

Structural and functional interpretation of QSAR models

P.G. Polishchuk¹, T.M. Khristova^{1,2}, L.N. Ognichenko¹,

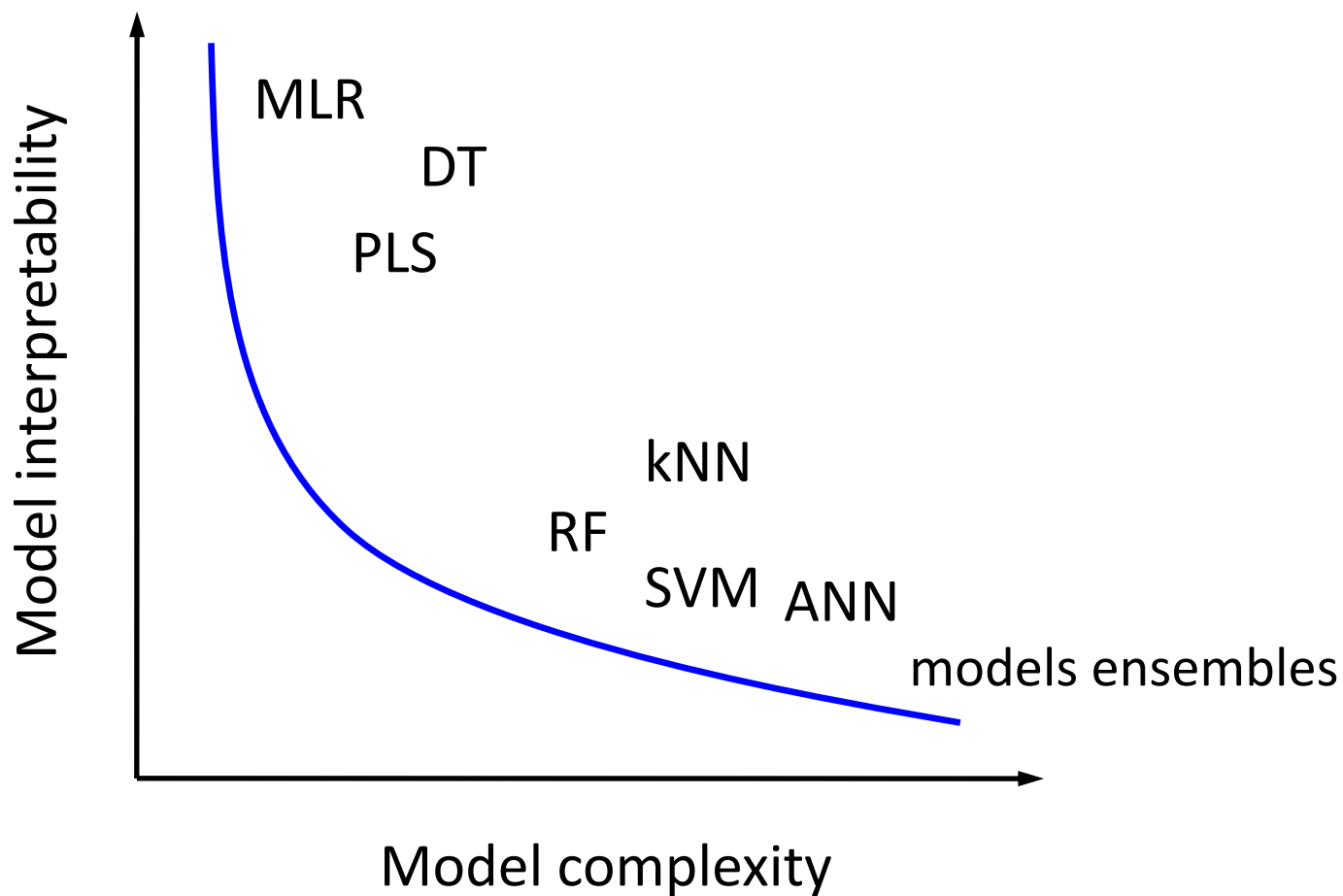
A.P. Kosinskaya¹, A.A. Varnek², V.E. Kuz'min¹

¹ A.V. Bogatsky Physico-Chemical Institute of NAS of Ukraine, Odessa, Ukraine

² Strasbourg University, Strasbourg, France

QSAR interpretation: interpretability vs. complexity

Popular misbelief



Model-specific approaches:

Rule-based (Decision tree)

Regression coefficients (MLR, PLS)

Latent variables (PLS)

Weights and biases (ANN)

...

Model-independent approaches:

Local gradients or partial derivatives

$$C_i = \frac{f(x_i) - f(x_i + \Delta x_i)}{\Delta x_i}$$

QSAR interpretation: common workflow

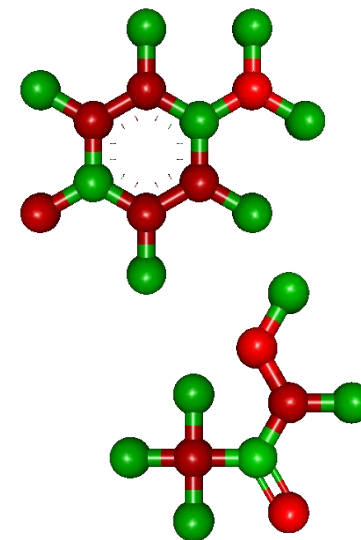
Model

Variables
contributions

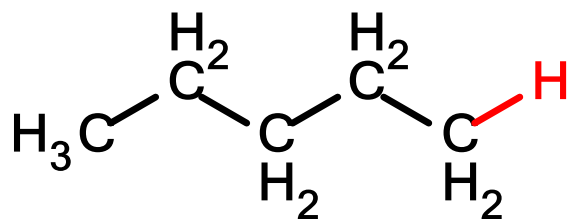
Structure-
property
relationship

$f(x)$

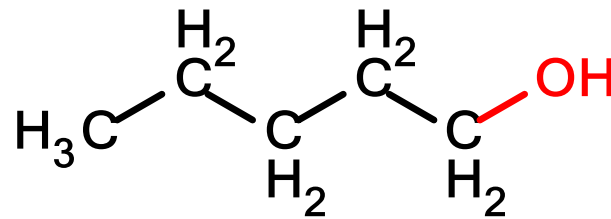
	Var_1	Var_2
Mol_1	-0.23	1.82
Mol_2	2.36	1.27
Mol_3	5.01	2.30
Mol_4	0.69	-0.58



Matched molecular pairs & molecular transformations



$$\log S = -3.18$$

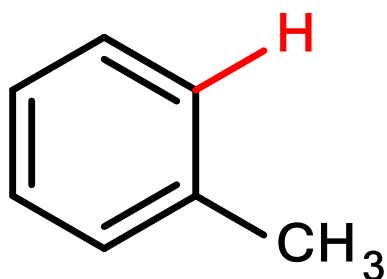


$$\log S = -0.60$$

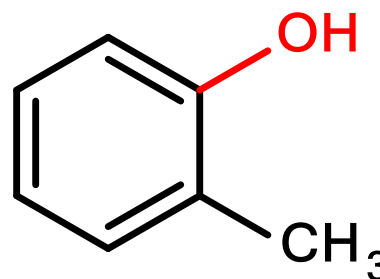
$$\Delta \log S = 2.58$$



$$\Delta \log S = 1.59$$

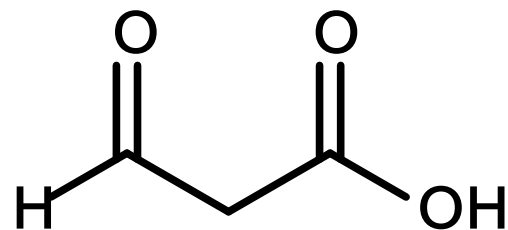
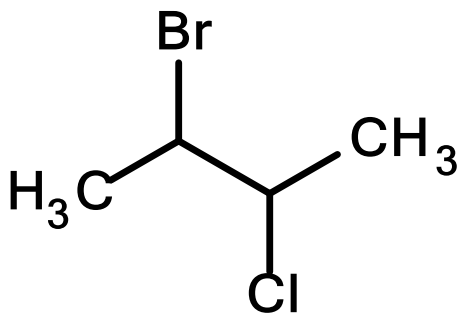
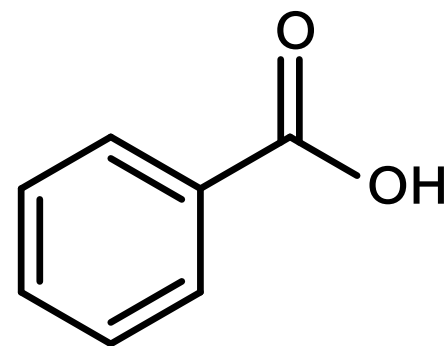
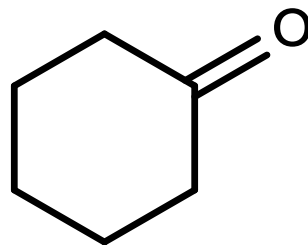
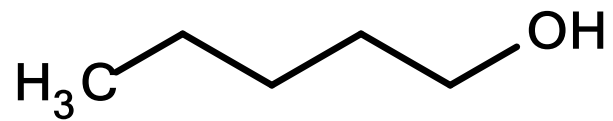
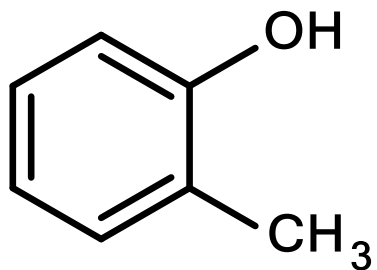


$$\log S = -2.21$$

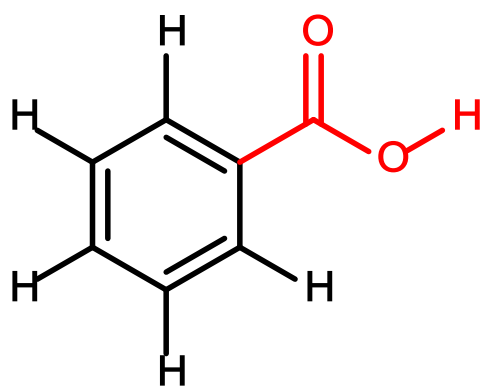


$$\log S = -0.62$$

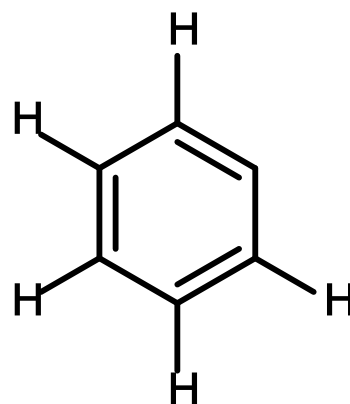
Exemplified dataset



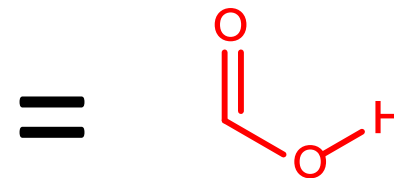
Structural QSAR interpretation



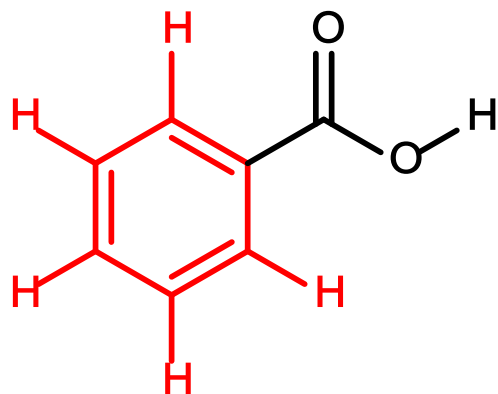
$$\log S_{\text{pred}} = -1.55$$



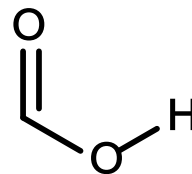
$$\log S_{\text{pred}} = -1.61$$



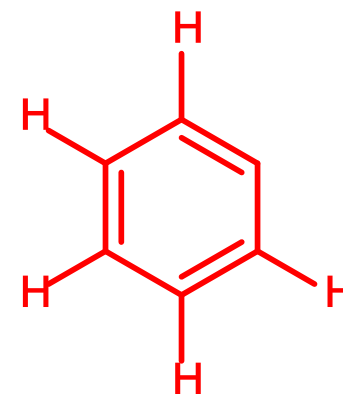
$$\Delta \log S_{\text{pred}} = 0.06$$



$$\log S_{\text{pred}} = -1.55$$

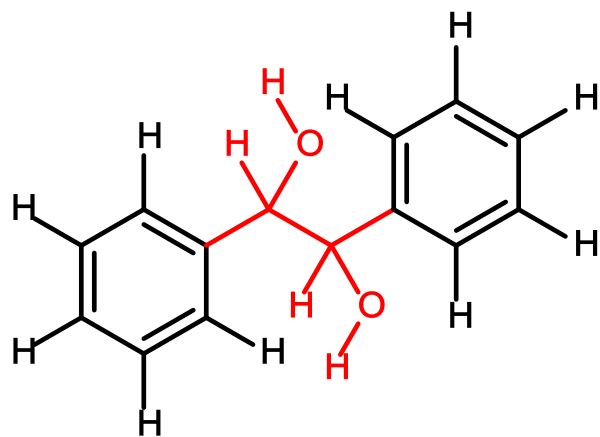


$$\log S_{\text{pred}} = -1.35$$



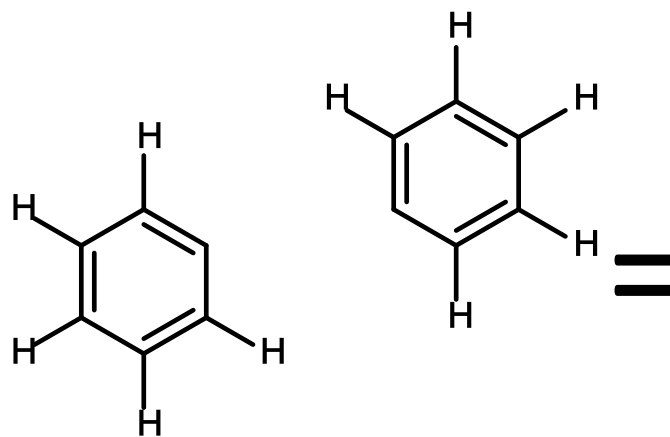
$$\Delta \log S_{\text{pred}} = -0.20$$

Structural QSAR interpretation



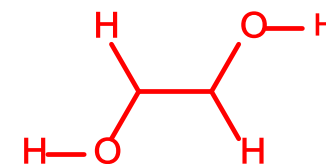
$$\log S_{\text{pred}} = -1.93$$

-



$$\log S_{\text{pred}} = -4.32$$

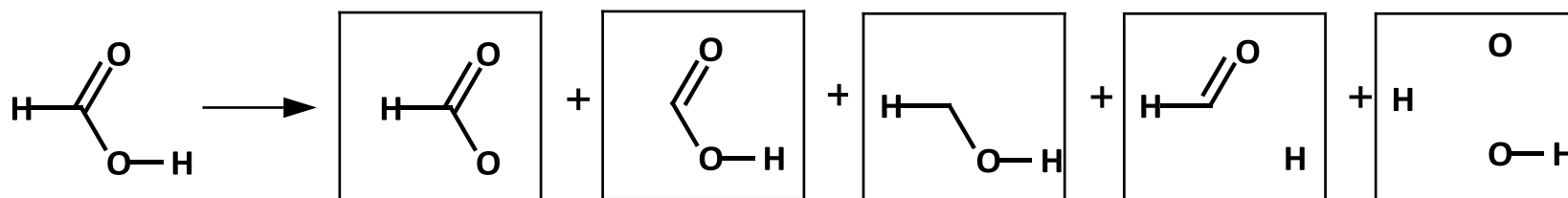
=



$$\Delta \log S_{\text{pred}} = 2.39$$

Simplex representation of molecular structure (SiRMS)

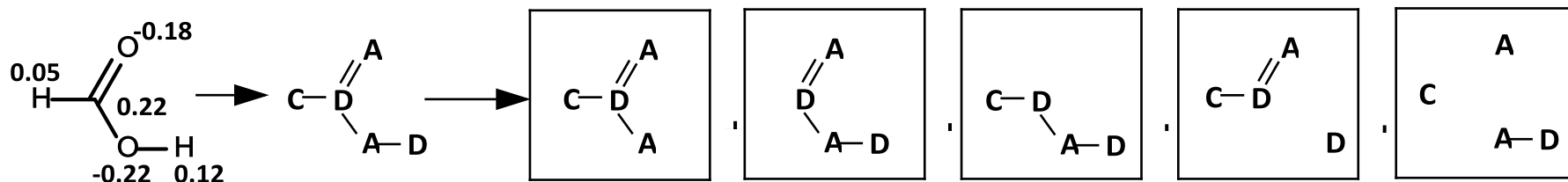
Simplex generation example



Atom-property labeling

Labeling of simplex vertexes by atom properties

(for example by partial charge, groups are $A \leq -0.05 < B \leq 0 < C \leq 0.05 < D$)

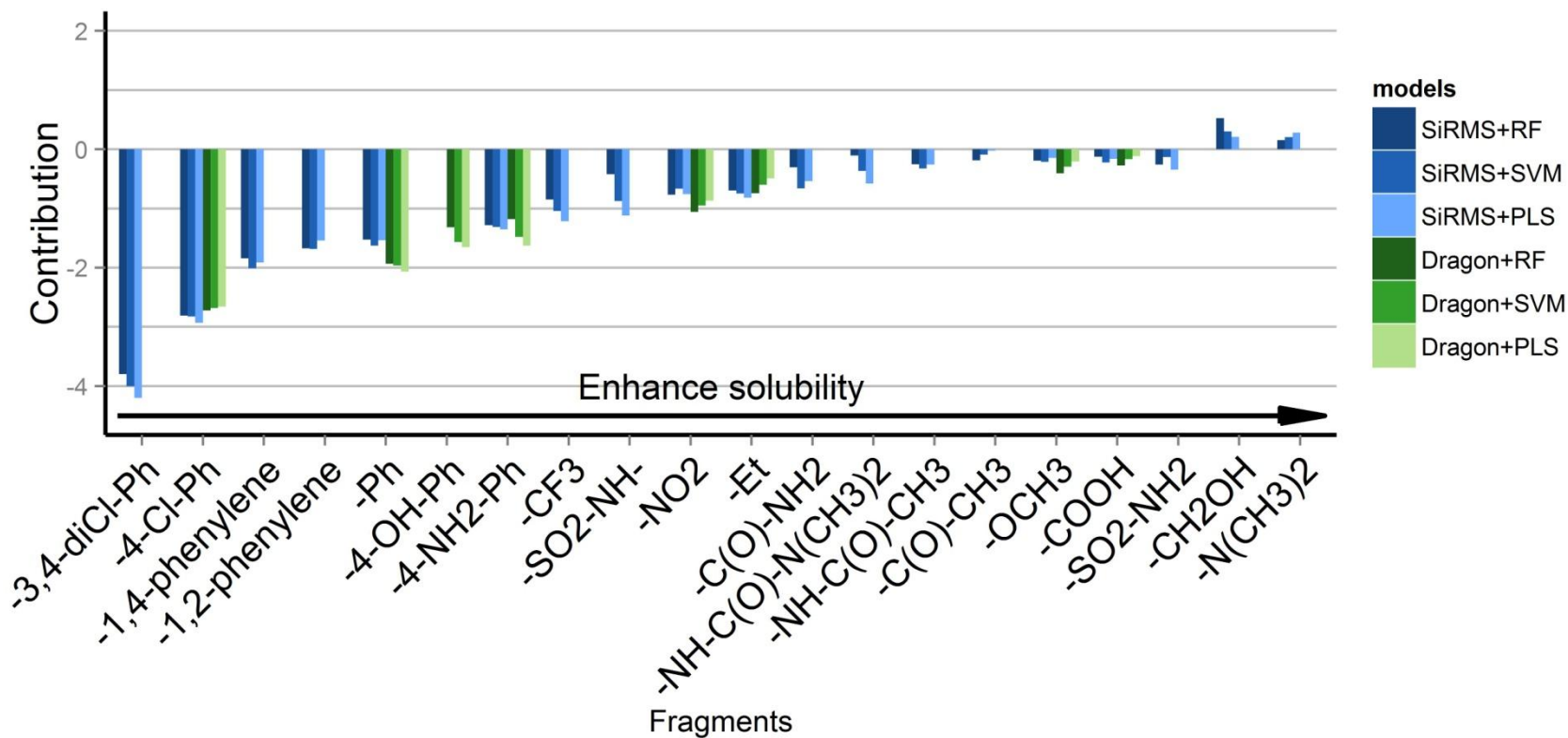


Examples of structural interpretation

Solubility (1033 compounds)

5-fold external cross validation results

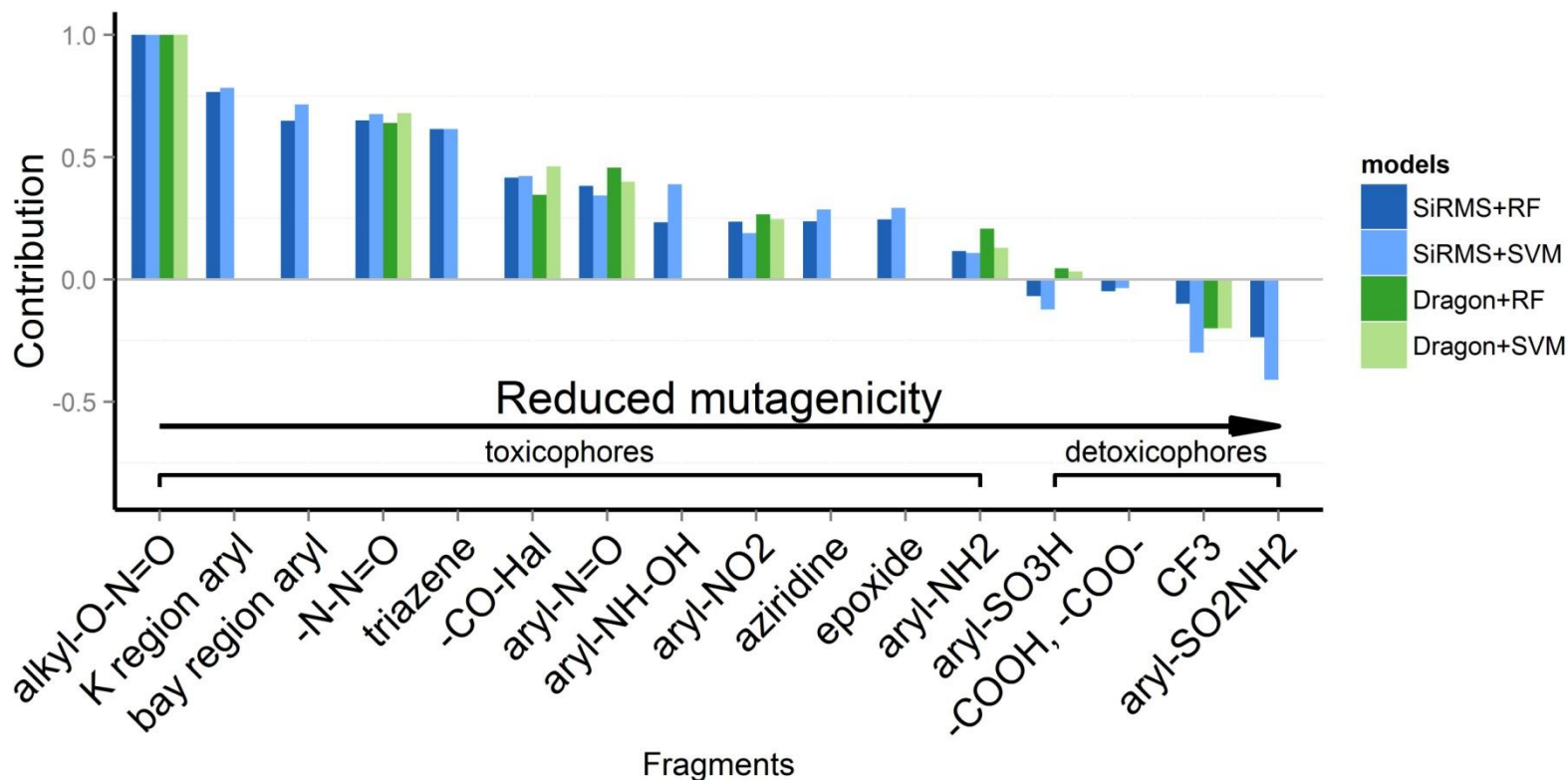
Endpoint	Model	SiRMS		Dragon	
		R^2_{cv}	RMSE	R^2_{cv}	RMSE
Solubility, logS	PLS	0.84	0.82	0.91	0.60
	RF	0.88	0.71	0.91	0.62
	SVM	0.87	0.72	0.92	0.59



Mutagenicity (Ames, 4361 compounds)

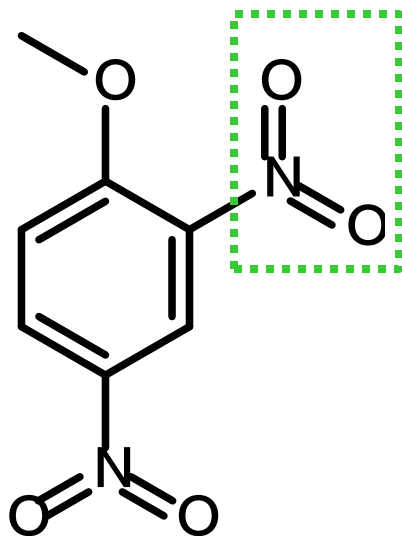
5-fold external cross validation results

Descriptors	Algorithm	Balanced Accuracy
SiRMS	RF	0.817
	SVM	0.800
Dragon	RF	0.816
	SVM	0.793

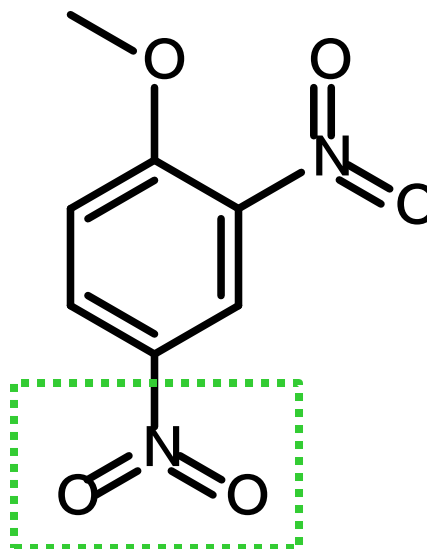


Combined contribution (effect) of fragments

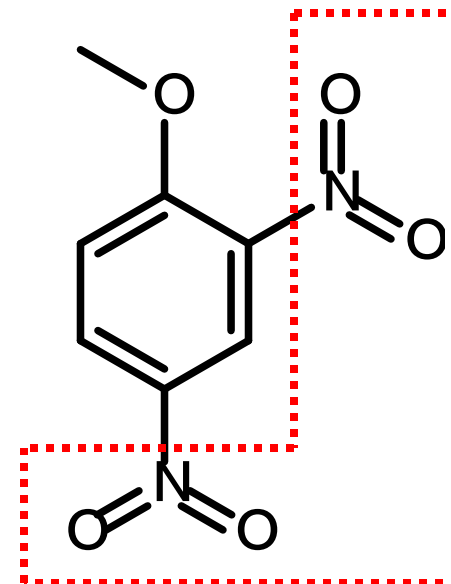
RF+SiRMS



Contribution = **0**
(non-mutagen)



Contribution = **0**
(non-mutagen)

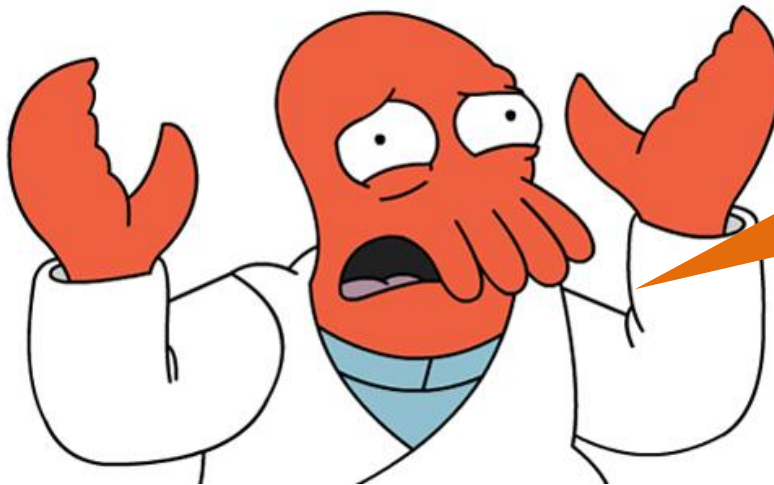


Contribution = **1**
(mutagen)

Questions

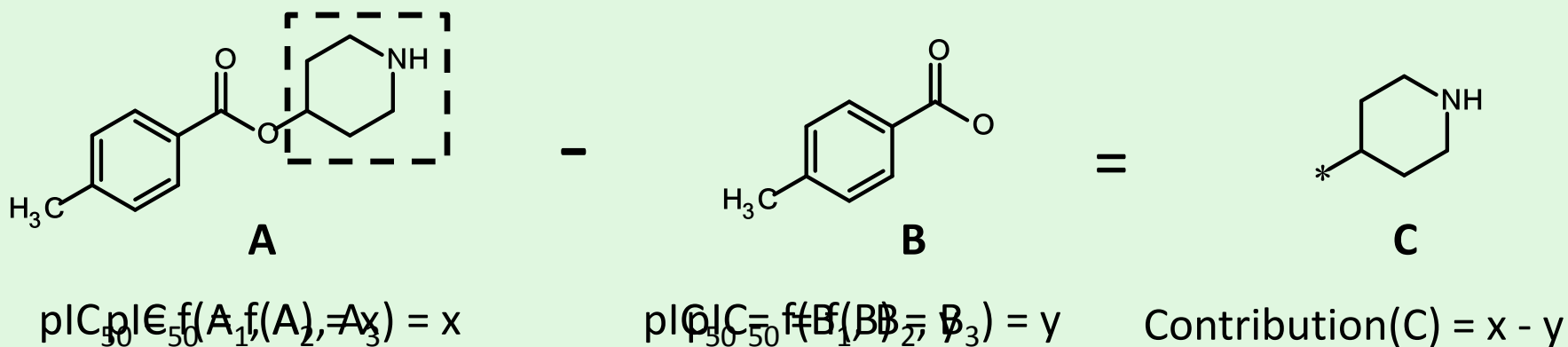


How does the fragment influence the property?

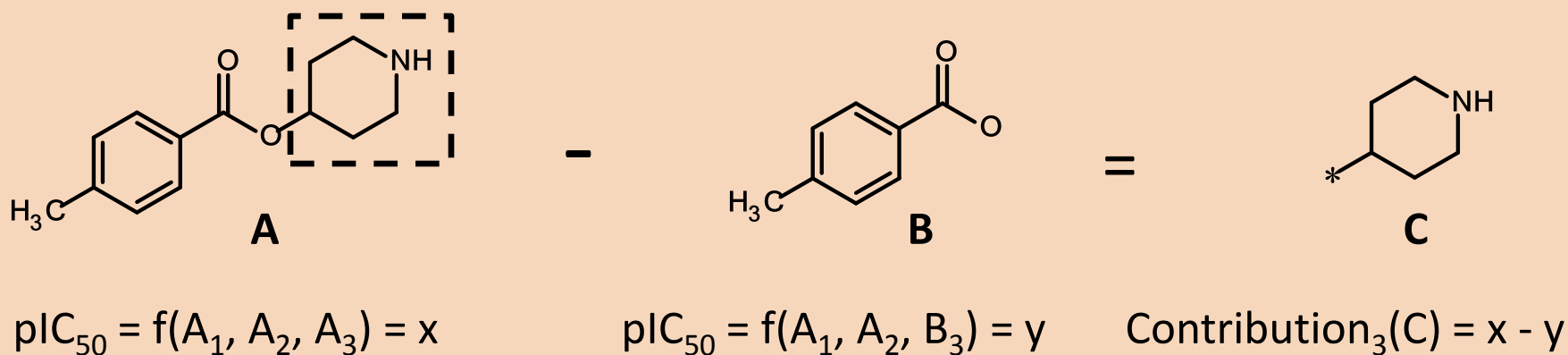


Functional interpretation of QSAR models

Structural interpretation



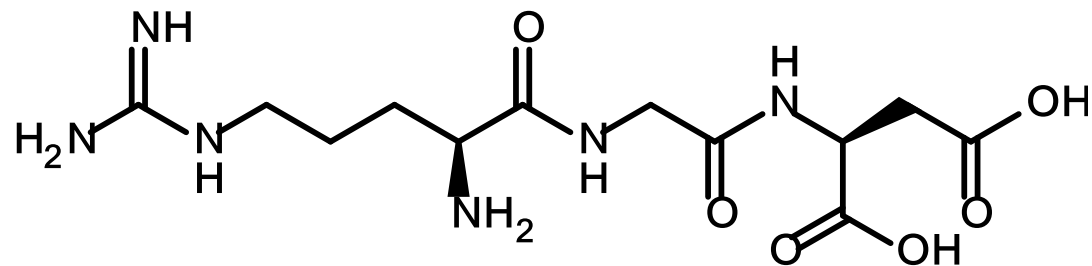
Functional interpretation



1, 2, 3 – groups of descriptors represented different physico-chemical factors (charge, H-bonding, etc) of compound A and B.

Antagonists of fibrinogen receptor

Antagonists of fibrinogen receptor: dataset



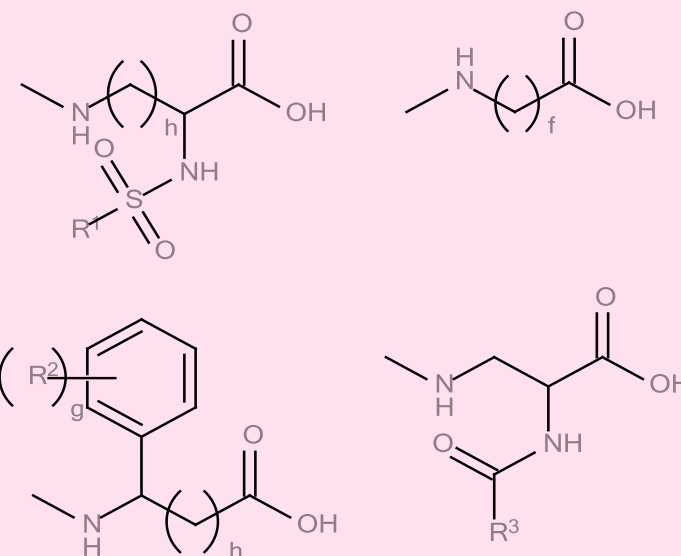
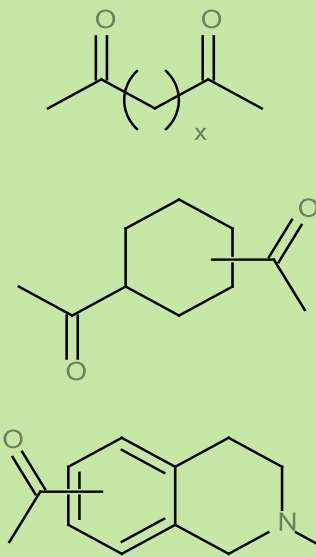
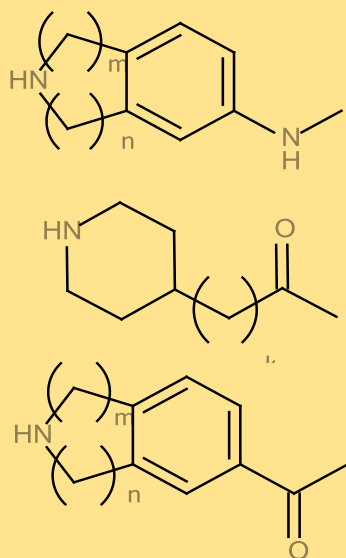
Arg-Gly-Asp

Arg-mimetic

Linker

Asp-mimetic

Fragment examples



338 compounds

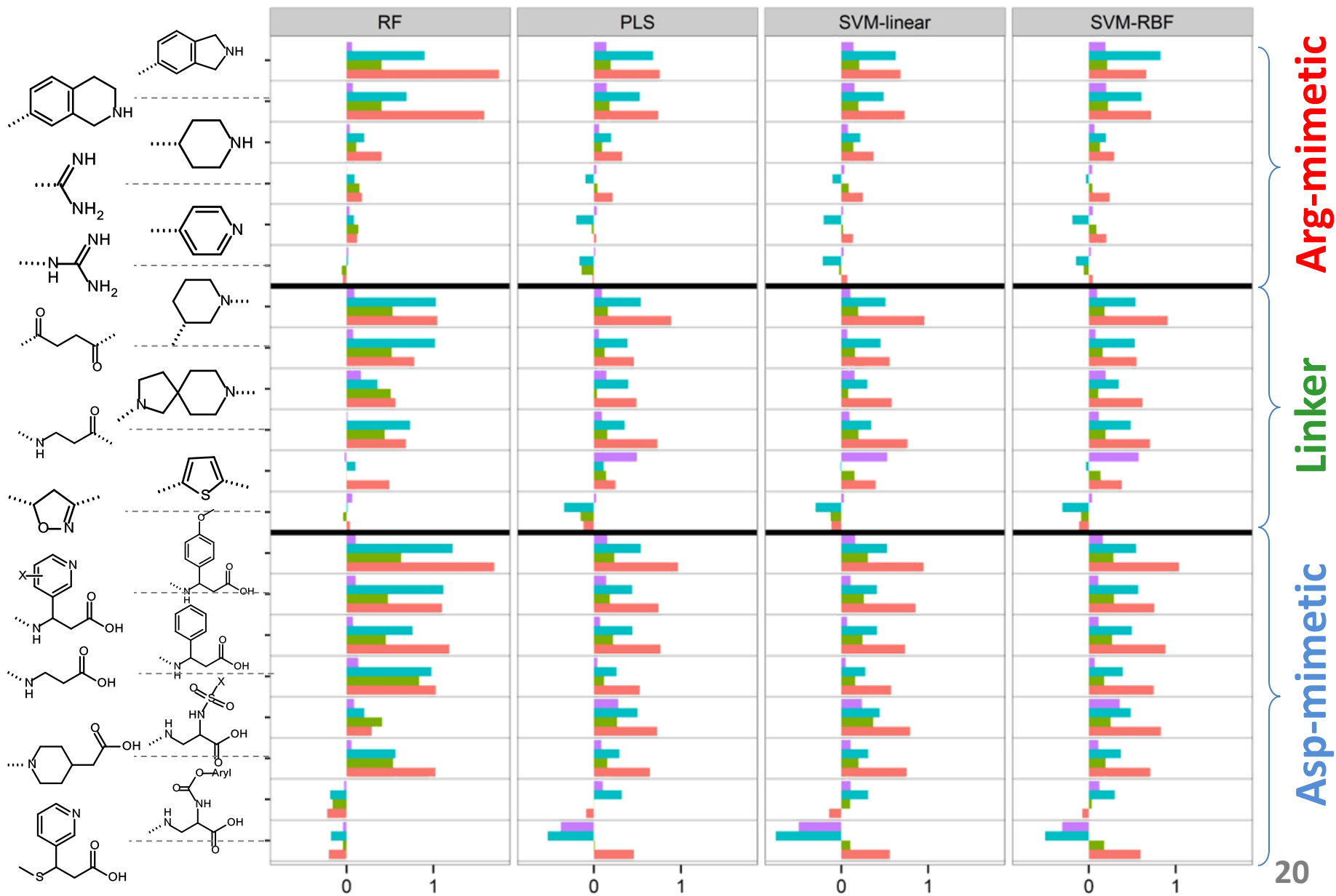
Antagonists of fibrinogen receptor: models

5-fold external cross validation results

Algorithm	R²	RMSE
RF	0.72	0.80
SVM (RBF kernel)	0.69	0.84
SVM (linear)	0.67	0.88
PLS	0.67	0.87

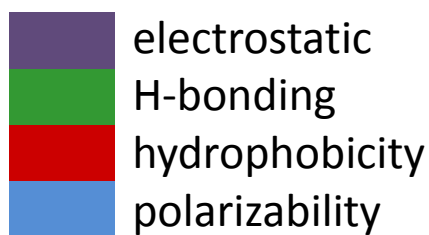
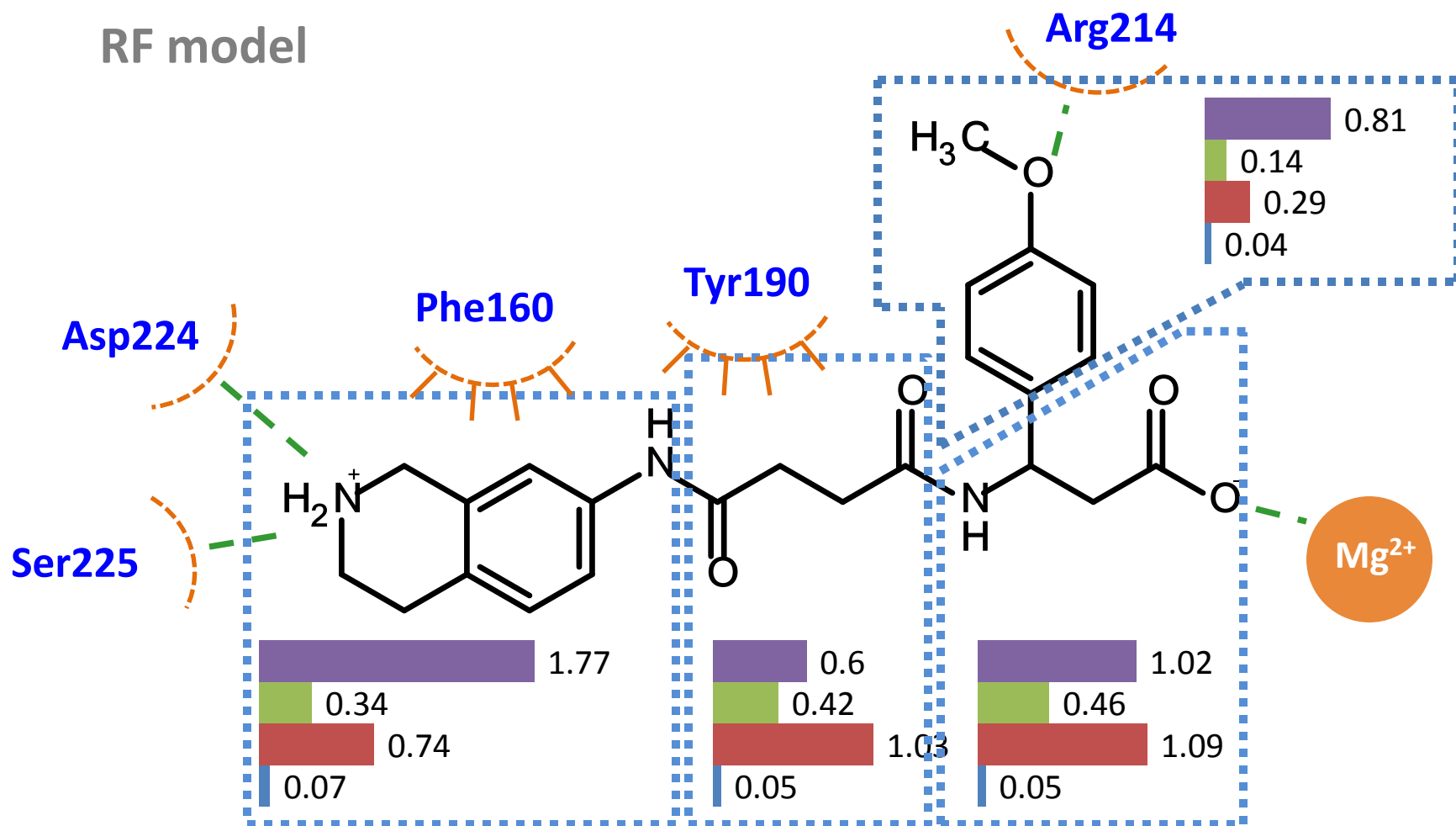
Functional interpretation

polarizability hydrophobic H-bonding electrostatic



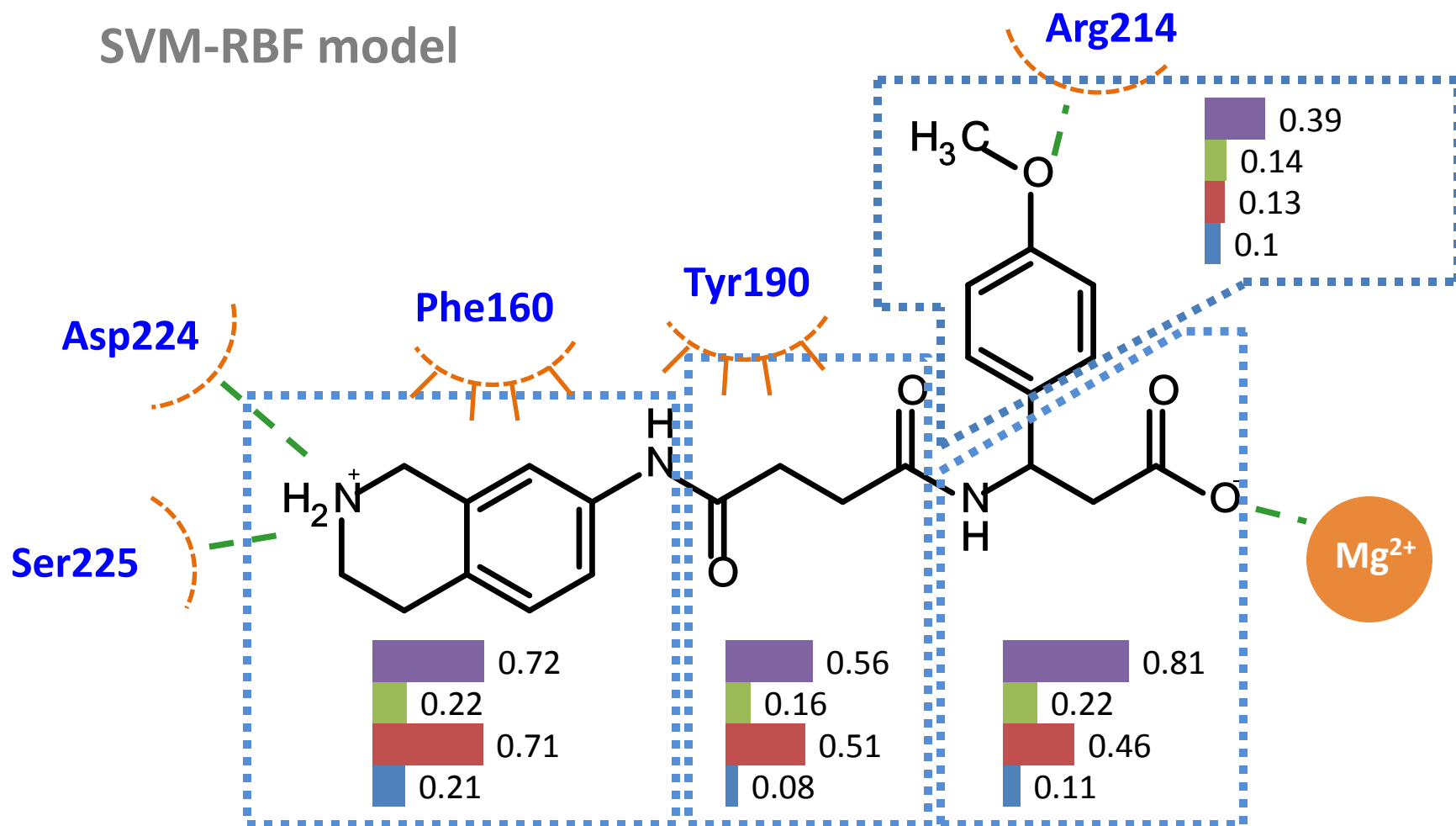
Functional interpretation of RF model

RF model



Functional interpretation of SVM model

SVM-RBF model



electrostatic
H-bonding
hydrophobicity
polarizability

Conclusions

Almost any QSAR model can be interpreted using the proposed schemes.

Results of structural and functional interpretation obtained from different models are well correlated.

Structural interpretation allows fragments ranking and finding potential alerts.

Functional interpretation may provide a guess which factors are dominated and influence on the investigated property.

Related projects

A.V. Bogatsky Physico-Chemical Institute,
Chemoinformatic group:

<http://qsar4u.com>



SiRMS project on GitHub:

<https://github.com/DrrDom/sirms>



DrrDom / sirms

Online structural interpretation tool:

<http://www.physchem.od.ua/compute>