

# **Molecular Descriptors Family on Structure-Activity and Structure-Property Relationships: Results**

**Lorentz JÄNTSCHI & Sorana BOLBOACĂ**

Technical University of Cluj-Napoca, Romania,

<http://lori.academicdirect.org>

&

„Iuliu Hatieganu“ University of Medicine and Pharmacy, Romania

<http://sorana.academicdirect.ro>

# Purpose

- To present the results obtained by utilization of an original approach called **Molecular Descriptors Family (MDF)** on **Structure-Property (SPR)** and **Structure-Activity Relationships (SAR)** applied on different classes of chemical compounds and its usefulness as precursors of models elaboration of new chemical compounds with better properties and/or activities.

# Methodology

## Molecular Descriptors Family

- **Preparing chemical compounds for molecular modeling**
- **Generating the molecular descriptors family**
- **Finding the MDF SPR/SAR models**
- **Validating the MDF SPR/SAR models**
- **Comparing the MDF SPR/SAR models with previous reported model(s)**

# Materials

<b>Set name</b>	<b>Observed/Measured Property/Activity</b>
IChr	retention chromatography index
PCB_rrf	relative response factor
23159	octanol/water partition coefficients
23159e	octanol/water partition coefficients
PCB_lkow	octanol/water partition coefficient
36638	water activated carbon adsorption
MR10	molar refraction
Ta395	cytotoxicity
52730	toxicity
Tox395	mutagenicity
41521	insecticidal activity
Triazines	herbicidal activity
52344	antioxidant efficacy
26449	antituberculosic activity
23151	antimalarial activity
22583	anti-HIV-1 potencies

# Results: database

vl.academicdirect.org >> localhost >> SARs | phpMyAdmin 2.6.1-pl3 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://vl.academicdirect.org/phpMyAdmin/index.php?lang=en-utf-8&server=1&collation\_connection=utf8\_general\_ci

phpMyAdmin

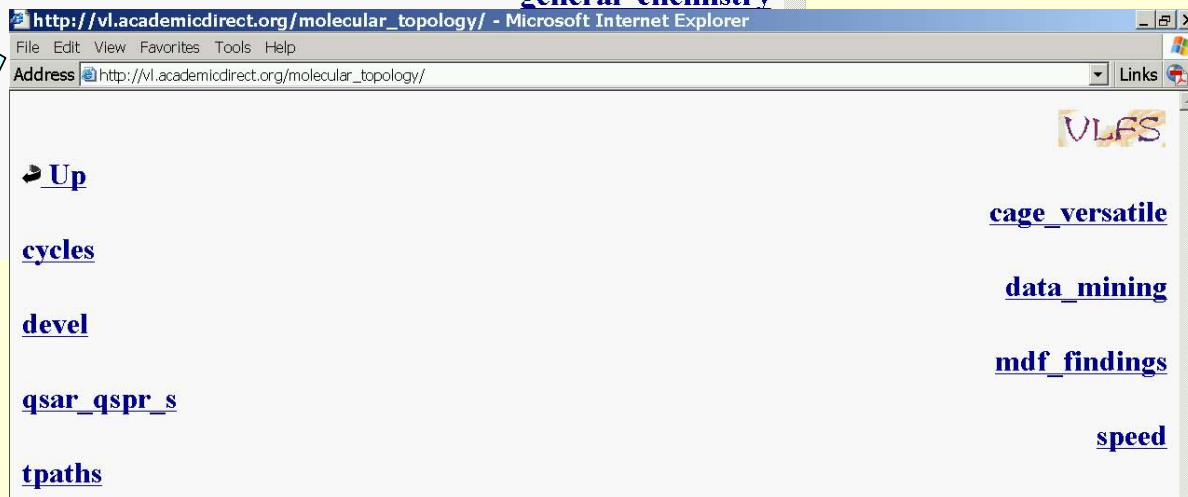
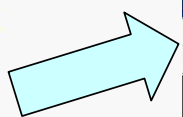
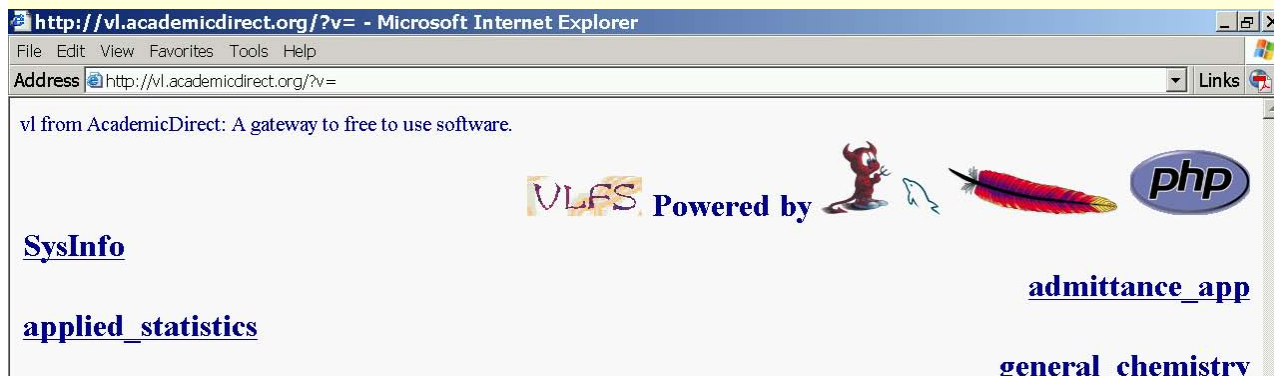
Database: SARs (194)

SARs

- 3300\_data
- 3300\_tmpx
- 3300\_xval
- 3300\_yval
- 19654\_data
- 19654\_tmpx
- 19654\_xval
- 19654\_yval
- 23158c\_data
- 23158c\_tmpx
- 23158c\_xval
- 23158c\_yval
- 23159e\_data
- 23159e\_tmpx
- 23159e\_xval
- 23159e\_yval
- 26449t\_data
- 26449t\_tmpx
- 26449t\_xval
- 26449t\_yval
- 31572\_data
- 31572\_tmpx
- 31572\_xval
- 31572\_yval
- 33504\_data
- 33504\_tmpx

<input type="checkbox"/>	MR10_xval									107,692	MyISAM	latin1_swedish_ci	8.3 MB	-		
<input type="checkbox"/>	MR10_yval									107,692	MyISAM	latin1_swedish_ci	4.9 MB	-		
<input type="checkbox"/>	PCB_rrf_data									209	MyISAM	latin1_swedish_ci	2.8 KB	-		
<input type="checkbox"/>	PCB_rrf_tmpx									131,328	MyISAM	latin1_swedish_ci	212.1 MB	-		
<input type="checkbox"/>	PCB_rrf_xval									62,873	MyISAM	latin1_swedish_ci	100.3 MB	-		
<input type="checkbox"/>	PCB_rrf_yval									62,873	MyISAM	latin1_swedish_ci	2.9 MB	-		
<input type="checkbox"/>	RRC433_lbr_data									30	MyISAM	latin1_swedish_ci	1.3 KB	-		
<input type="checkbox"/>	RRC433_lbr_tmpx									131,328	MyISAM	latin1_swedish_ci	32.7 MB	-		
<input type="checkbox"/>	RRC433_lbr_xval									86,409	MyISAM	latin1_swedish_ci	19.9 MB	-		
<input type="checkbox"/>	RRC433_lbr_yval									86,409	MyISAM	latin1_swedish_ci	4.0 MB	-		
<input type="checkbox"/>	RRC433_lkow_data									30	MyISAM	latin1_swedish_ci	1.3 KB	-		
<input type="checkbox"/>	RRC433_lkow_tmpx									131,328	MyISAM	latin1_swedish_ci	32.7 MB	-		
<input type="checkbox"/>	RRC433_lkow_xval									86,388	MyISAM	latin1_swedish_ci	19.9 MB	-		
<input type="checkbox"/>	RRC433_lkow_yval									86,388	MyISAM	latin1_swedish_ci	4.0 MB	-		
<input type="checkbox"/>	RRC433_pka_data									30	MyISAM	latin1_swedish_ci	1.3 KB	-		
<input type="checkbox"/>	RRC433_pka_tmpx									131,328	MyISAM	latin1_swedish_ci	32.7 MB	-		
<input type="checkbox"/>	RRC433_pka_xval									86,373	MyISAM	latin1_swedish_ci	19.9 MB	-		
<input type="checkbox"/>	RRC433_pka_yval									86,373	MyISAM	latin1_swedish_ci	4.0 MB	-		
<input type="checkbox"/>	Ta395_data									15	MyISAM	latin1_swedish_ci	1.1 KB	-		
<input type="checkbox"/>	Ta395_tmpx									131,328	MyISAM	latin1_swedish_ci	17.7 MB	-		
<input type="checkbox"/>	Ta395_xval									102,608	MyISAM	latin1_swedish_ci	11.8 MB	-		
<input type="checkbox"/>	Ta395_yval									102,608	MyISAM	latin1_swedish_ci	4.7 MB	-		
<input type="checkbox"/>	Tox395_data									14	MyISAM	latin1_swedish_ci	1.1 KB	-		
<input type="checkbox"/>	Tox395_tmpx									131,328	MyISAM	latin1_swedish_ci	16.7 MB	-		
<input type="checkbox"/>	Tox395_xval									103,411	MyISAM	latin1_swedish_ci	11.1 MB	-		
<input type="checkbox"/>	Tox395_yval									103,411	MyISAM	latin1_swedish_ci	4.7 MB	-		
<input type="checkbox"/>	a_acids_data									12	MyISAM	latin1_swedish_ci	1.1 KB	-		
<input type="checkbox"/>	a_acids_tmpx									131,328	MyISAM	latin1_swedish_ci	14.7 MB	-		
<input type="checkbox"/>	a_acids_xval									97,007	MyISAM	latin1_swedish_ci	9.0 MB	-		
<input type="checkbox"/>	a_acids_yval									97,007	MyISAM	latin1_swedish_ci	4.4 MB	-		
<input type="checkbox"/>	molecular_topology									86	MyISAM	latin1_swedish_ci	17.1 KB	-		
										<b>194 table(s)</b>	<b>Sum</b>	<b>14,305,820</b>	<b>--</b>	<b>latin1_swedish_ci</b>	<b>1.9 GB</b>	<b>0 Bytes</b>

# Results: web interface



# Results: MDF model

Address [http://vl.academicdirect.org/molecular\\_topology/mdf\\_findings/j\\_mdf\\_demo.php](http://vl.academicdirect.org/molecular_topology/mdf_findings/j_mdf_demo.php)

## MDF Demo Calculator

<b>Molecule filename:</b> 03_mr1003.hin		<b>Distance operator:</b> Topological distance, t <b>Geometrical distance, g</b>	<b>Atomic property:</b> Cardinality, C Count of directly bounded hydrogen's, H <b>Relative atomic mass, M</b> Atomic electronegativity, E Group electronegativity, G Partial charge, Q
<b>Descriptor (of interaction) formula:</b> 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' = p1p2 <sup>1/2</sup> Inverted Q, 'q' = 1/p1p2 <sup>1/2</sup> <b>First atom's Property multiplied by distance, 'J' = p1 d</b> 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' = d(p1p2) <sup>1/2</sup> Inverted L, 'l' = 1/p1p2d First atom's property potential, 'V' = p1/d First atom's property field, 'E' = p1/d <sup>2</sup> First atom's property work, 'W' = p1 <sup>2</sup> /d Properties work, 'w' = p1p2/d First atom's property force, 'F' = p1 <sup>2</sup> /d <sup>2</sup> Properties force, 'f' = p1p2/d <sup>2</sup> First atom's property weak nuclear force, 'S' = p1 <sup>2</sup> /d <sup>3</sup> Properties weak nuclear force, 's' = p1p2/d <sup>3</sup> First atom's property strong nuclear force, 'T' = p1 <sup>2</sup> /d <sup>4</sup> Properties strong nuclear force, 't' = p1p2/d <sup>4</sup>		<b>Interaction model:</b> Rare model and resultant relative to fragment's head, R Rare model and resultant relative to conventional origin, r Medium model and resultant relative to fragment's head, M <b>Medium model and resultant relative to conventional origin, m</b> Dense model and resultant relative to fragment's head, D Dense model and resultant relative to conventional origin, d	
<b>Fragmentation criteria:</b> Minimal fragments, m Maximal fragments, M <b>Szeged distance based fragments, D</b> Cluj path based fragments, P	<b>Molecular overall superposing formula:</b> Cond., smallest m Cond., highest M Cond., smallest absolute, n Cond., highest absolute, N Avg., sum, S Avg., average, A Avg., S/count(fragments), a <b>Avg., Avg.(Avg./atom)/count(atoms), B</b> <b>Avg., S/count(bonds), b</b> Geom., product, P Geom., mean, G Geom., P <sup>1</sup> /count(fragments), g Geom., Geom.(Geom./atom)/count(atoms), F Geom., P <sup>1</sup> /count(bonds), f Harm., sum, s Harm., mean, H Harm., s/count(fragments), h Harm., Harm.(Harm./atom)/count(atoms), l Harm., s/count(bonds), i		<b>Linearization operator:</b> Identity (no change), I Inversed l, i <b>Absolute l, A</b> Inversed A, a Logarithm of A, L Logarithm of l, l

Submit Query

# Results: investigated datasets

http://vl.academicdirect.org/molecular\_topology/mdf\_findings/k\_browse\_or\_query.php?database=dat - Microso

File Edit View Favorites Tools Help

Address http://vl.academicdirect.org/molecular\_topology/mdf\_findings/k\_browse\_or\_query.php?database=database0

Browse or Query MDF SARs by sets from vl.academicdirect.org web server.

Browse	Query
<input type="text" value="database0"/> <input type="button" value="Submit Query"/>	<input type="text" value="database0"/> <input type="button" value="Submit Query"/>
<ul style="list-style-type: none"><li>IChr10_</li><li>DevMTOp25_</li><li>RRC433_lbr_</li><li>RRC433_lkow_</li><li>RRC433_pka_</li><li>52730_</li><li>DHFR_</li><li>a_acids_</li><li>34121bad_</li><li>34121nopt_</li><li>19654_</li><li>3300_</li></ul>	<ul style="list-style-type: none"><li>IChr10_</li><li>36638_</li><li>PCB_rrf_</li><li>23159e_</li><li>Ta395_</li><li>Tox395_</li><li>41521_</li><li>MR10_</li><li>26449t_</li><li>52344_</li><li>DevMTOp00_</li></ul>



# Results: training vs. test application

Address [http://vl.academicdirect.org/molecular\\_topology/qsar\\_qspr\\_s/](http://vl.academicdirect.org/molecular_topology/qsar_qspr_s/)

## Training vs. Test Experiment

Please select a data file from the list of available data.

The experiment will perform a random split of experimental data in two sets: "trainig set" and "test set".

The QSAR/QSPR model are calculate using the data from training set.

The obtained QSAR equation are apply then on both sets, in order to calculate statistical parameters.

- 23167.txt
- 23167.txt
- 26449.txt
- 31572.txt
- 33504.txt
- 36638\_1.txt
- 36638\_2.1.txt
- 36638\_2.2.txt
- 36638\_3.1.txt
- 36638\_3.2.txt
- 40846\_1.txt
- 40846\_2.txt

Submit Query

You selected:

Set file: 23167.txt

Please select:

Training set count: 18

- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28

Submit Query

Set file: 23167.txt

Training set count: 18

Training set list:

- a-Endosulfan 1-3-5-CB 1-3-CB
- o-p1-DDT 1-2-3-5-CB p-p1-DDT
- Dieldrin 2-2p-5-PCB 1-4-CB 2-
- 2p-4-5-5p-PCB b-BHC 1-2-3-4-
- CB PCB 2-2p-4-4p-5-5p-PCB 1-2-

Test set list:

- Methoxychlor
- Heptachloroepoxide Clordane a-
- BHC g-BHC p-p1-DDE 2-5-PCB 3-
- 5-PCB 2-4-4p-PCB 2-2p-5-5p-
- PCB 1-2-CB Pentachlorobenzene

Submit Query

Address [http://vl.academicdirect.org/molecular\\_topology/qsar\\_qspr\\_s/result.php](http://vl.academicdirect.org/molecular_topology/qsar_qspr_s/result.php)

Set file: 23167.txt

Training set count: 18

Training set: a-Endosulfan 1-3-5-CB 1-3-CB o-p1-DDT 1-2-3-5-CB p-p1  
CB PCB 2-2p-4-4p-5-5p-PCB 1-2-4-CB 1-2-4-5-CB Heptachlor 1-2-3-C

Test set: Methoxychlor Heptachloroepoxide Clordane a-BHC g-BHC p-p1  
Pentachlorobenzene 2-3-5-6-chloronitrobenzene

Training set data:

Mol	imDrkQt	LHDROQg	iSPRtQg	Y
a-Endosulfan	4.9741e-2	-3.0834	-5.3905e-4	3.55
			-6.0117e-4	3.69

# Results: leave one out application

http://vl.academicdirect.org/molecular\_topology/mdf\_findings/loo/ - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://vl.academicdirect.org/molecular\_topology/mdf\_findings/loo/ Links

Leave one out analysis require a tabulated data in html format as input data with followings:

- column labels;
- row labels;
- independent variables - first set of columns;
- estimated dependent variable - following column;
- dependent variable;
- predicted variable - last column;

© November 2005, [Lorentz JÄNTSCHI](#)

Data example:

Mol	iSMmEQt	iSMmWHg	LADmkQt	INPRJQg	Y_Est	Y	Y_Pre
t1	2.0396e-2	7.0397e-4	1.4830	1.6344e-1	3.8295	3.82	3.8619
t2	1.9849e-2	5.212e-4	-1.5099	3.1038e-2	5.2003	5.2	5.2009
t3	1.7683e-2	4.2792e-4	1.8269	1.5977e-2	5.3298	5.34	5.3267
t4	1.5673e-2	3.2419e-4	2.3702	1.8778e-2	5.8340	5.83	5.8348
t5	1.9487e-2	4.0228e-4	1.0849	1.3049e-2	5.9636	6.01	5.9436

# Results: comparisons

Set name	Previous reported SAR			MDF SAR			
	r <sup>2</sup>	n	v	r <sup>2</sup>	r <sup>2</sup> <sub>cv(100)</sub>	n	v
<b>IChr10</b>	<b>0.900</b>	<b>10</b>	<b>2</b>	<b>0.999</b>	<b>0.999</b>	<b>10</b>	<b>2</b>
<b>PCB_rrf</b>	-	-	-	<b>0.628</b>	<b>0.619</b>	<b>209</b>	<b>1</b>
				<b>0.693</b>	<b>0.682</b>	<b>209</b>	<b>2</b>
				<b>0.737</b>	<b>0.717</b>	<b>209</b>	<b>4</b>
<b>PCB_lkow</b>	-	-	-	<b>0.873</b>	<b>0.87</b>	<b>206</b>	<b>1</b>
				<b>0.89</b>	<b>0.885</b>	<b>206</b>	<b>2</b>
				<b>0.917</b>	<b>0.909</b>	<b>206</b>	<b>4</b>
<b>36638</b>	<b>0.967</b>	<b>16</b>		<b>0.994</b>	<b>0.991</b>	<b>16</b>	<b>3</b>
<b>23159</b>	<b>0.388</b>	<b>18</b>	<b>1</b>	<b>0.755</b>	<b>0.684</b>	<b>18</b>	<b>1</b>
	<b>0.839</b>	<b>18</b>	<b>3</b>	<b>0.982</b>	<b>0.974</b>	<b>18</b>	<b>2</b>
<b>23159e</b>	-	-	-	<b>0.899</b>	<b>0.758</b>	<b>8</b>	<b>1</b>
				<b>0.968</b>	<b>0.898</b>	<b>8</b>	<b>2</b>
<b>Ta395</b>	<b>0.870</b>	<b>13</b>	<b>2</b>	<b>0.977</b>	<b>0.961</b>	<b>15</b>	<b>2</b>
<b>Tox395</b>	<b>0.800</b>	<b>13</b>	<b>2</b>	<b>0.957</b>	<b>0.934</b>	<b>14</b>	<b>2</b>

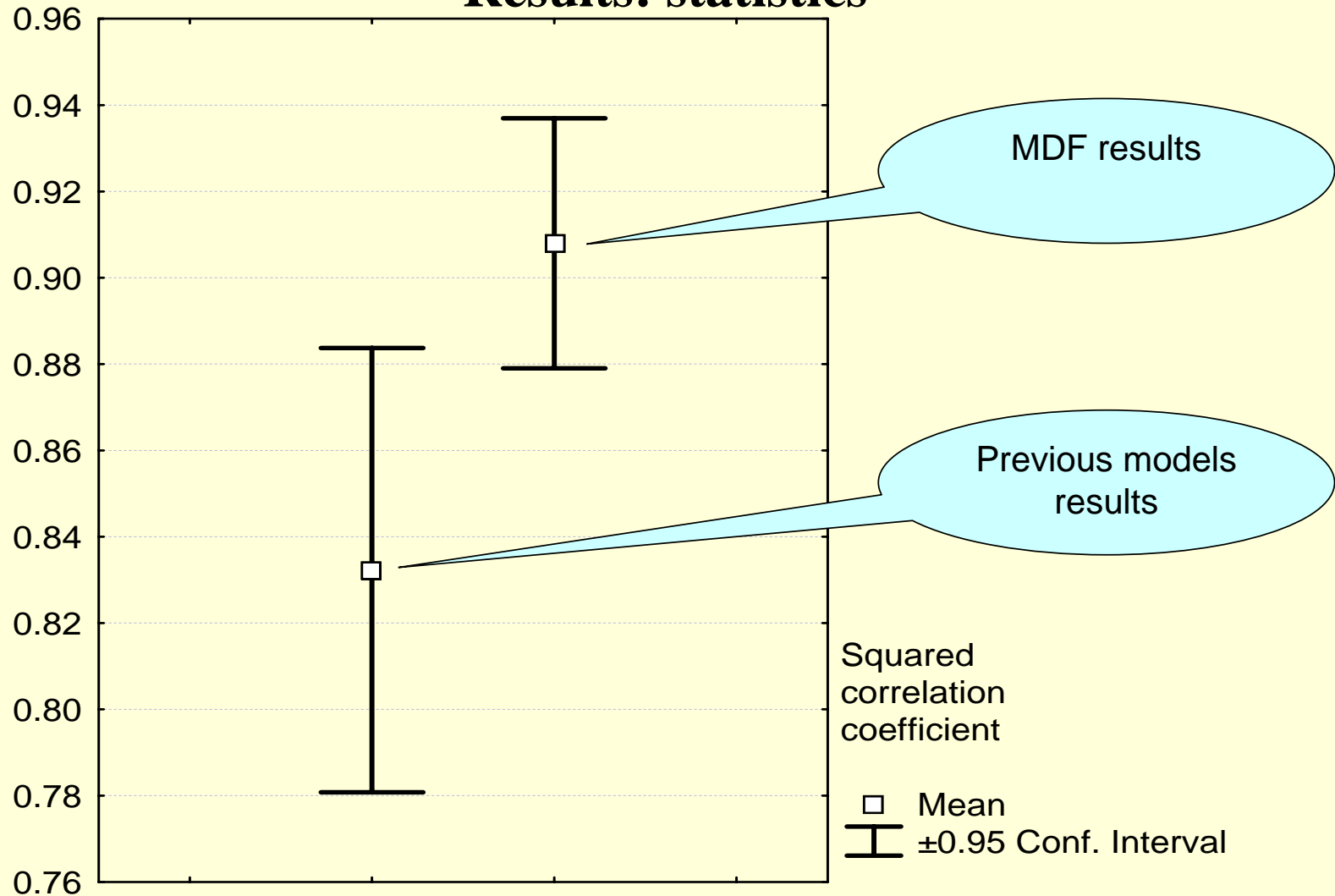
# Results: comparisons

Set name	Previous reported SAR			MDF SAR			
	$r^2$	n	v	$r^2$	$r^2_{cv(100)}$	n	v
41521	0.913	8	3	0.999	0.998	8	2
	0.985	8	5				
26449	0.991		1	0.961	0.954	10	1
	0.998		2	0.99	0.988	10	2
	0.993	10	4	0.998	0.997	10	4
MR10	0.976	10	2	0.999	0.999	10	2
23151	0.741	16	4	0.997	0.995	16	3
	0.985	13	4				
52344	0.780	8	1	0.904	0.832	8	1
	0.710	8	1	0.999	0.999	8	2
	0.810	8	2	0.999	0.999	8	2
	0.970	8	4	0.999	0.999	8	2

# Results: comparisons

Set name	Previous reported SAR			MDF SAR			
	r <sup>2</sup>	n	v	r <sup>2</sup>	r <sup>2</sup> <sub>cv(100)</sub>	n	v
52730	-	-	-	0.966	0.947	10	1
				0.998	0.996	10	2
Triazines	0.970	30	3	0.951	0.946	30	1
				0.975	0.971	30	2
				0.983	0.976	30	3
				0.989	0.985	30	4
22583	0.888	37	5	0.783	0.766	57	2
				0.885	0.809	57	3
				0.883	0.884	57	4
				0.918	0.9	57	5

## Results: statistics



## **Conclusions: SPR**

- **Molecular Descriptors Family on Structure-Property Relationships obtained very good performances in estimation and prediction of compound's properties as are for example retention chromatographic index, molar refraction and water activated carbon organics adsorption.**

## Conclusions: SAR

- **Good performances are also obtained by Molecular Descriptors Family on Structure-Activity Relationships in estimation and prediction of compound's activities as are for example toxicity of alkyl metal compounds, insecticidal activity of neonicotinoid compounds, antituberculosic activity of polyhydroxyxanthenes, antioxidant efficacy of 3-indolyl derivatives, or antimalarial activity of some 2,4-diamino-6-quinazoline sulfonamide derivatives.**



# Conclusions: new compounds virtual building

http://vl.academicdirect.org/molecular\_topology/mdf\_findings/sar/ - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://vl.academicdirect.org/molecular\_topology/mdf\_findings/sar/ Links

Predict activity based on

- a learning set and
- a set of previous obtained MDF SAR models for
- any molecule submitted as HIN file by the user.

Learning set:

22583 Submit Query

22583  
23110  
23151  
23158  
23159  
23167  
26449t  
36638  
40846\_1  
40846\_2  
40846\_4

© November 2005, Lorentz JÄNTSCHI

# Conclusions

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

# Acknowledgments

- **Research was partially supported by the M. Ed. & Res., through ET/36/2005(-2008) project.**
- **AcademicDirect.Org domain is sponsored by Register.Com registrar.**
- **Servers IP's are from T.U. C-N addresses space.**

**Thank you for your attention!**