Online System for Molecular Descriptors Family on Structure-Activity Relationships: Assessment and Characterization of Biologic Active Compounds

Lorentz JÄNTSCHI[1], Sorana Daniela BOLBOACĂ[2]

[1] *Technical University of Cluj-Napoca, 15 Constantin Daicoviciu Street, 400020 Cluj-Napoca, http://lori.academicdirect.org*
[2] *„Iuliu Hatieganu“ University of Medicine and Pharmacy, 13 Emil Isac Street, 400023 Cluj-Napoca, Romania, http://sorana.academicdirect.ro*

Starting with an original methodology of molecular descriptor family (MDF) on structure-activity relationships [1], which was applied to sets of biological active compounds, an open system was created and assessed in order to provide a virtual experimental environment useful in characterization of compounds activities.

A number of thirty-one samples of biologic active compounds were studied by the use of MDF methodology. The best performing models (models with highest abilities in estimation and prediction) were integrated into an online system.

The system integrates the models of thirty one sets of biological active compounds. The range of sets sample sizes vary from 8 to 209 with an average of 45.71 (95% CI [24.87, 66.55]). The previous reported models had an average of the squared correlation coefficients of 0.86 (95% CI [0.82, 0.90]), and a median of the number of variables used equal with 4. The MDF models had an average of squared correlation coefficients equal with 0.90 (95% CI [0.87, 0.92]), an average of cross-validation leave-one-out scores equal with 0.88 (95% CI [0.85, 0.90]), and a median of variable equal with 2. The best performing MDF models had significantly greater correlation coefficients comparing with the previous reported models ($p < 0.05$).

The open system provide effective models which can be used in studying the 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. The future development of the system will allow the access to exhaustive sets of compounds, opening a new pathway in study of theirs activities.

[1]. Jäntschi L. *Leonardo Electronic Journal of Practices and Technologies* 6 (2005) 76-98.

ONLINE SYSTEM FOR MOLECULAR DESCRIPTORS FAMILY ON STRUCTURE-ACTIVITY RELATIONSHIPS: ASSESSMENT AND CHARACTERIZATION OF BIOLOGICAL ACTIVE COMPOUNDS

Lorentz JÄNTSCHI[1], Sorana Daniela BOLBOACĂ[2]

[1] *Technical University of Cluj-Napoca, 15 Constantin Daicoviciu, 400020 Cluj-Napoca, http://lori.academicdirect.org*
[2] *"Iuliu Hatieganu" University of Medicine and Pharmacy, 13 Emil Isac, 400023 Cluj-Napoca, Romania, http://sorana.academicdirect.ro*

ABSTRACT

Starting with an original methodology of molecular descriptor family on structure-activity relationships (MDF-SAR), which was applied to sets of biological active compounds, an open system was created and assessed in order to provide a virtual experimental environment useful in characterization of compounds activities.

The system integrates the models of thirty-one sets of compounds. The range of sets sample sizes vary from 8 to 209 with an average of 45.71 (95% CI [24.87, 66.55]). The previous reported models had an average of the squared correlation coefficients of 0.86 (95% CI [0.82, 0.90]), and a median of the number of variables used equal with 4. The MDF models had an average of squared correlation coefficients equal with 0.90 (95% CI [0.87, 0.92]), an average of cross-validation leave-one-out scores equal with 0.88 (95% CI [0.85, 0.90]), and a median of variable equal with 2. The best performing MDF models had significantly greater correlation coefficients comparing with the previous reported models ($p < 0.05$).

The open system provide reliable models which can be used in studying the 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. The future development of the system will allow the access to exhaustive sets of compounds, opening a new pathway in study of theirs activities.

INTRODUCTION

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

QSAR is used nowadays in drug investigations being seen as a useful tool in design of new compounds [2,3], in characterization of activity by the use of gene expression programming [4], and in analysis of the relationships between compounds structure and their biological activities [5,6].

An original approach called Molecular Descriptors Family on Structure-Property (MDF-SPR) and Structure-Activity Relationships (MDF-SAR) was developed [7]. The MDF-SPR/SAR methodology, a unitary approach based on minimal complex knowledge obtained from the compound's structure, was applied on different classes of chemical compounds. Starting with the MDF-SAR models, an opens system was developed and assessed in order to provide a virtual experimental environment with applicability in analysis and characterization of compounds activities.

METHOD

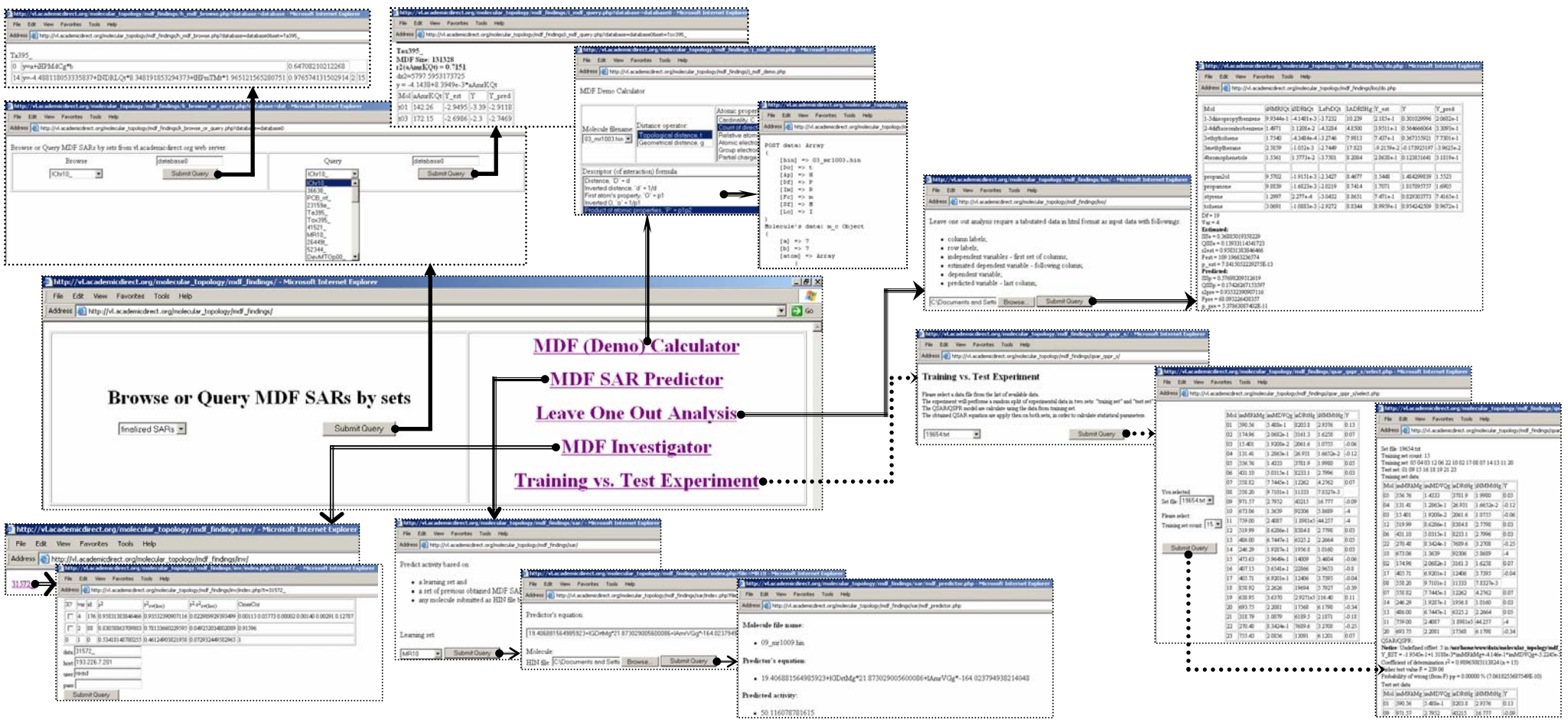
A number of thirty-one classes of compounds were investigated with MDF-SPR/SAR methodology. Twenty-one sets contained biological active compounds. The studied classes of biological active compounds are: substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides, substituted N 4-methoxyphenyl benzamides, HEPTA and TIBO derivatives, taxoids, 2,4-diamino-6-quinazoline sulfonamides, polyhydroxyxanthones, quinolines, alkyl metal compounds, para substituted phenols, benzene derivates, mono-substituted nitrobenzenes, polychlorinated organic compounds, para substituted phenols, volatile organic compounds, quinolines, neonicotinoids, substituted triazines, dipeptides, and 3-indolyl derivates.

The steps of molecular descriptors family on structure-activity/structure-property relationships integrates [7] (1) the approaches of compounds preparing for molecular modeling, (2) the methodology of molecular descriptor family generation, (3) the methodology of finding the best performing MDF-SAR/SPR models, (4) the MDF-SAR/SPR validation, and (5) the comparison of the MDF-SAR/SPR models with previous reported models.

PHP (Hypertext Preprocessor), MySQL and Apache were used in creation of the open system.

RESULTS

The open system integrates six distinct programs useful in analysis and characterization of compounds activities/properties. The system is hosted by AcademicDirect domain being available at: http://vl.academicdirect.org/molecular_topology/mdf_findings/



Statistical characteristics of MDF-SAR/SPR and comparisons with previously reported models are in the following tables:

Statistical characteristics and comparisons: previous reported models vs. MDF-SAR models

Set abb.	Previously reported			MDF-SAR			$Z_{prev-MDF}$
	n_{prev}	v_{prev}	r^2_{prev}	n_{MDF}	v_{MDF}	r^2_{MDF}	
40846_1	20	7	0.9170	40	4	0.9180	0.021
40846_2	20	6	0.9020	40	4	0.9040	0.039
40846_4	20	4	0.8220	40	4	0.9200	0.068
RRC_lbr	30	2	0.9550	30	4	0.9739	1.027
52344	8	4	0.9700	8	2	0.9998	3.972*
19654	23	3	0.8865	23	4	0.9978	6.329*
3300	35	5	0.9790	35	4	0.9665	0.939
31572	24	4	0.9530	24	4	0.9583	0.197
23151	13	4	0.9850	16	3	0.9970	1.917*
23158	40	5	0.8000	40	2	0.9510	3.206*
41521	8	5	0.9850	8	2	0.9990	2.144*
Ta395	13	2	0.8700	15	2	0.9770	2.086*
Tox395	13	2	0.8000	14	2	0.9570	1.86*
Triazines	30	3	0.9700	30	4	0.9890	1.864*
23167	27	3	0.9300	31	3	0.9390	0.253
Dipeptides	58	2	0.7820	58	5	0.9250	3.011*
22583	57	5	0.8830	57	5	0.9180	0.97
23110	69	5	0.9000	69	5	0.9360	1.339

number of compounds used by n_{prev} = previously reported model and n_{MDF} = MDF-SAR model;
number of variables used by v_{prev} = previously reported model and v_{MDF} = MDF-SAR model;
squared correlation coefficient of r^2_{prev} = previously reported model and r^2_{MDF} = MDF-SAR model;
$Z_{prev-MDF}$ = Fisher's Z parameter of comparison between correlation coefficients;
* $p < 0.05$; n.a. = not available

Statistical characteristics of the previously reported models and MDF-SAR/SPR models

Characteristic		Previously reported	MDF-SAR/SPR
Sample size			
	Min	8	8
	Max	73	209
	Average [95%CI]	32.06 [26.20-37.92]	45.71 [24.87-66.55]
Number of variable			
	Min	1	1
	Max	7	5
	Median	4	2
	Mode	5	2
Squared correlation coefficient [95%CI]		0.86 [0.82-0.90]	0.90 [0.87-0.92]
Leave-one-out score [95%CI]		n.a.	0.88 [0.85-0.90]

n.a. = not available

CONCLUSIONS

The open system provide effective models which can be used in studying the 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.

The future development of the system will allow the access to exhaustive sets of compounds, opening a new pathway in study of biological active compounds.

ACKNOWLEDGEMENT

The research was partly supported by UEFISCSU Romania through ET/36/2005 project.

REFERENCES

[1]. C. Hansch, *Acct. Chem. Res.* 2 (1969), 232-239.; [2]. M.P. González, C. Terán, M. Teijeira, A.M. Helguera, *Curr. Med. Chem.* 13 (2006), 2253-2266.; [3]. B. Narasimhan, V. Mourya, A. Dhake, *Bioorg. Med. Chem. Lett.* 16 (2006), 3023-3029.; [4]. H.Z. Si, T. Wang, K.J. Zhang, Z.D. Hu, B.T. Fan, *Bioorg. Med. Chem.* 14 (2006), 4834-4841.; [5]. A. Gallegos Saliner, *Curr. Comput.-Aided Drug Des.* 2(2006), 105-122.; [6]. C. Chang, P.W. Swaan, *Eur. J. Pharm. Sci.* 27 (2006), 411-424.; [7]. L. Jäntschi, *Leonardo Electronic Journal of Practices and Technologies* 6 (2005) 76-98.