1st International Electronic Conference on Medicinal Chemistry

2-27 November 2015

chaired by Dr. Jean Jacques Vanden Eynde

sponsored by pharmaceuticals

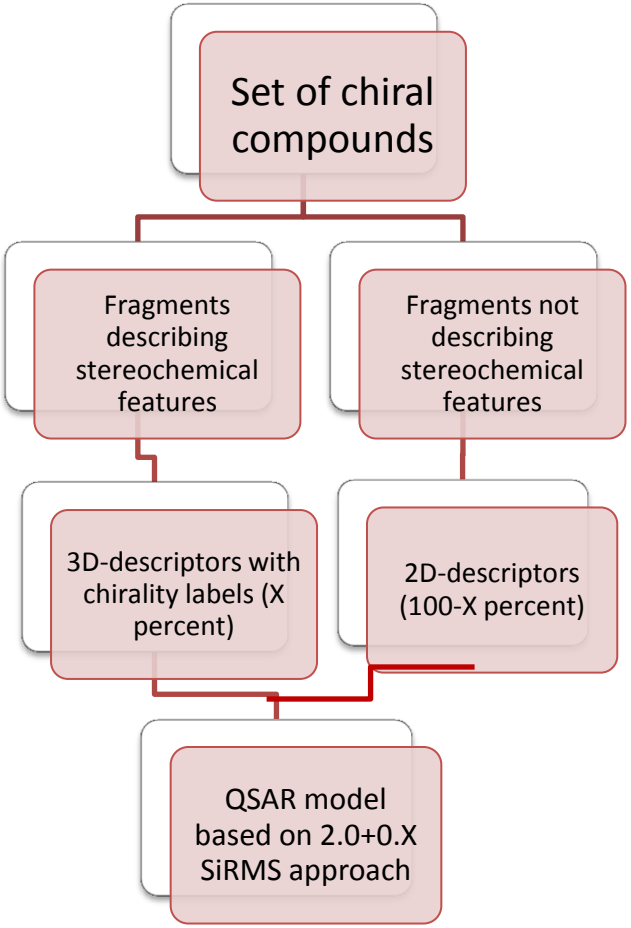
Consideration of the stereochemical features of compounds in QSAR models. 2D+0.X molecular descriptors.

Adlen Mouats [1,*], Victor E.Kuz'min[2], and Anatoliy G. Artemenko[2]

1. I.I. Mechnikov Odessa National University, Odessa, Ukraine;
2. 2. A.V. Bogatsky Physico-Chemical Institute of the National Academy of Sciences of Ukraine, Odessa, Ukraine;

* nandorua92@gmail.com

Consideration of the stereochemical features of compounds in QSAR models. 2D+0.X molecular descriptors.

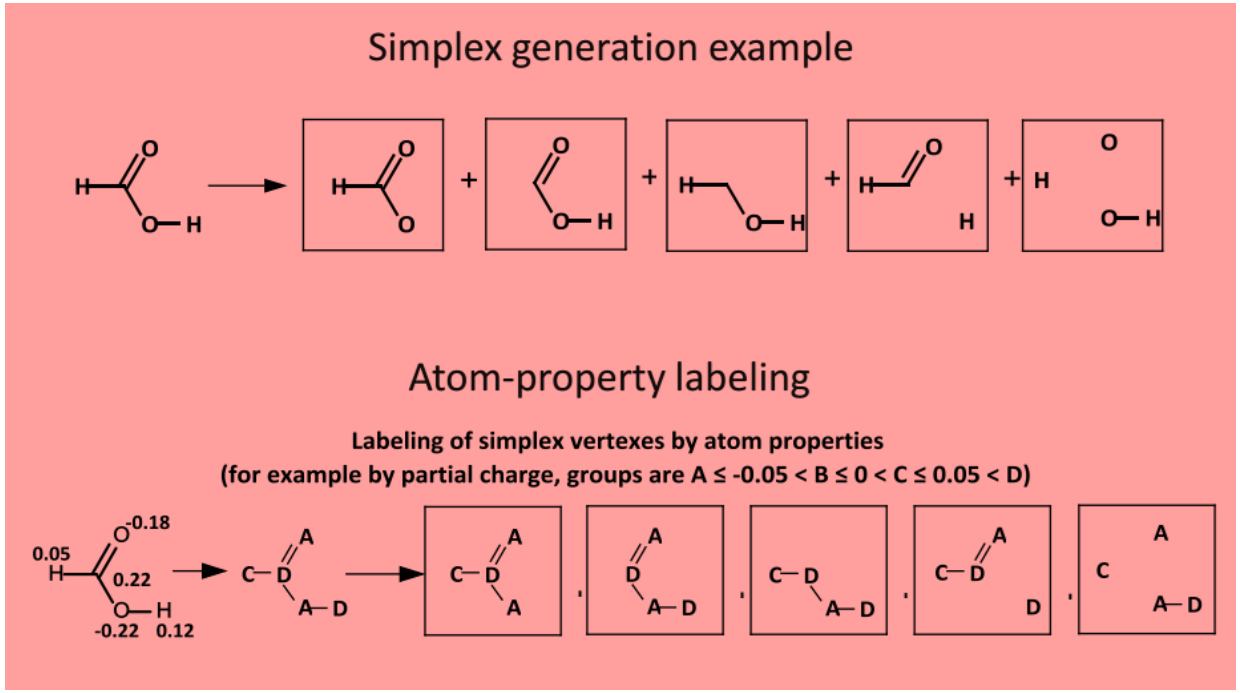


Abstract:

In chemoinformatics, stereochemical attributes are commonly taken into account only by direct description of spatial structures via 3D-QSAR approaches which are applied for one fixed conformer of each molecule. That can be undesirable if we don't know the spatial structure of the molecule interacting with a biological target. In this study we show how to solve this problem in terms of simplex representation of the molecular structure (SiRMS).

In the SiRMS approach, every molecule is represented as a system of different simplexes (tetratomic fragments with fixed composition and structure). The advantages of that approach are the absence of "molecular alignment" problems, consideration of different physical-chemical properties of atoms (e.g. charge, lipophilicity, etc.), the high adequacy and good interpretability of obtained models etc. In this study, all molecular fragments which don't determine stereochemistry of a molecule are described in terms of 2D molecular representation (structural formula). Structural elements which determine molecular stereoisomerism are described by respective 3D chiral conformation-independent simplexes It should be noted that chiral simplexes allow us to describe the molecular system of any stereochemical complexity. In the proposal (2.0+0.X)D - QSAR approach parameter (0.X) is determined by the ratio of 2D achiral and 3D chiral simplexes.

Simplex representation of molecular structure (SiRMS)



Atom properties used for labeling

- ❖Sybyl types
- ❖Atom charges
- ❖VDW-interaction descriptors
- ❖Polarizability
- ❖Lipophilicity
- ❖H-bond donor/acceptor property

‘2.X’D-SIRMS description

The approach described in this work allows to use the combination of 2D and 3D QSAR approaches. Each molecule can be divided into two parts:

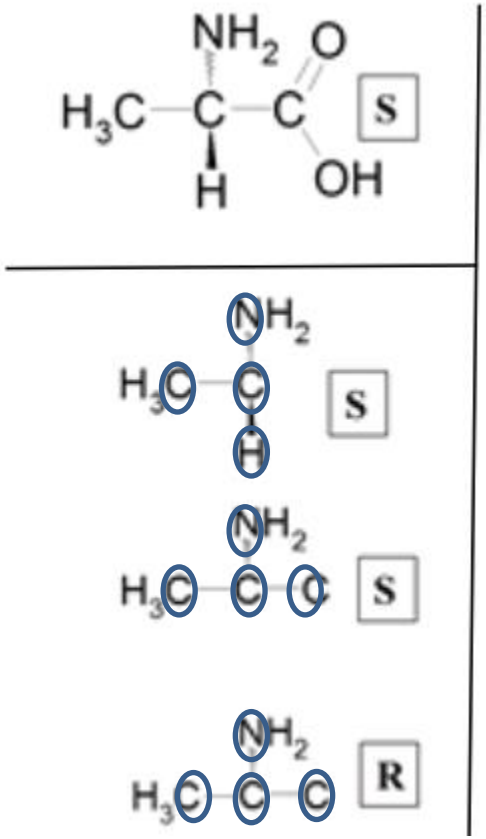
– atoms which determine stereochemical features;

– rest of the molecule.

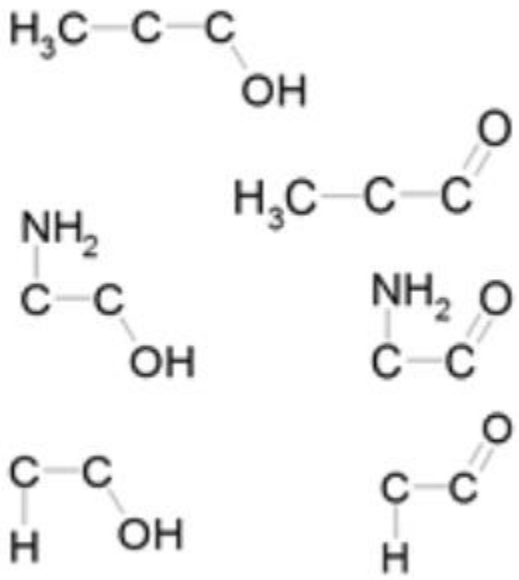
For the first group, we use <u>conformation-independent</u> simplexes with labels (R) or (S) given according to Khan-Ingold-Prelog rules. Also, in essence, all the molecular fragments that does not determine its stereochemistry, described in terms of 2D-QSAR model (structural formula).

Scheme of this approach is given in graphical abstract

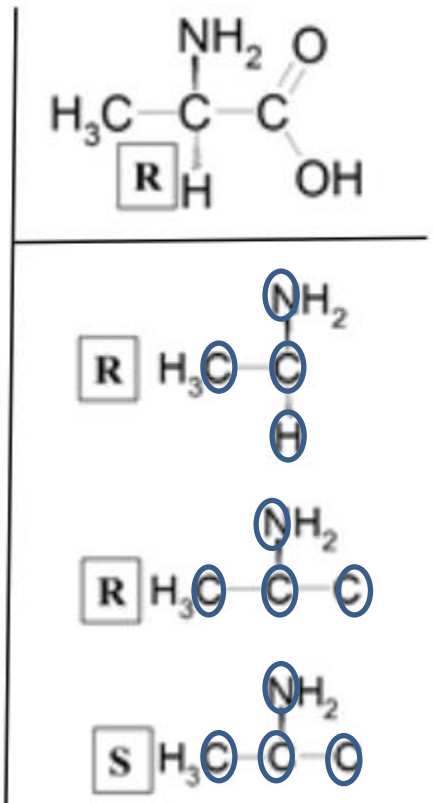
Chiral simplex generation scheme



Common 3D simplexes for S-isomer



Common 2D-simplexes



Common 3D simplexes for R-isomer

Please note that only atoms in circles are used to generate corresponding chiral simple descriptors

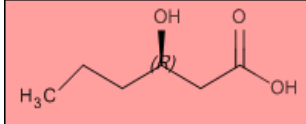
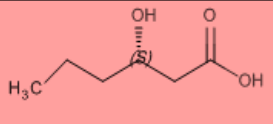
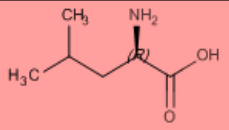
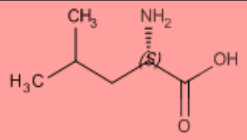
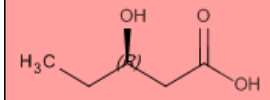
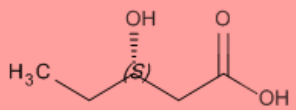
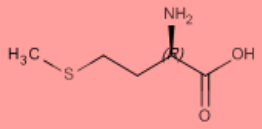
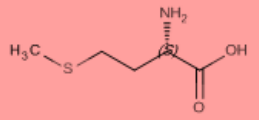
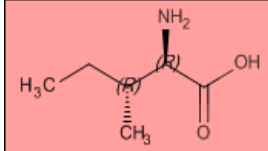
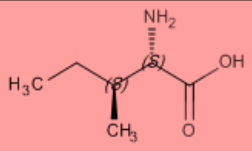
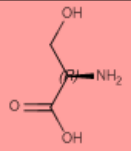
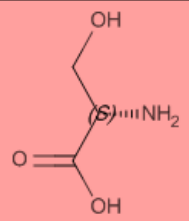
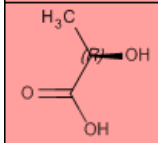
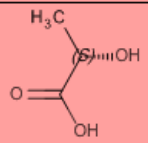
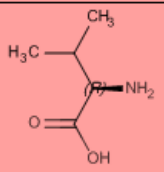
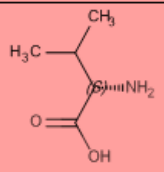
All of the QSAR-studies represented here had common scheme of the research

1. Calculation of simplex descriptors
2. Separating compounds to training and test sets based on 5-fold cross-validation
3. Calculation of QSAR models using statisticap approach (e.g. PLS, MLR etc)
4. Creation of consensus model based on data of models developed at previous step and its validation
5. Functional and/or structural interpretation of the consensus model

To evaluate our approach, we have solved five different QSAR-tasks.

1. Structure-chromatographic retention for enantiomers
2. Structure – CBG affinity for Kramer steroids
3. Structure- CCR2 affinity for CCR2 antagonists
4. Structure-drosophila BII cell line for ecdysteroids
5. Structure-antimalarial activity for naphtylisoquinoline alcaloids

Structure-chromatographic retention [1] for enantiomers

1R	1S	5R	5S
			
2R	**2S**	**6R**	**6S**
			
3R	**3S**	**7R**	**7S**
			
4R	**4S**	**8R**	**8S**
			

We used this relatively simple dataset to evaluate if this approach can separate compounds which differ only at 3D-level. The results were satisfying (see next slide)

Task 1

Statistical characteristics of the obtained model

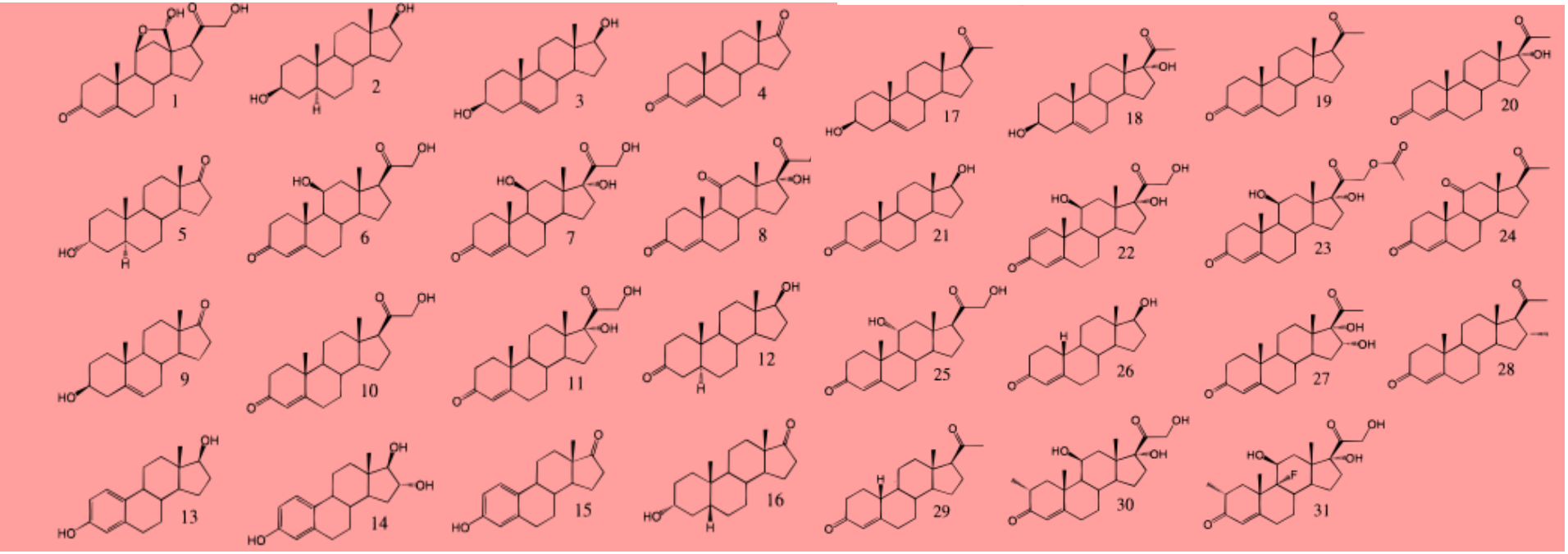
R^2	0.97
Q^2	0.95
RMSE	0.04

Here and further R^2 is for the coefficient of determination (R^2 ts is for the coeffisient of determination of test set), Q^2 is for the cross-validation coefficient of determination and RMSE for root mean square error

Observed vs Predicted data

Name	Set	(Obs.)	(Pred.)	(Resid.)
1R	ws	-0.3500	-0.3353	-0.0147
2R	ws	-0.3300	-0.3360	0.0060
3R	ws	-0.2300	-0.2347	0.0047
4R	ws	-0.5000	-0.4902	-0.0098
5R	ws	-0.2700	-0.2076	-0.0624
6R	ws	-0.1900	-0.1833	-0.0067
7R	ws	-0.5500	-0.5590	0.0090
8R	ws	-0.2300	-0.2296	-0.0004
1S	ws	-0.2100	-0.1959	-0.0141
2S	ws	-0.1700	-0.1959	0.0259
3S	ws	-0.0200	-0.0016	-0.0184
4S	ws	-0.3700	-0.3782	0.0082
5S	ws	-0.0300	-0.0957	0.0657
6S	ws	-0.0900	-0.0714	-0.0186
7S	ws	-0.4300	-0.4378	0.0078
8S	ws	-0.1000	-0.1177	0.0177

Structure – CBG affinity for Kramer steroids



Set of 31 steroid structures described by Kramer is often used as a benchmark of descriptional approaches for 3D-QSAR because of wide range of structural differences as well as range of activity. That's why it was necessary to use this set to validate our approach as well

Statistical characteristics of '2.X'D-SiRMS Models (statistical method – PLS)

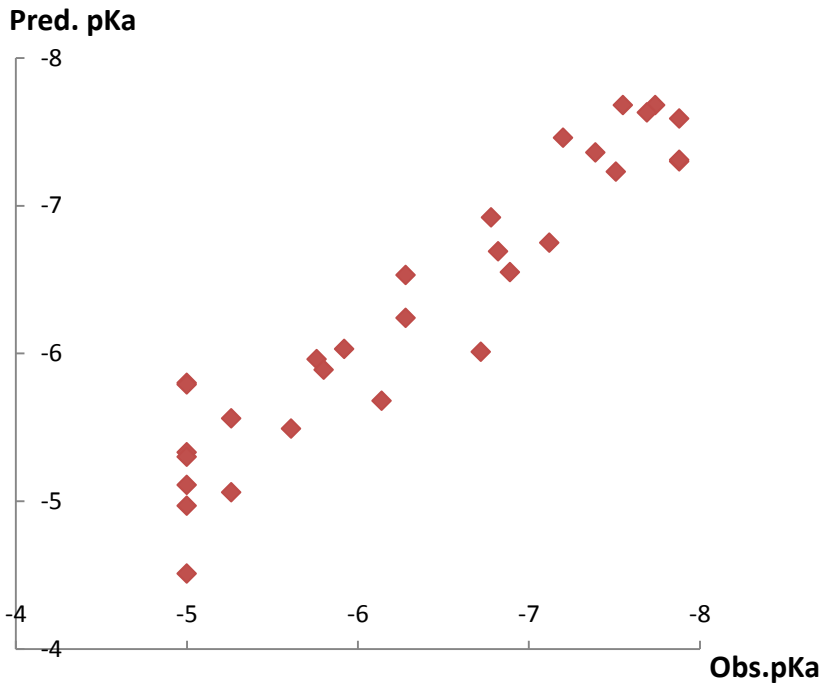
Model	R^2	Q^2	R^2ts	RMSE
1	0.86	0.68	0.85	0.62
2	0.85	0.77	0.89	0.52
3	0.85	0.73	0.87	0.58
4	0.86	0.76	0.78	0.55
5	0.85	0.78	0.90	0.49
Consensus	0.87	0.79	0.84	0.51

Comparison of some 3D-QSAR researches for this set

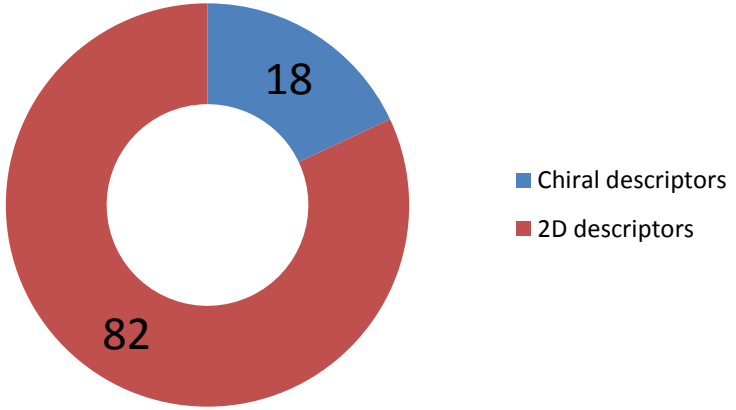
Descriptors	Statistical method	Q^2	Source
Similarity matrices	GA+ANN	0.94	[2]
TOMOCOMD-bilinear indices	MLR	0.83	[3]
MEDV	MLR+GA	0.77	[4]
TQSI	MLR	0.76	[5]
CoMSIA	PLS	0.73	[6]

The only model that showed significantly higher Q^2 is similarity-matrices based. We suggest that its' results are higher because ANN often fits great for models based on matrices . So our approach shows reliable results compared to most 3D-QSAR models

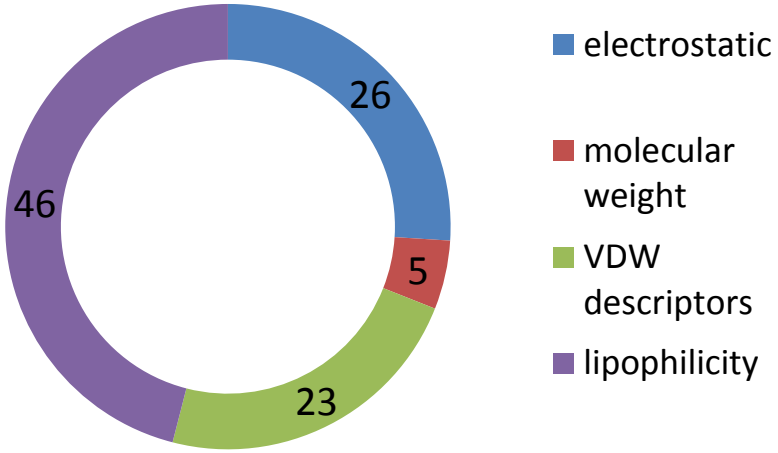
Observed vs Predicted diagram



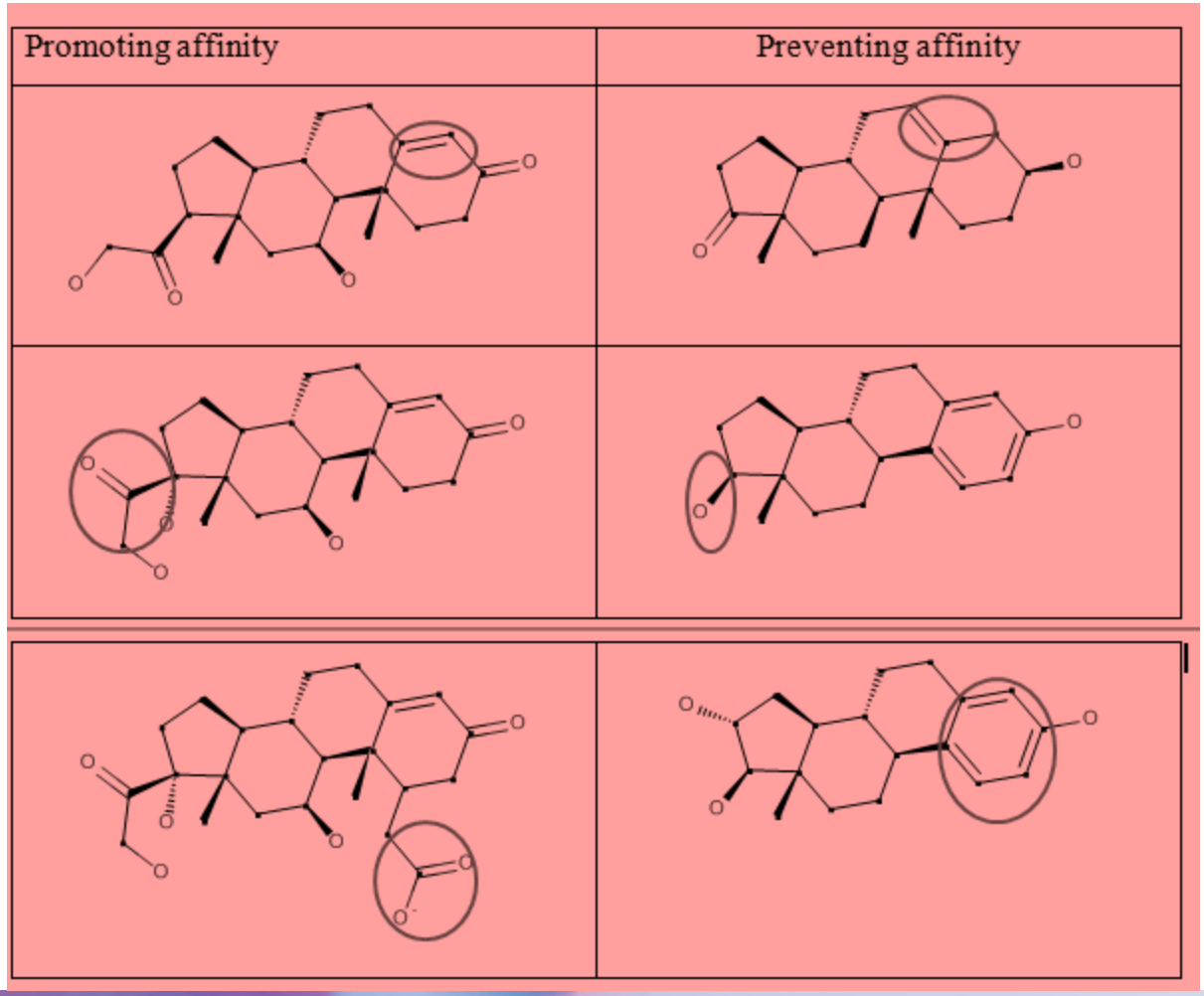
Relative influence of different descriptors to consensus model



Relative influence of 2D-descriptors

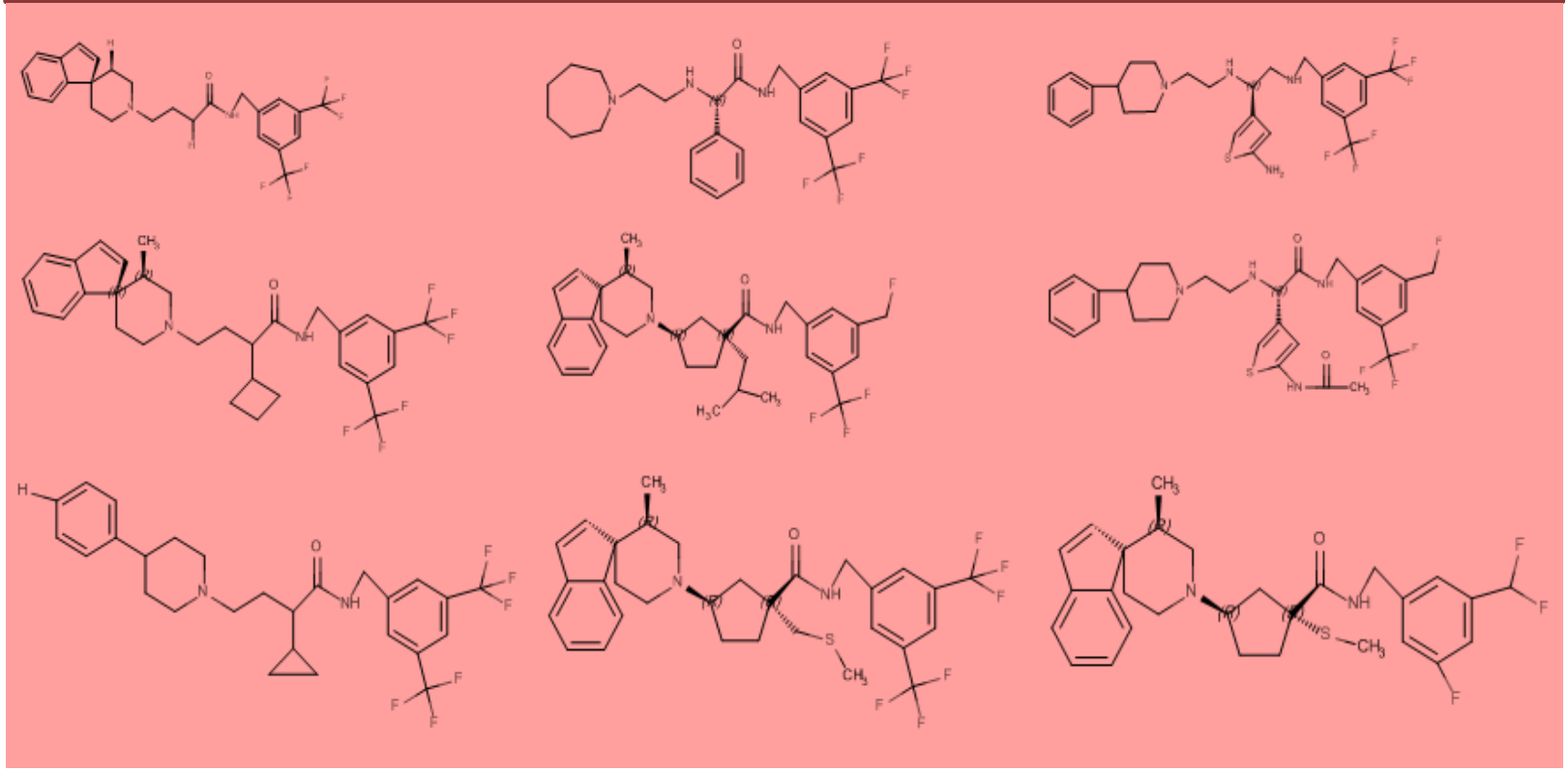


Influence of different molecular fragments



Structure-affinity for CCR2 antagonists

Examples of compounds used for training



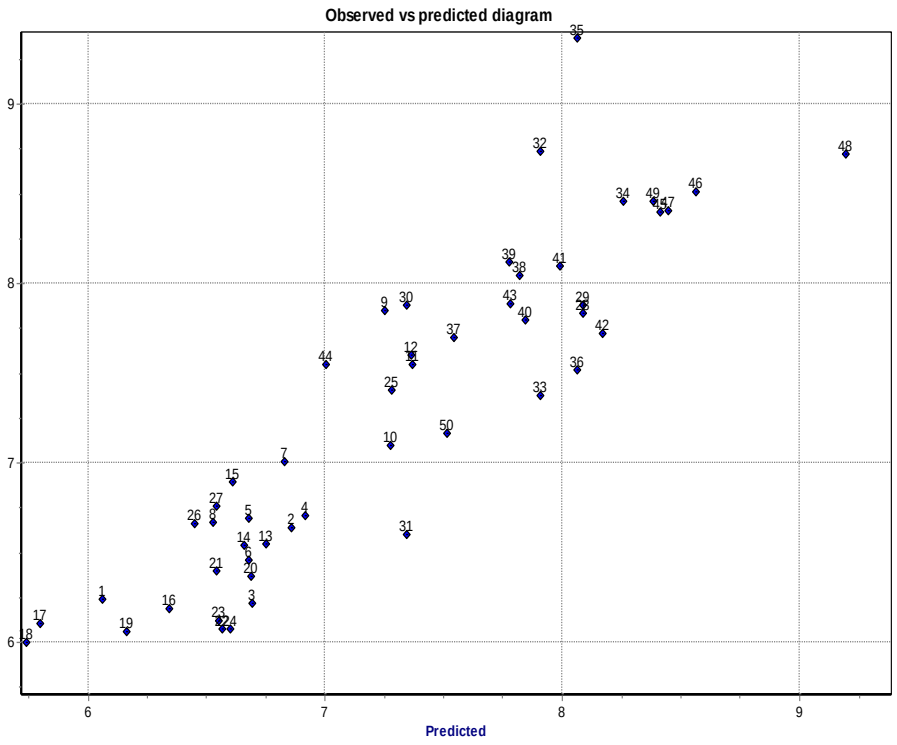
This set was selected for research as containing both chiral and achiral compounds. It was previously researched by HQSAR approach. [7]

Statistical characteristics of '2.X'D-SiRMS Models (statistical method – PLS)

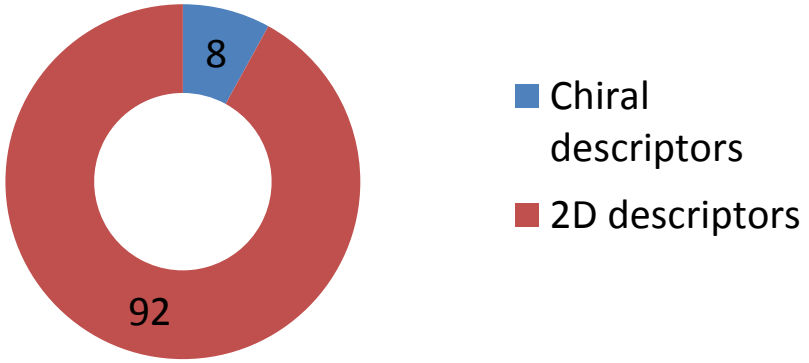
Model	R^2	Q^2	R^2ts	RMSE
2D-SiRMS	0.84	0.80	0.78	0.37
'2.X D'-SiRMS	0.88	0.83	0.81	0.29
HQSAR[7]	0.94	0.84	0.80	0.47

These results show that using of chiral descriptors allows to boost statistical parametres for the models and describe given structures better, so we cannot ignore this data even though it has relatively low influence (as shown is slide below).
Also it shows similar efficiency of '2.X'D-SIRMS approach compared to Hologram QSAR.

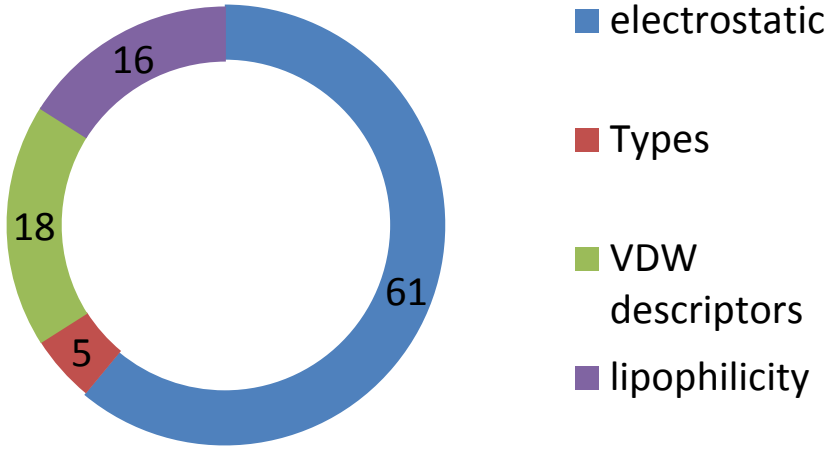
Observed vs Predicted diagram for consensus model



Relative influence of different descriptors for consensus model

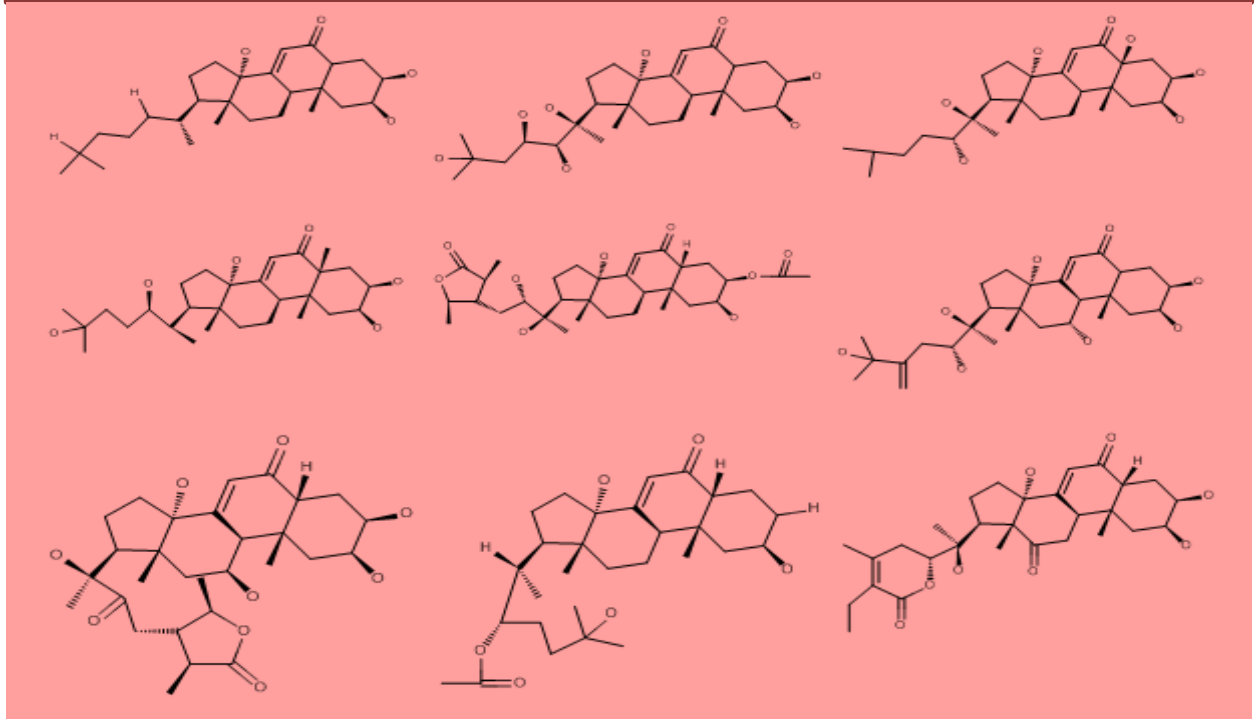


Relative influence of 2D-descriptors



Structure-drosophila BII cell line for ecdysteroids

Examples of compounds used for training



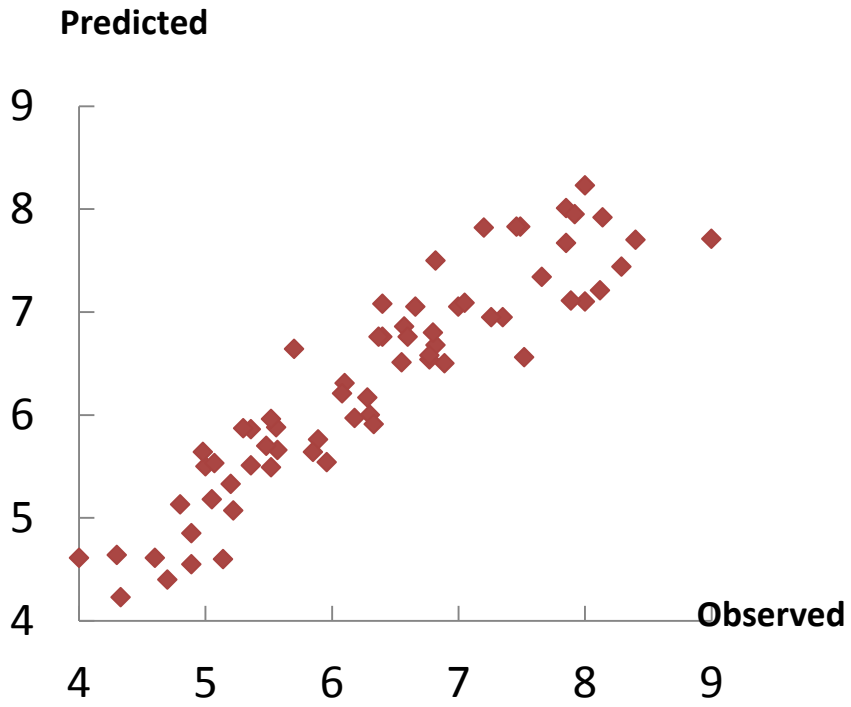
Compounds used in this set were previously studied via CoMFA approach [8]. This set was selected as containing compounds with multiple chiral centers

Statistical characteristics of '2.X'D-SiRMS Models (statistical method – PLS)

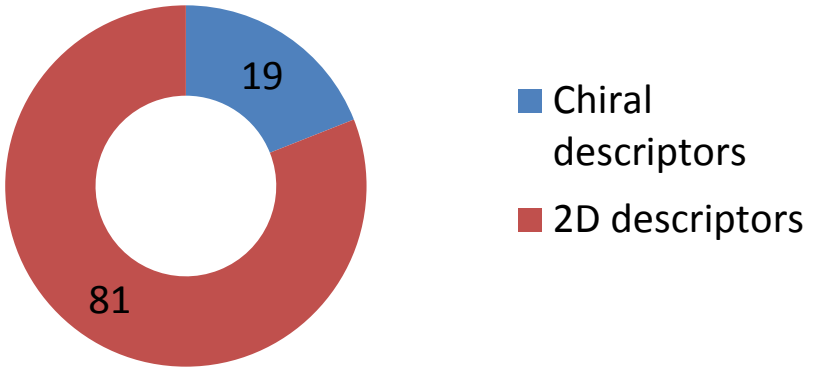
Model	R^2	Q^2	R^2ts	RMSE
1	0.83	0.70	0.76	0.49
2	0.84	0.70	0.84	0.49
3	0.87	0.79	0.71	0.42
4	0.82	0.72	0.88	0.52
5	0.86	0.76	0.84	0.47
Consensus	0.88	0.79	0.78	0.44
CoMFA(PLS)	0.89	0.69	0.39	0.44
Golbraikh descriptors (kNN)[9]	N/A	0.61	0.89	0.42

Again, '2.X'D-SIRMS model shows comparible results to those obtained via 3D-approach, and, in terms of cross-validation, even exceeds them.
NB: there were 4 outliers as well as in CoMFA study.

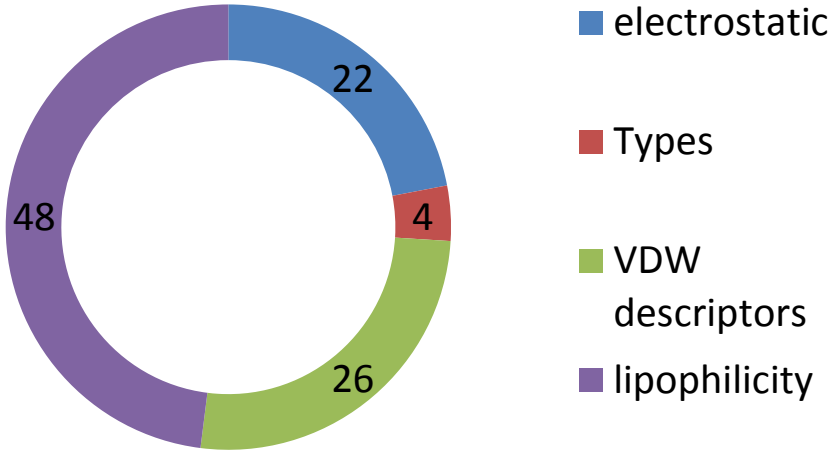
Observed vs Predicted Diagram for –logED50 for consensus model



Relative influence of different descriptors for consensus model



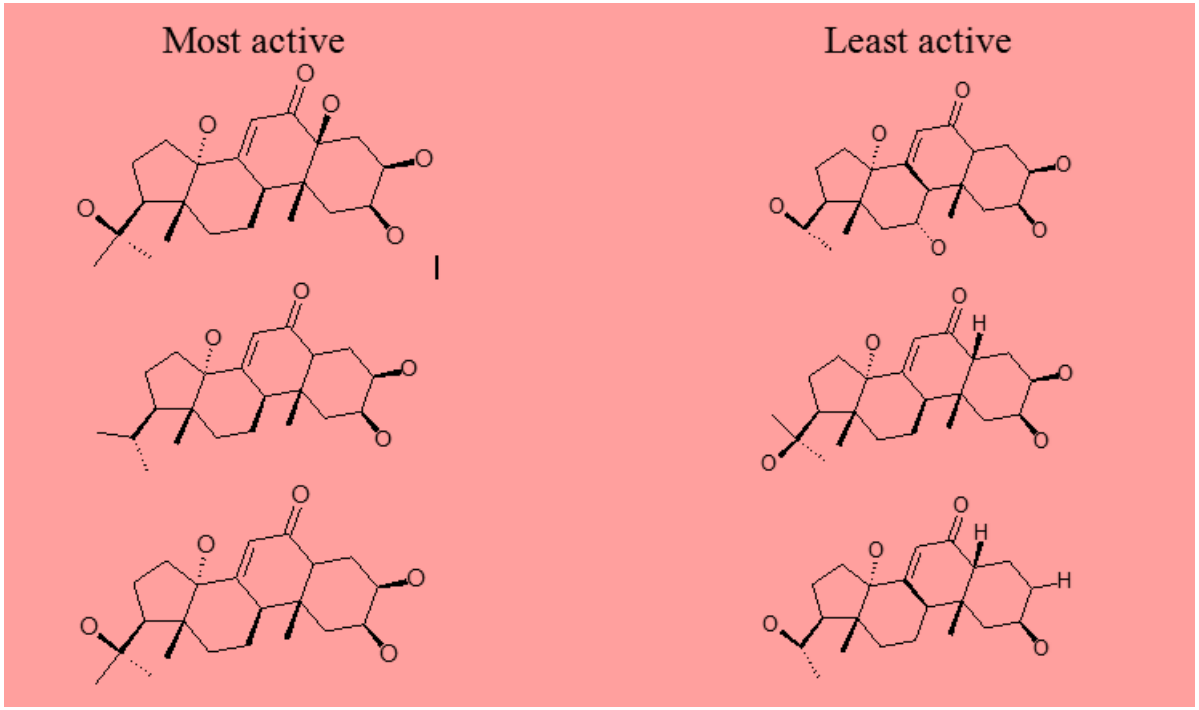
Relative influence of 2D-descriptors



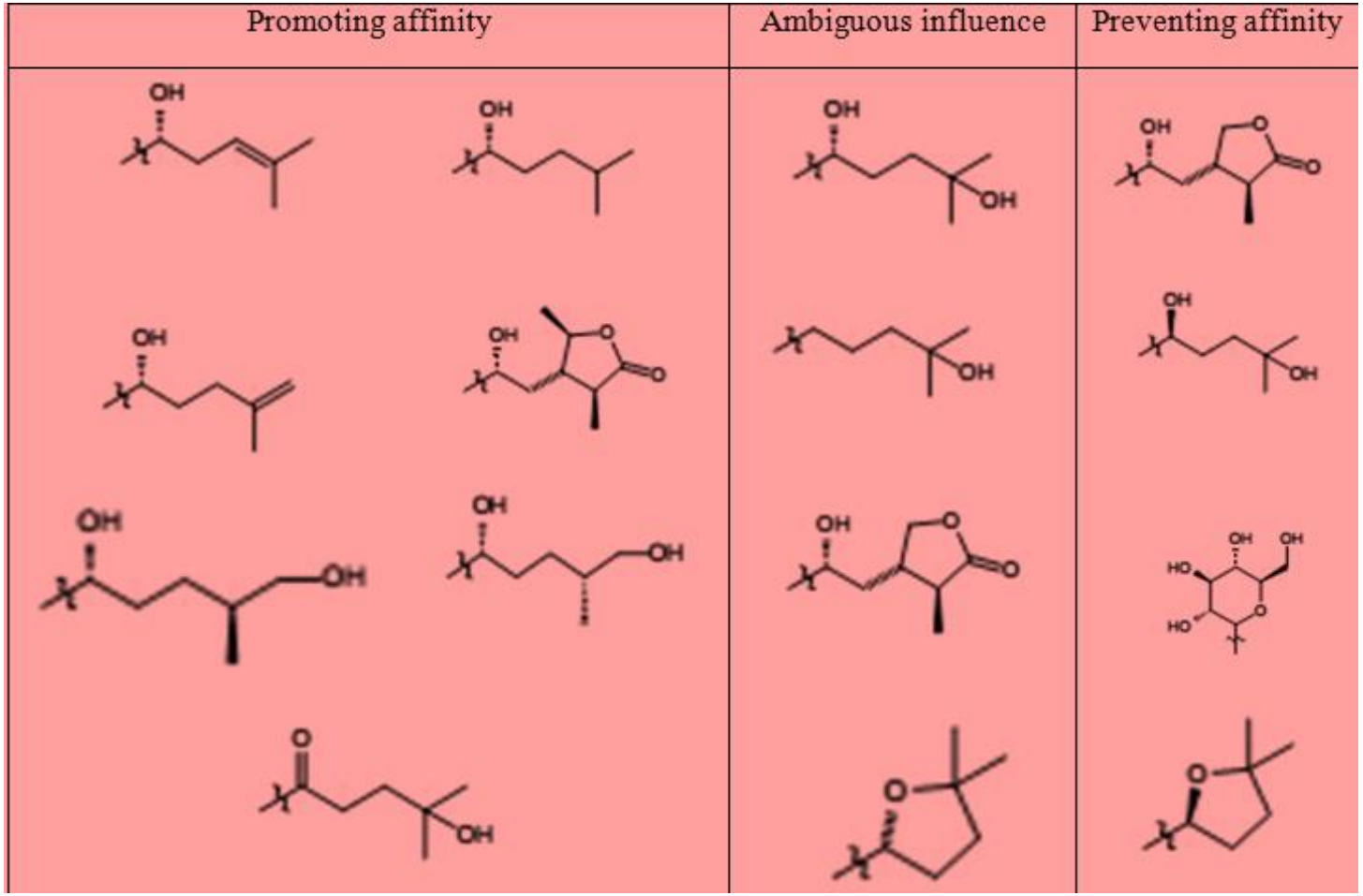
Structural interpretation of obtained data

Simplex descriptors allow us to find structural fragments which prevent
Or, to the contrary, promote studied ability. For this model we separated fragments
Into two groups – to study influence of different molecular scaffolds and different substituents

a) Influence of the scaffolds

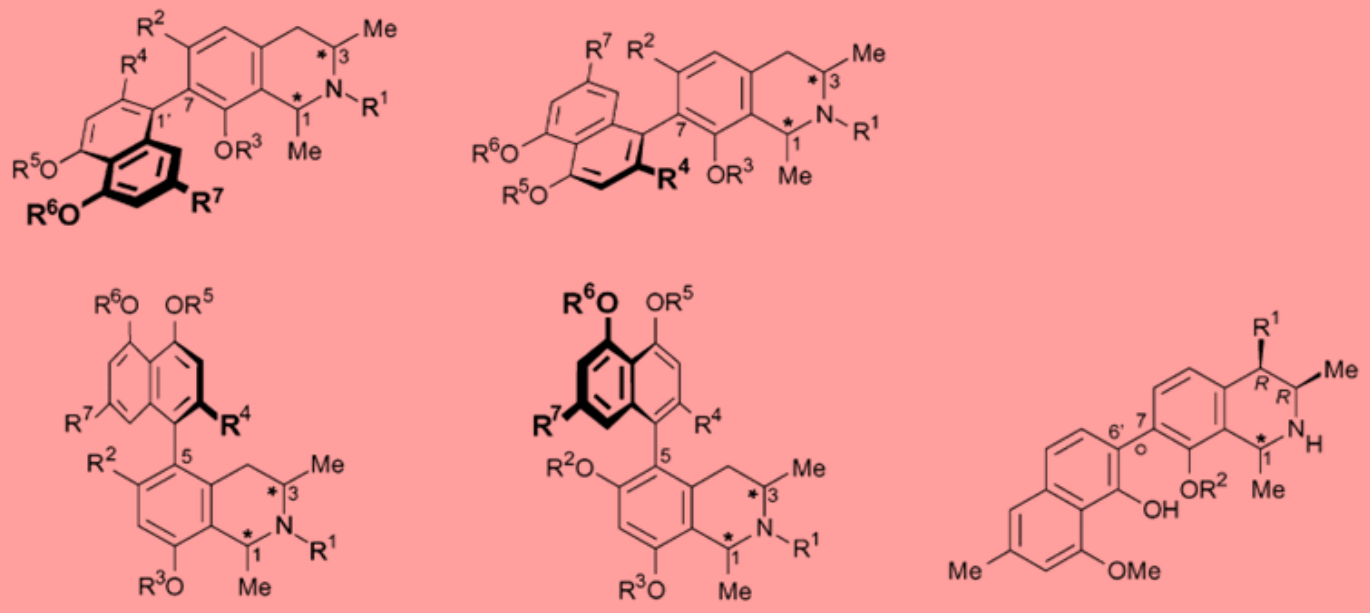


b)Influence of different substituent groups



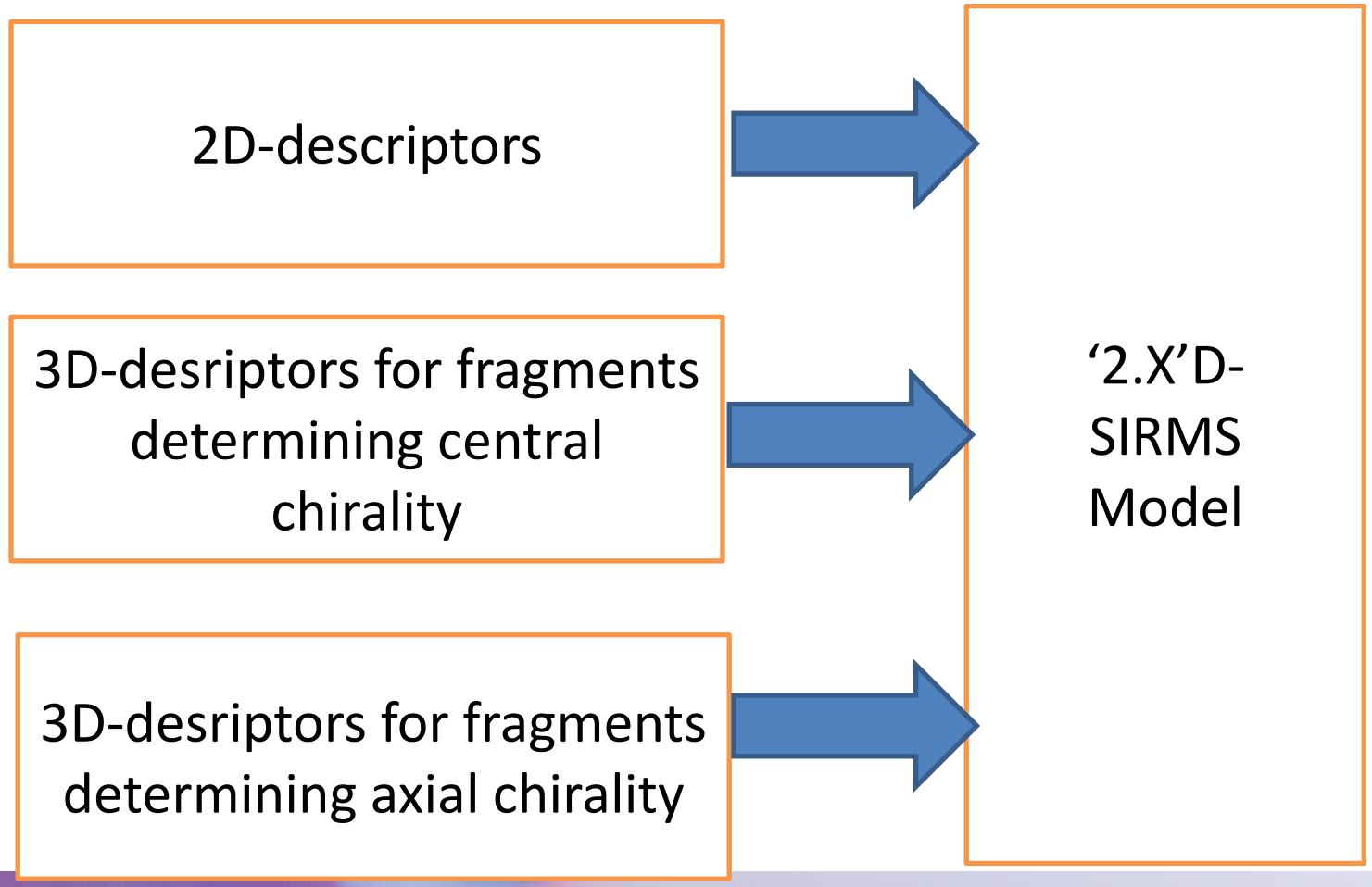
Structure-antimalarial activity for 45 naphtylisoquinoline alcaloids

Examples of the studied compounds



This set was previously studied by Bringmann et al. Via CoMSIA approach[10]. We included it into our study because there are compounds containing two types of stereoisomery – compounds with central and axial chirality

For this model, we had to modify our approach to include all chiral data

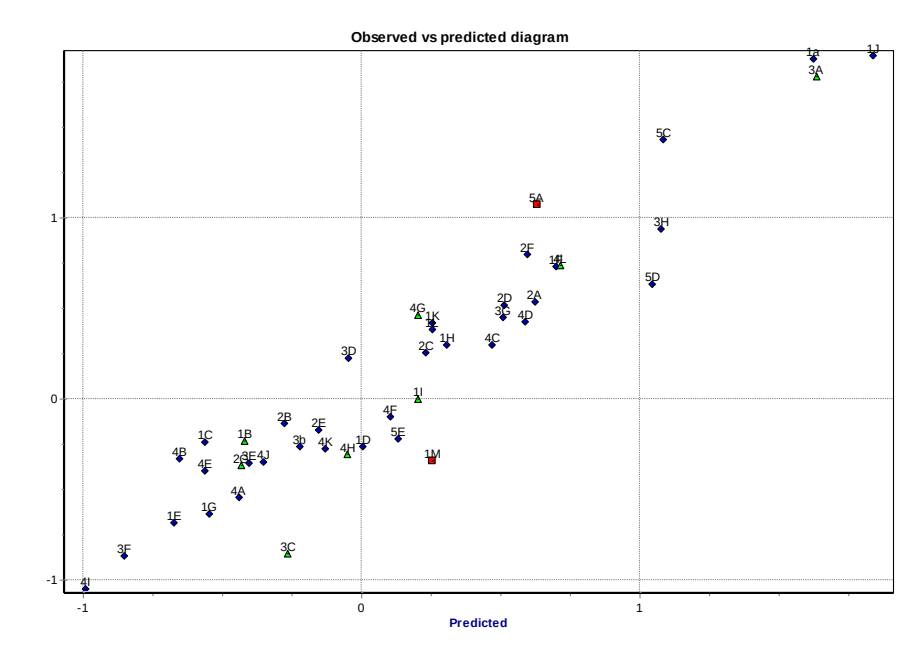


Statistical characteristics of '2.X'D-SiRMS Models (statistical method – PLS)

Model	R^2	Q^2	R^2ts	RMSE
1	0.89	0.81	0.81	0.3
2	0.86	0.75	0.73	0.36
3	0.90	0.82	0.87	0.29
4	0.87	0.8	0.82	0.32
5	0.77	0.68	0.84	0.39
SiRMS Consensus	0.9	0.78	0.82	0.31

We used only part of the training set used in [10] so it would be incorrect to compare results. However, their models on set including those alkaloids studied by CoMSIA are: Q^2=0.82, RMSE=0,67

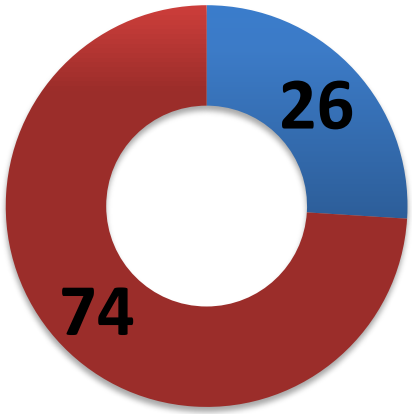
Observed vs Predicted diagram for consensus model

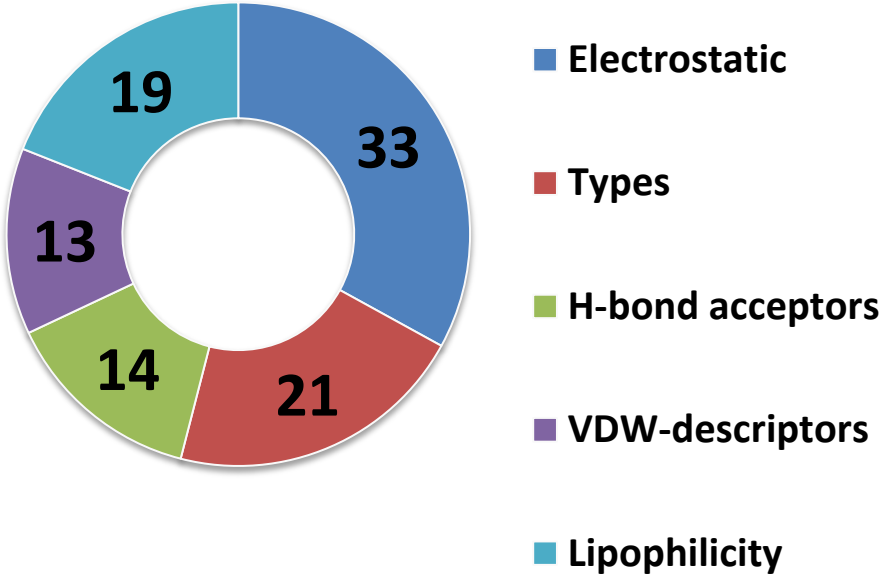


Example of separation for couples of atropoisomeres (1A and 2A, 1K and 2D, respectively)

	-lgIC50(exp)	-lgIC50(pred)		-lgIC50(exp)	-lgIC50(pred)
1A	1.8	1.62	2A	0.53	0.62
1K	0.41	0.29	2D	0.51	0.5

Relative influence of different descriptors to consensus model





Only 3D descriptors determining axial chirality were selected via QSAR processing

Relative influence of 2D-descriptors

Models Summary

Property	Compounds	Chirality type	Quality of our models		Level of our models	Best model by different approach		
			Q^2	R^2ts		Q^2	R^2ts	Source
Chromatographic retention	Hydroxy acids and amino acids(16)	Central	0.95	-	2.76D	N/A	N/A	[1]
CBG affinity	Steroids (31, Kramer set)	Central	0.79	0.84	2.18D	0.83	N/A	[2-6]
CCR2 affinity	CCR2 antagonists(50)	Central + achiral	0.83	0.81	2.08D	0.84	0.80	[7]
Drosophila BII cell line for ecdysone receptor	Ecdysteroids(71)	Central (multiple centers)	0.79	0.78	2.19D	0.69	N/A	[8-9]
Antimalarial Activity	Naphtylisoquinoline alkaloids (45)	Central+ axial	0.78	0.82	2.26D	N/A	N/A	[10]

Conclusions

- ✓ **Results obtained in this study show that '2.X'D-SiRMS" approach can be equally or even more efficient than 3D-QSAR approaches.**
- ✓ **Also this approach helps to get models for compounds with specifical stereochemical features (e.g. atropoisomers) as well as for enantiomers.**
- ✓ **It allows to get structural and functional interpretation what can be useful for further researches of these properties and compounds**

References

1. I.Lukovits, W.Linert. A topological account of chirality. J. Chem. Inf. Comput. Sci. 2001, #41, P.1517-1520
2. S-S.So, M,Karplus. Three-Dimensional Quantitative Structure-Activity Relationships from Molecular Similarity Matrices and Genetic Neural Networks. 1. Method and Validations. J. Med. Chem.1997,40,4347-4359
3. Y..Marreno-Ponce et al. 3D-chiral (2.5) atom-based TOMOCOMD-CARDD descriptors: theory and QSAR applications to central chirality codification.J Math Chem (2008) 44:755–786
4. S.S. Liu, C.S. Yin, L.S. Wang, Combined MEDV-GA-MLR method for QSAR of three panels of steroids, dipeptides, and COX-2 inhibitors. J. Chem. Inf. Comput. Sci.42, 749–756 (2002)
5. M. Lobato, L. Amat, E. Besalu, R. Carbo-Dorca. Structure-activity relationships of a steroid family using quantum similarity measures and topological quantum similarity indices.Quant. Struct.-Act. Relat.16, 465–472 (1997)
6. M.F. Parretti, R.T. Kroemer, J.H. Rothman, W.G. Richards Alignment of molecules by the Monte Carlo optimization of molecular similarity indices. J. Comput. Chem.18, 1334–1353 (1997)
7. P.C.Nair, K.Srikanth, M.E.Sobhia. QSAR studies on CCR2 antagonists with chiral sensitive hologram descriptors. Bioorg. Med. Chem. Lett. 2008. #18. P.1323–1330
8. L.Dinan, R.E.Hormann, T.Fujimoto. An extensive ecdysteroid CoMFA. Journal of Computer-Aided Molecular Design, 1999, #13, P185–207.
9. A.Golbraikh, D.Bonchev, A.Tropsha. Novel Chirality Descriptors Derived from Molecular Topology/J. Chem. Inf. Comput. Sci., Vol. 41, No. 1, 2001
10. G.Bringmann, C.Runney. 3D QSAR Investigations on Antimalarial Naphthylisoquinoline Alkaloids by Comparative Molecular Similarity Indices Analysis (CoMSIA), Based on Different Alignment Approaches //J. Chem. Inf. Comput. Sci.2003,43,304-316