statistica software. The correlation coefficients obtained by the previous reported models were compared with the correlation coefficients obtained by MDF-SPR/MDF-SAR models (the Fisher's Z test at a significance level of 5% was applied [32]).

3. Results

The summaries of characteristics of the MDF-SPR/MDF-SAR and previous reported results models are presented in Table 2.

Table 2. Statistical characteristics of the MDF-SPR/MDF-SAR and previous reported models

Characteristic	Property (10 sets)		Activity (20 sets)	
	Previous	MDF-SPR	Previous	MDF-SAR
Sample size				
<i>Min</i>	10	8	8	8
<i>Max</i>	73	209	69	69
<i>Average</i>	38.36	87.95	26.10	34.09
<i>[95% CI]</i>	[21.01 - 55.71]	[48.78 - 127.12]	[21.09 - 31.11]	[28.90 - 39.28]
Number of variable				
<i>Min</i>	1	1	1	1
<i>Max</i>	4	4	7	5
<i>Median</i>	2	2	4	2
<i>Mode</i>	2	2	5	2
Squared correlation coefficient				
<i>Min</i>	0.3880	0.6288	0.3660	0.6280
<i>Max</i>	0.9986	0.9999	0.9860	0.9998
<i>Average</i>	0.8851	0.8987	0.8472	0.8994
Leave-one-out score				
<i>Min</i>	n.a.	0.6198	n.a.	0.6060
<i>Max</i>	n.a.	0.9999	n.a.	0.9994
<i>Average</i>	n.a.	0.8754	n.a.	0.8783

n.a. = not available

The characteristics of the previous reported models (where were available), of the best performing MDF-SPR and MDF-SAR models are presented in Table 3. There was also included in Table 3 the Z parameter resulted from comparison between the correlation coefficient of previous model and of best performing MDF-SPR, or respectively MDF-SAR model.

Table 3. Characteristics of previous reported models and MDF-SPR, respectively MDF-SAR models

Set abb.	Previous reported			Ref.	MDF				Ref.	$Z_{r_{\text{prev}} \text{ vs. } r_{\text{MDF}}}$
	n_{prev}	v_{prev}	r^2_{prev}		n_{MDF}	v_{MDF}	r^2_{MDF}	$r^2_{\text{cv-100}}$		
<i>Structure-property relationships</i>										
IChr	10	2	0.9000	[33]	10	2	0.9992	0.9985	[34]	4.565 [‡]
23159	18	3	0.8390	[35]	18	2	0.9817	0.9740	[36]	3.081 [†]
36638	16	4	0.6650	[37]	16	3	0.9950	0.9812	[38]	5.605 [‡]
33504	73	3	0.9986	[39]	73	2	0.9982	0.9980	n.a.	0.744 [*]
MR10	10	2	0.9760	[40]	10	2	0.9999	0.9999	[41]	4.492 [‡]
PCB_rrf	n.a.	n.a.	n.a.	[42]	209	4	0.7367	0.7169	[43]	n.a.
PCB_lkow	n.a.	n.a.	n.a.	[42]	206	4	0.9168	0.9093	[44]	n.a.
PCB_rrt	n.a.	n.a.	n.a.	[42]	206	2	0.9970	0.9970	n.a.	n.a.
23159e	n.a.	n.a.	n.a.	[35]	8	2	0.9681	0.8989	n.a.	n.a.
RRC_lko			n.a.	[45]			0.9781	0.9680	[46]	n.a.
w	n.a.	n.a.			30	4				
<i>Structure-activity relationships</i>										
Ta395	13	2	0.8700	[47]	15	2	0.9766	0.9614	[48]	2.066 ⁺
Tox395	13	2	0.8000	[47]	14	2	0.9568	0.9343	[48]	1.893 ⁺
RRC_lbr	30	2	0.9550	[45]	30	4	0.9737	0.9650	n.a.	1.013 [*]
23110	25	5	0.9180	[49]	69	5	0.9360	0.9280	n.a.	0.526 [*]
23167	27	3	0.9300	[50]	31	3	0.9390	0.9240	n.a.	0.253 [*]
23158	40	5	0.8000	[51]	40	2	0.9510	0.9450	n.a.	3.206 [‡]
41521	8	5	0.9850	[52]	8	2	0.9991	0.9982	[53]	2.144 ⁺
Triaz	30	3	0.9700	[54]	30	1	0.9885	0.9850	[55]	1.766 ⁺
Dipep	58	2	0.7820	[56]	58	5	0.9250	0.9100	n.a.	3.011 [†]
52344	8	4	0.9700	[57]	8	2	0.9998	0.9994	[58]	3.972 [‡]
26449	10	4	0.9860	[59]	10	2	0.9974	0.9948	[60]	1.577 [*]
23151	13	4	0.9850	[61]	16	3	0.9972	0.9959	[62]	1.999 ⁺
22583	37	5	0.8830	[63]	57	5	0.9179	0.8994	[64]	0.749 [*]
19654	23	3	0.8865	[65]	23	4	0.9973	0.9956	[66]	5.948 [‡]
3300	35	5	0.9584	[67]	35	4	0.9655	0.9564	n.a.	0.380 [*]
40846_1	40	6	0.7530	[68]	40	4	0.9180	0.8910	n.a.	2.573 [†]
40846_2	40	7	0.7190	[68]	40	4	0.9037	0.8804	[69]	2.535 [†]
40846_4	36	5	0.7690	[68]	40	4	0.9175	0.8911	[70]	3.591 [†]
RRC_pk	n.a.	n.a.	n.a.	[45]	30	4	0.9605	0.9490	n.a.	n.a.
52730	n.a.	n.a.	n.a.	[71]	10	2	0.9998	0.9993	[72]	n.a.

number of compounds used by previous reported model (n_{prev}) and MDF-SPR/MDF-SAR model (n_{MDF});

number of variables used by previous reported model (v_{prev}) and MDF-SPR/MDF-SAR model (v_{MDF});

squared correlation coefficient of previous reported model (r^2_{prev}) and MDF-SPR/MDF-SAR model (r^2_{MDF});

cross validation leave-one-out score ($r^2_{\text{cv-100}}$); $Z_{r_{\text{prev-MDF}}}$ = Fisher's Z parameter of comparison between

correlation coefficients; n.a. = not available; [‡] $p \leq 0.001$; [†] $0.001 < p < 0.01$; ⁺ $0.01 < p < 0.05$; ^{*} $p > 0.05$

As was specified in material and method section, the MDF-SPR and MDF-SAR results were integrated into an open system. The open system incorporates distinct programs useful in analysis and characterization of compounds properties/activities. The system is hosted on AcademicDirect domain and it is available at the following URL:

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

The named and the functions of the programs are presented in Table 4.

Table 4. The open system: programs and characteristics

Program	Function
<i>BorQ SARs by sets (BorQ: Browse or Query)</i>	
Browse	Display for a set of data the MDF-SPR/MDF-SAR equations, with some statistical parameters (the squared correlation coefficient, the number of descriptors, and the sample size).
Query	Display the following characteristics for the MDF-SPR/MDF-SAR investigation: the size of the molecular descriptor family, the MDF-SPR/MDF-SAR equations, the number of descriptors used by the models, the sample size for each model, the values of each descriptor, the squared correlation coefficient, the leave-one-out score, the squared correlation coefficient between each descriptor and measured/observed property/activity.
<i>DC Predictor (DC: demo calculator)</i>	
	Provide a demo calculation of the Molecular Descriptors Family for a specified compound (a *.hin file) based on characteristics choused by the user.
<i>SARs (SAR: Structure-Activity Relationship)</i>	
	A previous obtained MDF-SPR/MDF-SAR model(s) is used (learning set). The property/activity of a new compound from the same class as learning compounds can be predicted based on its structure. A *.hin file of the compound of interest is necessary.
<i>LOO Analysis (LOO: leave one out)</i>	
	Based on the data resulted form MDF-SPR/MDF-SAR investigation the program is able to compute the leave-one-out score and to display statistical characteristics of the estimated and predicted property/activity of interest (number of descriptors used by the model, degree of freedom, standard error, standard deviation, squared correlation coefficient, Fisher parameter and associated significance). The program is able to work just with tabulated data (with labels on columns and rows). The columns must be organized as followed: independent variables (first sets of columns), estimated dependent variable, measured/observed dependent variable, and predicted variable.
<i>Investigator</i>	
	Display the characteristics of the sets of molecules which are in analysis. The administrator of the system is able to delete the MDF-SPR/MDF-SAR models which are considered not being at the level of imposed conditions and desires.
<i>TvT Experiment (TvT: Training vs. Test)</i>	
	Based on previous analyzed set of compounds, the program randomly split the compounds into training and test sets (the user can impose the number of compounds in training set). The MDF-SPR/MDF-SAR equation is calculated on the training set and applied on the test set. The program display the molecular descriptors and associated values for compounds in training and test sets, the MDF-SPR/MDF-SAR equations, the squared correlation coefficient, the Fisher parameter and associated significance for training and test sets.

4. Discussion

The paper presented the estimation and prediction abilities of the Molecular Descriptors Family in characterization of property/activity of chemical compounds. Based on the obtained results a virtual environmental library has been created.

Four observations can be notice looking at the entire ensemble of chemical compounds sets. First, the squared correlation coefficients and associated correlation coefficients, measures of statistical fit of the regression, had always values greater in MDF-SPR/MDF-SAR models comparing with previous reported models. With one exception (for the relative response factor of polychlorinated biphenyls PCB_rrf set) the MDF-SPR/MDF-SAR models obtained squared correlation coefficients greater than 0.9 (see Table 3). The results of the squared correlation coefficient and of the leave-one-out score being greater than 0.8 in the majority of the cases sustain in accordance with the Basak at all criteria [73] the predictive abilities of the models. With one exception (for alkanes set with boiling point as property, the 33504 set), the squared correlation coefficients obtained by best performing MDF-SPR model was greater than the squared correlation coefficient previous reported (see Table 3). Note that, the lowest performance was obtained for the relative response factor of polychlorinated biphenyls (PCB_rrf set, see Table 3). The relative response factor is a complex property that depends by many factors not just by the compound's structure (all external factors of gas chromatography and/or HRGC/ECD methods). In seventy percent of cases the squared correlation coefficient was statistical significant greater comparing with previous reported models (see Table 3).

Second observation refers the number of descriptors used in the MDF-SPR/MDF-SAR models. It is well known that the fitting power of the model become greater by increasing the number of descriptors, being generally accepted that a regression model with ν descriptors for a sample size of n could be acceptable only if the following criterion is satisfied: $n \geq 4-5 \cdot \nu$ [74]. As it can be observed, with three exceptions (RRC_lbr, Dipep, and 19654 sets), the number of descriptors used by MDF-SPR/MDF-SAR models was less than the number of variables used in previous reported models. More, with two exceptions (41521, 52344 sets, both of them with 8 compounds), the condition imposed by Hawkins [74] was respected by the MDF-SPR/MDF-SAR models. In five cases, the previous reported models did not respect the above describe condition (compound sets: 36638 - $5 \cdot \nu$ condition, 41521 - both conditions, 52344 - both conditions, 26449 - both conditions, and 23151 - both conditions).

The third observation that can be noted regards the sample sizes used by the previous reported models and by the MDF-SPR/MDF-SAR models. There were seven sets in which the previous reported model was obtained after exclusion of one (Tox395 set), two (Ta395 set), three (23151 set), four (40846_4 and 23167 sets), twenty (22583 set), or forty-four (23110) compounds, while the MDF-SPR/MDF-SAR models were using in all cases the whole sample size (see Table 3).

The last but not the least observation regards the stability of the MDF-SPR/MDF-SAR models (defines as the differences between squared correlation coefficient and the cross-validation leave-one-out score) which sustained the prediction abilities (see Table 3).

The performances of the MDF-SPR/MDF-SAR methodology were good in estimation and prediction of properties of different chemical classes. For example, retention chromatography index (IChr set) and molar refraction (MR10 set), are two properties which can be estimated and predicted

with high accuracy (squared correlation coefficients and leave-one-out cross-validations scores greater than 0.99, see Table 3).

Good performances in estimation and prediction are obtained for octanol/water partition coefficients (23159, 23159e, RRC433_lkow, PCB_lkow sets) with squared correlation coefficients from 0.9168 to 0.9817, and cross-validation leave-one-out scores from 0.8989 (23159e set) to 0.9740 (23159 set). The abilities of estimation and prediction of water activated carbon organics adsorption (36638 set) is significantly greater comparing with previous reported models ($Z = 15.605$, see Table 3), sustaining that the studied property of organics is related with compounds structure (see Table 2, $r^2 = 0.9950$, $r^2_{cv-100} = 0.9812$).

These results suggest that the physicochemical properties of compounds are in relationships with compounds structure and the information obtained from their structure can be useful in property characterization.

Regarding the toxicity of chemical compounds it can be say that the MDF-SAR models estimate and predict almost perfect the toxicity of alkyl metal compounds (52730 set, $r^2 = 0.9998$, $r^2_{cv-100} = 0.9993$, see Table 3) and obtains high performances in estimation of cytotoxicity of studied quinolines (Ta395 set, Table 3). Good performances (Tox395 set, see Table 3) and significantly greater comparing with previous reported model (Tox395 set, $Z = 1.893$, Table 3) are also obtained in estimation and prediction of mutagenicity of studied quinolines.

Looking at the performances of the MDF-SAR models obtained on insecticidal (41521 set) and herbicidal (Triaz set) activities it can be observed that, even if the previous reported models had squared correlation coefficients close to one (see Table 3), the MDF-SAR models obtained statistical significant greater correlation coefficients ($Z = 2.144$ for 41521 set, and $Z = 1.766$ for Triazines set, Table 3).

High performances are obtained by MDF-SAR models in estimation and prediction of antioxidant efficacy of studied 3-indolyl derivates (52344 set), where the obtained values for squared correlation coefficient and cross-validation leave-one-out score are greater than 0.999 (see Table 3).

The abilities of MDF-SAR methodology in investigations of drugs were revealing in the study of antituberculosic activity of some polyhydroxyxanthenes (26449 set), antimalarial activity of some 2,4-diamino-6-quinazoline sulfonamide derivates (23151 set) and carbonic anhydrase inhibitors (40846_1, 40846_2, and 40846_4 sets). The squared correlation coefficients were greater than 0.9 and the cross-validation leave-one-out scores were greater than 0.88 (see Table 3).

Even if the MDF-SAR model obtained greater squared correlation coefficient in investigation of anti-HIV-1 potencies of HEPTA and TIBO derivatives (22583 set), there was not identify statistical significance between MDF-SAR correlation coefficient and correlation coefficients obtained by previous reported models ($p \geq 0.05$, see Table 3).

The MDF-SPR/MDF-SAR proved to offer reliable and valid models in characterization of property/activity of chemical compounds. The results indicate that important information regarding compounds property/activity can be obtained by analyzing the compounds structure.

Base on the above results, the developed open system provides an environment of modeling the property/activity of chemical compounds assisted by a computer, offering to the researchers the

alternative of free risks experiments. The analysis of the system can be done through its advantages and disadvantages.

The advantages offered by the system can be summarizing as follows:

- ÷ Provides the values of molecular descriptors family calculated based on information obtained strictly from the compounds structure for studied classes of compounds;
- ÷ Identify the best performing models based on generated molecular descriptors family;
- ÷ Display a summary report of statistical characteristics of the best performing models;
- ÷ Provides parameters of measures the goodness-of-fit, the robustness and the predictivity of the obtained models;
- ÷ Allows to the user to visualize a demo of how the program calculate a molecular descriptor;
- ÷ Predict the property/activity of new compounds from a class previously studied based on the best performing MDF-SPR/MDS-SAR model.

Note that the costs of virtual experiments are less comparing with real experiments. In addition, the experiment risks are withdrawn. Comparing with the experimental approach, the proposed online system provides a stable and valid alternative in studying of relationships between compounds structure and associated activity/property.

In order to use the system facilities, the user must to have a computer connected to the Internet and browses skills. This can be considering at least for the researchers from developing countries a disadvantage of the system.

The open system provide effective models which can be used in studying the property/activity of new compounds in real time, without any experiments, and with low costs, being necessary just building up as *.hin files the three dimensional structure of the new compound and a previous study on the same class of compound. The future development of the system will allow the access to as many sets of compounds as possible, opening a new pathway in study of relationship between property/activity and structure of chemical compounds.

The MDF-SPR/MDF-SAR methodology opens a new pathway in understanding the relationships between compounds structure and property/activity, in characterization, investigation and development of new compounds, more competitive as production costs and property/activity.

Acknowledgements

We would like to thank Prof. Dr. I. Černušák and Prof. Dr. J. Noga from Comenius University, Bratislava, Slovak Republic, for the discussions and suggestions on the topic presented in this article. The research was partly supported by UEFISCSU Romania through project ET/36/2005.

References and Notes

1. Hansch, C. Quantitative approach to biochemical structure-activity relationships. *Accounts of Chemical Research* **1969**, 2(8), 232-239.
2. Crum-Brown, A.; Fraser, T.R. On the Connection between Chemical Constitution and Physiological Action. Part I. On the Physiological Action of the Salts of the Ammonium Bases, derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia. *Philosophical Transactions of the Royal Society of London* **1868**, 25, 151-2003.

3. Richet, C.R. *Comptes Rendus des Seances de la Societe de Biologie et de ses Filiales*, **1893**, 9, 775.
4. Meyer, H. Zur Theorie der Alkoholnarkose Erste Mittheilung. Welche Eigenschaft der Anästhetica bedingt ihre narkotische Wirkung? *Naunyn-Schmiedeberg's Archives of Pharmacology* **1899**, 42(2-4), 109-118.
5. Ferguson, J. *Proceedings of the Royal Society B: Biological Sciences* **1939**, 127, 387.
6. Hammett, L. P. Some Relations between Reaction Rates and Equilibrium Constants. *Chemical Reviews* **1935**, 17, 125-136.
7. Taft, R.W. Polar and Steric Substituent Constants for Aliphatic and o-Benzoate Groups from Rates of Esterification and Hydrolysis of Esters. *Journal of the American Chemical Society* **1952**, 74(12), 3120 - 3128.
8. Hansch, C.; Leo, A. Substituent Constants for Correlation Analysis in Chemistry and Biology. John Wiley & Sons: New York, **1979**.
9. Aptula, A.; Roberts, D.W.; Cronin, M.T.D.; Schultz, T.W. Chemistry-Toxicity Relationships for the Effects of Di- and Trihydroxybenzenes to *Tetrahymena pyriformis*. *Chemical research in toxicology* **2005**, 18, 844-854.
10. Johnson-Restrepo, B.; Pacheco-Londoño, L.; Olivero-Verbel, J. Molecular Parameters Responsible for the Melting Point of 1,2,3-Diazaborine Compounds. *Journal of Chemical Information and Modeling* **2003**, 43, 1513-1519.
11. Ponce, Y.M.; Marrero, R.M.; Castro, E.A.; De Armas, R.R.; Díaz, H.G.; Zaldivar, V.R.; Torrens, F. Protein Quadratic Indices of the Macromolecular Pseudograph's α -Carbon Atom Adjacency Matrix. 1. Prediction of Arc Repressor Alanine-mutant's Stability. *Molecules* **2004**, 9, 1124-1147.
12. Huibers, P.D.T.; Shah, D.O.; Katritzky, A.R. Predicting Surfactant Cloud Point from Molecular Structure. *Journal of Colloid and Interface Science* **1997**, 193, 132-136.
13. Juvale, D.C.; Kadam, S.S.; Kulkarni, V.M. Comparative 2D and 3D-QSAR of antifungal griseofulvin analogues. *Indian Journal of Chemistry Sect. A: Inorganic, Bio-inorganic, Physical, Theoretical & Analytical* **2006**, 45, 194-201.
14. De Lima Ribeiro, F.A.; Ferreira, M.M.C. QSPR models of boiling point, octanol-water partition coefficient and retention time index of polycyclic aromatic hydrocarbons. *Journal of Molecular Structure: THEOCHEM* **2003**, 663(1-3), 109-126.
15. Jin, A.Y.; Kohn, H.; Béguin, C.; Andurkar, S.V.; Stables, J.P.; Weaver, D.F. A quantitative structure-activity relationship study for α -substituted acetamido-N-benzylacetamide derivatives - A novel anticonvulsant drug class. *Canadian Journal of Chemistry* **2005**, 83, 37-45.
16. Acevedo-Martínez, J.; Escalona-Arranz, J.C.; Villar-Rojas, A.; Téllez-Palmero, F.; Pérez-Rosés, R. González, L., Carrasco-Velaz, R. Quantitative study of the structure-retention index relationship in the imine family. *Journal of Chromatography A* **2006**, 1102(1-2), 238-244.
17. Lather, V.; Madan, A.K. Topological models for the prediction of HIV-protease inhibitory activity of tetrahydropyrimidin-2-ones. *Journal of Molecular Graphics and Modelling* **2005**, 23(4), 339-345.
18. Khadikar, P.V.; Deshpande, N.V.; Kale, P.P.; Dobrynin, A.; Gutman, I.; Domotor, G.J. The Szeged index and an analogy with the Wiener index. *Journal of Chemical Information and Computer Science* **1995**, 35, 547-550.

19. Jäntschi, L.; Katona, G.; Diudea, M.V. Modeling Molecular Properties by Cluj Indices. *MATCH - Communications in Mathematical and in Computer Chemistry* **2000**, *41*, 151-188.
20. Diudea, M.; Gutman, I.; Jäntschi, L. Molecular Topology. Nova Science, Huntington: New York, **2002**, Chapter 6. p. 173-196.
21. González, M.P.; Terán, C.; Teijeira, M.; Helguera, A.M. Radial distribution function descriptors: An alternative for predicting A2 A adenosine receptors agonists. *European Journal of Medicinal Chemistry* **2006**, *41*(1), 56-62.
22. Narasimhan, B.; Mourya, V.; Dhake, A. Design, synthesis, antibacterial, and QSAR studies of myristic acid derivatives. *Bioorganic and Medicinal Chemistry Letters* **2006**, *16*(11), 3023-3029.
23. Si, H.Z.; Wang, T.; Zhang, K J.; Hu, Z.D.; Fan, B.T. QSAR study of 1,4-dihydropyridine calcium channel antagonists based on gene expression programming. *Bioorganic and Medicinal Chemistry* **2006**, *14*(14), 4834-4841.
24. Saliner, A.G.; Netzeva, T.I.; Worth, A.P. Prediction of estrogenicity: Validation of a classification model. *SAR and QSAR in Environmental Research* **2006**, *17*(2), 195-223.
25. Chang, C.; Swaan, P. W. Computational approaches to modeling drug transporters. *European Journal of Pharmaceutical Sciences* **2006**, *27*(5), 411-424.
26. Jäntschi, L. Molecular Descriptors Family on Structure Activity Relationships 1. Review of the Methodology, *Leonardo Electronic Journal of Practices and Technologies* **2005**, *6*, 76-98.
27. HyperChem , Molecular Modelling System [online]; ©2003, Hypercube [cited 2007 Feb]. Available from: URL: <http://hyper.com/products/>
28. Jäntschi, L.; Katona, G.; Diudea, M. V. Modeling Molecular Properties by Cluj Indices. *Commun. Math. Comput. Chem. (MATCH)*, **2000**, *41*, 151-188.
29. Diudea, M.; Gutman, I.; Jäntschi, L. Molecular Topology, 2nd Edition, Nova Science, Huntington: New York, **2002**.
30. Jäntschi, L. Free Software Development. 1. Fitting Statistical Regressions. *Leonardo Journal of Sciences* **2002**, *1*, 31-52.
31. Bolboacă, S.D.; Jäntschi, L.; Deneş, C.; Achimaş Cadariu, A. Skeletal Maturity Assessment Client-Server Application. *Roentgenologia & Radiologia* **2005**, *XLIV*(3), 189-193.
32. Steiger, J.H. Tests for comparing elements of a correlation matrix. *Psychological Bulletin* **1980**, *87*, 245-251.
33. Jäntschi, L.; Mureşan, S.; Diudea, M.; Modeling Molecular Refraction and Chromatographic Retention by Szeged Indices. *Studia Universitatis Babeş-Bolyai, Chemia* **2000**, *XLV*(1-2), 313-318.
34. Jäntschi L. MDF - A New QSAR/QSPR Molecular Descriptors Family. *Leonardo Journal of Sciences* **2004**, *4*, 67-84.
35. Baker, J.R.; Mihelcic, J.R.; Sabljic, A. Reliable QSAR for estimating KOC for persistent organic pollutants: correlation with molecular connectivity indices. *Chemosphere* **2001**, *45*, 213-221.
36. Jäntschi, L. Delphi Client - Server Implementation of Multiple Linear Regression Findings: a QSAR/QSPR Application. *Applied Medical Informatics* **2004**, *15*, 48-55.
37. 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.

38. Jäntschi, L. Water Activated Carbon Organics Adsorption Structure - Property Relationships. *Leonardo Journal of Sciences* **2004**, 5, 63-73.
39. Toropova, A.; Toropova, A.; Ismailov, T.; Bonchev, D. 3D weighting of molecular descriptors for QSPWQSAR by the method of ideal symmetry (MIS). 1. Application to boiling points of alkanes. *Journal of Molecular Structure (THEOCHEM)* **1998**, 424, 237-247.
40. Jäntschi, L.; Mureşan, S.; Diudea, M.V. Modeling molar refraction and chromatographic retention by Szeged Indices. *Studia Universitatis Babeş-Bolyai, Chemia* **2000**, XLV(1-2), 313-318.
41. Jäntschi, L.; Bolboacă, S. Molecular Descriptors Family on Structure Activity Relationships 4. Molar Refraction of Cyclic Organophosphorus Compounds. *Leonardo Electronic Journal of Practices and Technologies* **2005**, 7, 55-102.
42. Eisler, R.; Belisle, A.A. Planar PCB Hazards to Fish, Wildlife, and Invertebrates: A Synoptic Review, Biological Report 31 and Contaminant Hazard Reviews Report **1996**, 31(IV), 75. http://www.pwrc.usgs.gov/infobase/eisler/CHR_31_Planar_PCBs.pdf
43. Jäntschi, L. QSPR on Estimating of Polychlorinated Biphenyls Relative Response Factor using Molecular Descriptors Family. *Leonardo Electronic Journal of Practices and Technologies* **2004**, 5, 67-84.
44. Jäntschi, L.; Bolboacă, S. Molecular Descriptors Family on Structure Activity Relationships 6. Octanol-Water Partition Coefficient of Polychlorinated Biphenyls. *Leonardo Electronic Journal of Practices and Technologies* **2006**, 8, 71-86.
45. 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) 255-260.
46. Jäntschi, L.; Bolboacă, S.D. Modeling the Octanol-Water Partition Coefficient of Substituted Phenols by the Use of Structure Information. *International Journal of Quantum Chemistry*. In Press, Published Online: 3 Jan 2007.
47. Smith, C.J.; Hansch, C.; Morton, M.J. QSAR treatment of multiple toxicities: the mutagenicity and cytotoxicity of quinolines. *Mutation Research* **1997**, 379, 167-175.
48. Jäntschi, L.; Bolboacă, S. Molecular Descriptors Family on QSAR Modeling of Quinoline-based Compounds Biological Activities. *The 10th Electronic Computational Chemistry Conference*. April **2005**, <http://eccc.monmouth.edu>.
49. Toporov, A.A.; Toporova, A.P. QSAR modeling of toxicity on optimization of correlation weights of Morgan extended connectivity. *Journal of Molecular Structure (THEOCHEM)* **2002**, 578, 129-134.
50. 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, 1421-1428.
51. Agrawala, V.K.; Khadikarb, P.V. QSAR Prediction of Toxicity of Nitrobenzenes. *Bioorganic & Medicinal Chemistry* **2001**, 9, 3035-3040.
52. Hasegawa, K.; Arakawa, M.; Funatsu, K. 3D-QSAR study of insecticidal neonicotinoid compounds based on 3-way partial least squares model. *Chemometrics and Intelligent Laboratory Systems* **1999**, 47, 33-40.

53. Bolboacă, S.; Jäntschi, L. Molecular Descriptors Family on Structure Activity Relationships 2. Insecticidal Activity of Neonicotinoid Compounds. *Leonardo Journal of Sciences* **2005**, *6*, 78-85.
54. Diudea, M.; Jäntschi, L.; Pejov, L. Topological Substituent Descriptors. *Leonardo Electronic Journal of Practices and Technologies* **2002**, *1*, 1-18.
55. Bolboacă, S.; Jäntschi, L. Molecular Descriptors Family on Structure-Activity Relationships: Modeling Herbicidal Activity of Substituted Triazines Class. *Bulletin of University of Agricultural Sciences and Veterinary Medicine - Agriculture* **2006**, *62*, 35-40.
56. Opris, D.; Diudea, M.V. Peptide Property Modeling by Cluj Indices. *SAR and QSAR in Environmental Research* **2001**, *12*, 159-179.
57. Shertzer, G.H.; Tabor, M.W.; Hogan, I.T.D.; Brown, J.S.; Sainsbury, M. Molecular modeling parameters predict antioxidant efficacy of 3-indolyl compounds. *Archives of Toxicology* **1996**, *70*, 830-834.
58. Bolboacă, S.; Filip, C.; Țigan, Ș.; Jäntschi, L. Antioxidant Efficacy of 3-Indolyl Derivates by Complex Information Integration. *Clujul Medical* **2006**, *Issue LXXIX(2)*, 204-209.
59. Ungwitayatorn, J.; Pickert, M.; Frahm, A.W. Quantitative structure-activity relationship (QSAR) study of polyhydroxyxanthenes. *Pharmaceutica Acta Helvetiae* **1997**, *72*, 23-29.
60. Bolboacă, S.; Jäntschi, L. Molecular Descriptors Family on Structure Activity Relationships 3. Antituberculous Activity of some Polyhydroxyxanthenes. *Leonardo Journal of Sciences* **2005**, *7*, 58-64.
61. Agrawal, V.K.; Srivastava, R.; Khadikar, P.V. QSAR Studies on Some Antimalarial Sulfonamides. *Bioorganic & Medicinal Chemistry* **2001**, *9*, 3287-3293.
62. Jäntschi, L.; Bolboacă, S. Molecular Descriptors Family on Structure Activity Relationships 5. Antimalarial Activity of 2,4-Diamino-6-Quinazoline Sulfonamide Derivates. *Leonardo Journal of Sciences* **2006**, *8*, 77-88.
63. Castro, E.A.; Torrens, F.; Toropov, A.A.; Nesterov, I.V.; Nabiev, O.M. QSAR Modeling ANTI-HIV-1 Activities by Optimization of Correlation Weights of Local Graph Invariants. *Molecular Simulation* **2004**, *30(10)*, 691-696.
64. Bolboacă, S.; Țigan, Ș.; 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.
65. Zhou, Y.X.; Xu, L.; Wu, Y.P.; Liu, B.L. A QSAR study of the antiallergic activities of substituted benzamides and their structures. *Chemometrics and Intelligent Laboratory Systems* **1999**, *45*, 95-100.
66. Jäntschi, L.; Bolboacă, S.D. Antiallergic Activity of Substituted Benzamides: Characterization, Estimation and Prediction. *Clujul Medical* **2006**, *LXXIX(4)*, In Press.
67. 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)*, 2387-2392.

68. Supuran, C.T.; Clare, B.W. 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*, 41-50.
69. Jäntschi, L.; Ungureșan, M.L.; Bolboacă, S.D. Integration of Complex Structural Information in Modeling of Inhibition Activity on Carbonic Anhydrase II of Substituted Disulfonamides. *Applied Medical Informatics* **2005**, *17*(3, 4), 12-21.
70. 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*, 22-33.
71. Ade, T.; Zaucke, F.; Krug, H.F. The structure of organometals determines cytotoxicity and alteration of calcium homeostasis in HL-60 cells. *Fresenius Journal of Analytical Chemistry* **1996**, *354*, 609-614.
72. Bobloacă, S.D.; Jäntschi, L. Modeling of Structure-Toxicity Relationship of Alkyl Metal Compounds by Integration of Complex Structural Information. *Therapeutics, Pharmacology and Clinical Toxicology* **2006**, *X*(1), 110-114.
73. Basak, S.C., Niemi, G.J., Regal, R.R., Veith, G.D. Topological Indices: their Nature, Mutual Relatedness, and Applications. *Mathematical Modelling* **1987**, *8*, 300-305.
74. Hawkins, D.M. The Problem of Overfitting. *Journal of Chemical Information and Computer Science* **2004**, *44*, 1-12.