MOLECULAR FIELD TOPOLOGY ANALYSIS AS AN ADVANCED TOOL FOR QSAR/QSPR STUDIES

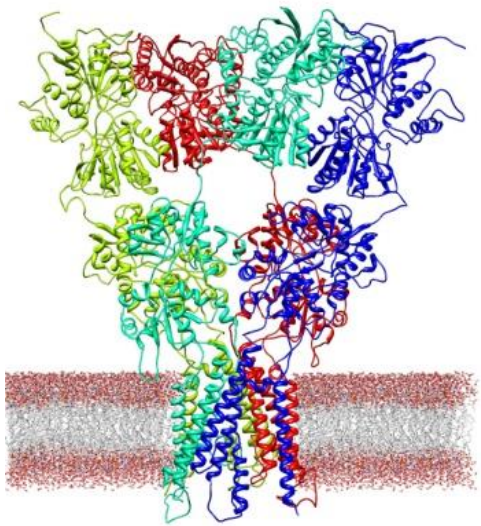
Vladimir A. Palyulin,
Eugene V. Radchenko,
Nikolay S. Zefirov

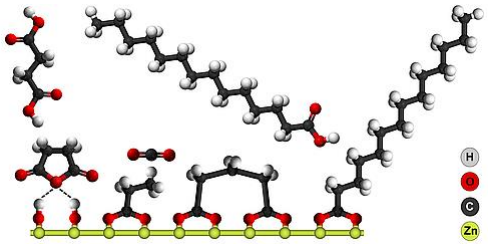
Department of Chemistry,
Lomonosov Moscow State University,
Moscow, 119991 Russia

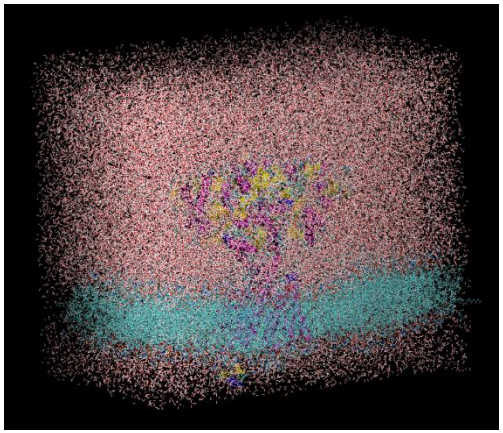
Design of Compounds/Materials with Required Activities/Properties:

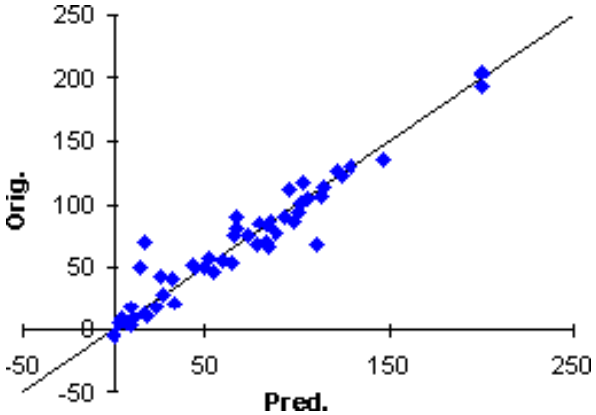
Molecular Modelling, Docking, Molecular Dynamics

QSAR/QSPR

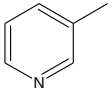
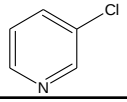
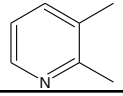
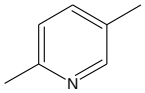
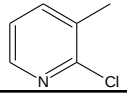
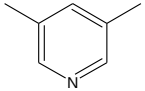
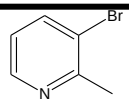
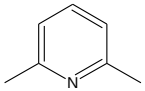








QSAR/QSPR: General Approach

	A	Structure	Descriptors			
Training	–		–	–	–	–
	–		–	–	–	–
	–		–	–	–	–
	–		–	–	–	–
Test	–		–	–	–	–
	–		–	–	–	–
New	?		–	–	–	–
	?		–	–	–	–

Model

F: *A*=F(*S*)

Predictivity

Δ*A*

Prediction

PROPERTIES

Physico-chemical properties:

Boiling points, melting points, density, viscosity, surface tension, solubility in various solvents, lipophilicity, magnetic susceptibility, retention indices, enthalpy of formation, etc.

Biological activity:

IC_{50}, EC_{50}, LD_{50}, MEC, ILS, etc.

DESCRIPTORS (“global” and “local”)

Topological indices

Quantum-chemical

Physico-chemical

Fragmental

Solvation Index. Prediction of Non-Specific Solvation Enthalpy of Organic Compounds

Solvation enthalpy (kJ/mol)

$$\Delta H_{solv}^{A/Y} = 4.52 + 9.04\,{}^{1}\chi^{S}$$

n **= 141** ***R*** **= 0.985** ***s*** **= 2.1**

Vaporization enthalpy (kJ/mol)

$$\Delta H_{vap}^{A} = 4.13 + 9.52\,{}^{1}\chi^{S} + 0.827\,\mu^{2}$$

n **= 528** ***R*** **= 0.989** ***s*** **= 2.0**

μ – dipole moment

${}^{1}\chi^{S}$ – 1-st order solvation topological index

Z_i – period number (measure of atom size)

δ_i – number of non-hydrogen neighbors

$$ {}^{1}\chi^{S} = \frac{1}{4}\sum_{(bonds)} \frac{Z_i Z_j}{\sqrt{\delta_i \delta_j}} $$

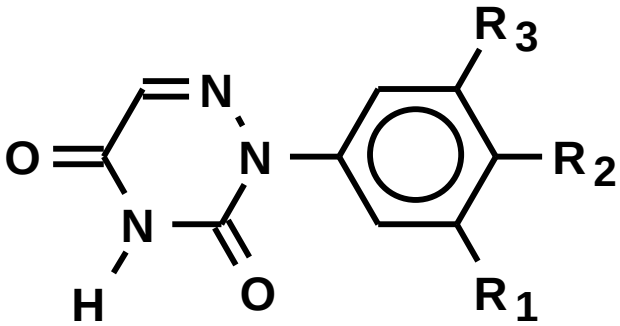
I.S.Antipin, N.A.Arslanov, V.A.Palyulin, A.I.Konovalov, N.S.Zefirov
Dokl. Akad. Nauk,1993, 331(2), 173-176.

Fragmental Descriptors

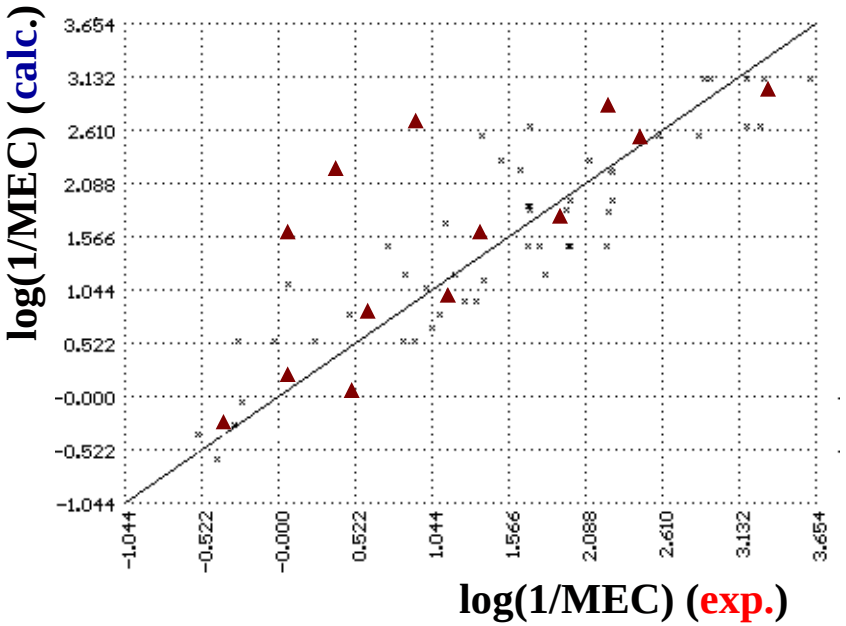
The numbers of fragments of various kind and various size (chains, cycles, branched fragments) in a molecule with several levels of classification of atoms. For each molecule hundreds of fragmental descriptors can be computed.

Good correlations with various physico-chemical properties (enthalpy of formation, boiling points, density of organic liquids, surface tension, lipophilicity, etc.), could be used for prediction of mutagenicity, logBB, etc.

Fragmental Approach: Anticoccidial Activity of Triazines



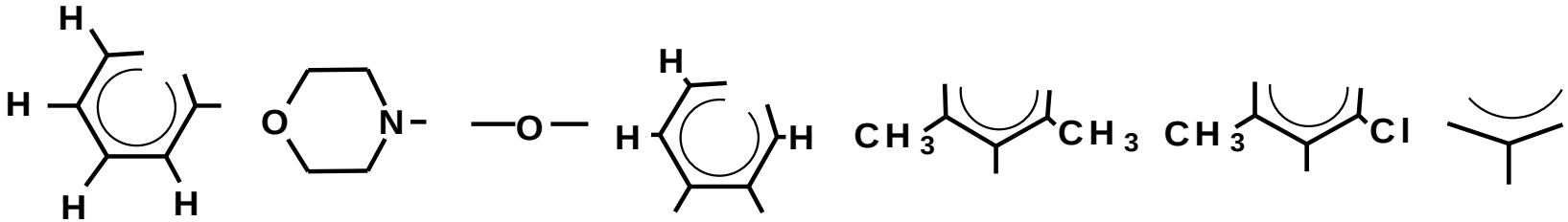
Training and Test (▲) sets:



log(1/MEC) = -0.258 Fr1 + 0.594 Fr2 - -0.734 Fr3 - 0.255 Fr4 - 0.558 Fr5 - -0.452 Fr6 + 0.639Fr7 -0.728,

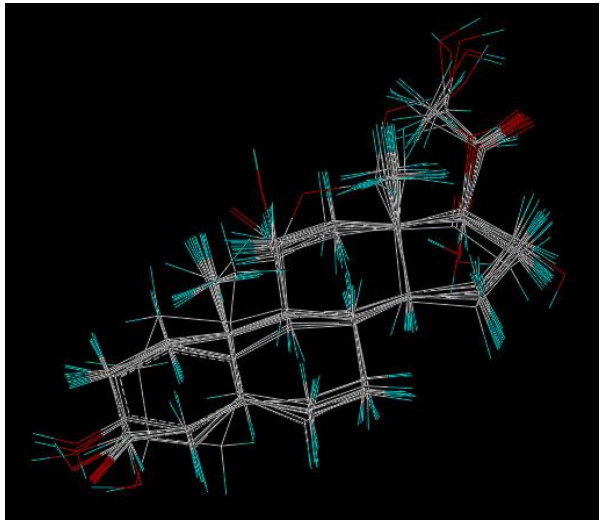
n=54, R=0.90, s=0.48

where Fr i - the number of the fragments :



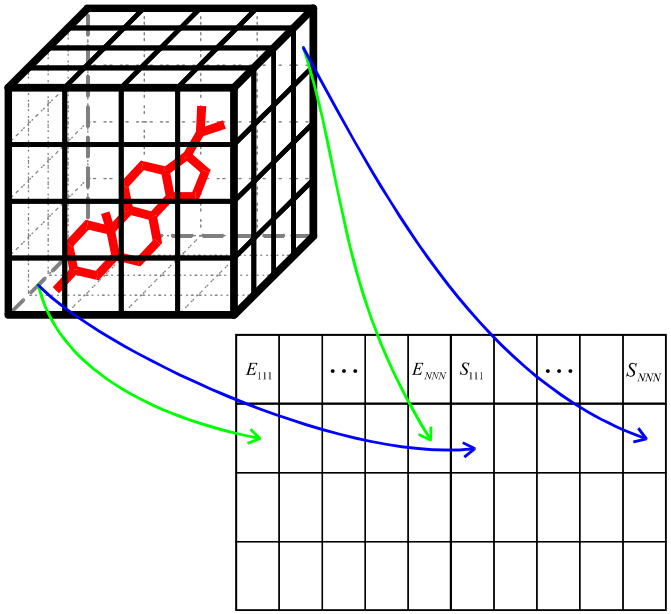
Fr1 Fr2 Fr3 Fr4 Fr5 Fr6 Fr7

Comparative Molecular Field Analysis (CoMFA)



$$E = \frac{1}{4\pi\varepsilon_0\varepsilon}\sum_i \frac{q_i\, q_p}{d_{ip}}$$

$$S = \sum_i \left[-\frac{A_{ip}}{d_{ip}^6} + \frac{B_{ip}}{d_{ip}^{12}} \right]$$



3D alignment could be difficult for flexible structures

Molecular Field Topology Analysis (MFTA)

The main steps of an MFTA study:

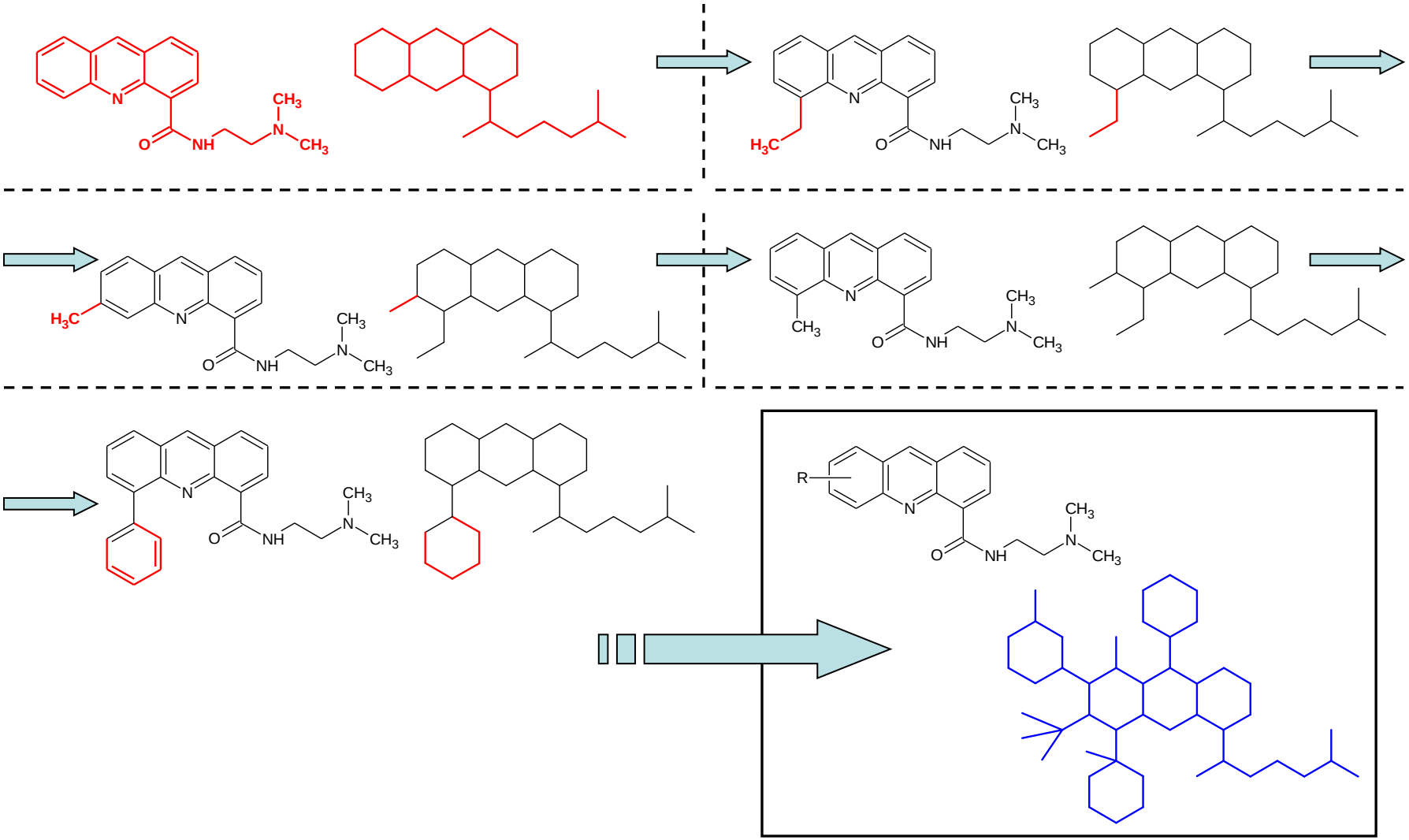
Construction of a ***molecular supergraph*** *(MSG)* that allows the superposition of any training set structure

Analysis of ***local physico-chemical molecular characteristics*** (atom and bond properties) in a uniform manner for all training set structures leading to a predictive PLS model

Radchenko E.V., Palyulin V.A., Zefirov N.S., in *Chemoinformatics Approaches to Virtual Screening*, ed. by A.Varnek, A.Tropsha, RSC, 2008, 150-181.

Palyulin V.A., Radchenko E.V., Zefirov N.S. J. Chem. Inf. Comp. Sci. 2000, 40, 659-667.

Molecular Supergraph



Local Molecular Descriptors

- **Electrostatic**
  - Atomic charge Q (electronegativity equalization approach: Gasteiger)
  - Absolute atomic charge $Q_a = abs(Q)$
- **Steric**
  - Van der Waals radius R (Bondi)
  - Group van der Waals radius R_g (atom+hydrogens)
  - Van der Waals radius of first environment R_e (atom+neighbours)
  - Exposed van der Waals surface of an atom S
  - Relative steric exposure $A=S/S_{free}$
- **Lipophilic**
  - Atomic lipophilicity contribution L_a (environment-dependent: Ghose, Crippen)
  - Group lipophilicity L_g (atom+hydrogens)
- **Hydrogen bonding**

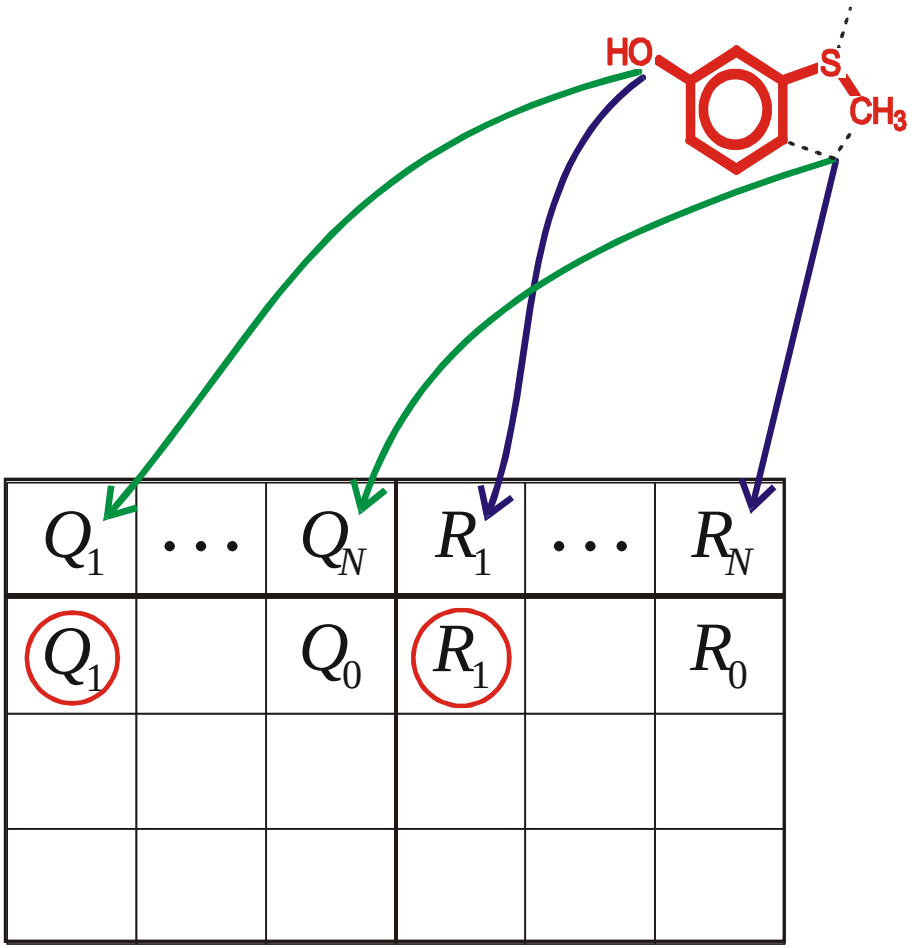
Hydrogen bond donor (H_d) and acceptor (H_a) ability of an atom (Abraham)

Construction of a Descriptor Matrix

Controlled by structure mapping to the supergraph

For the occupied positions: local property on a corresponding atom

For unoccupied positions: neutral value of local property



Statistical analysis

- Partial Least Squares Regression (PLS) Modelling
- Analysis of contribution to activity for each local property in every position of molecular supergraph
- Two-phase descriptor selection process

  Removal of low-variable descriptors

  Optional Q^2-guided selection of optimal descriptor subset using genetic algorithm
- Stable cross-validation (SCV) procedure to eliminate dependence on test set split and chance correlations

MFTA-Oriented Structure Generation

MFTA (and most other QSAR techniques) provide meaningful predictions only within the same general chemical class

Need for a specialized MFTA-oriented generation

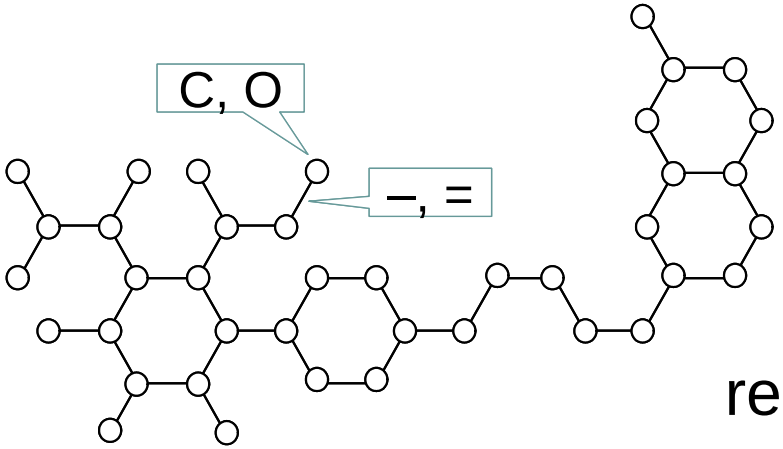
Structural constraints:

- The number of substituted positions in the central fragment
- Forbidden molecular fragments (“bad list”)
- Molecular mass

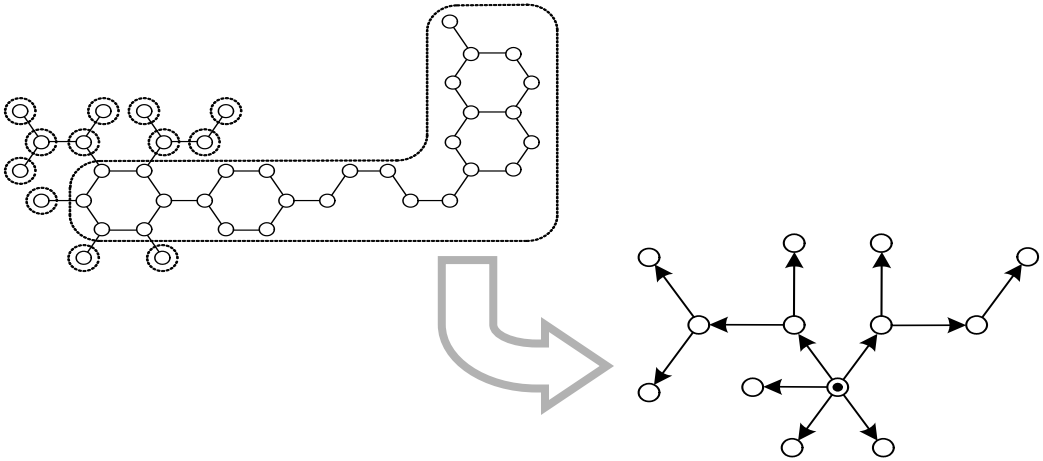
Generation modes: Deterministic or Stochastic

MFTA-Oriented Structure Generation

Analysis of a training set



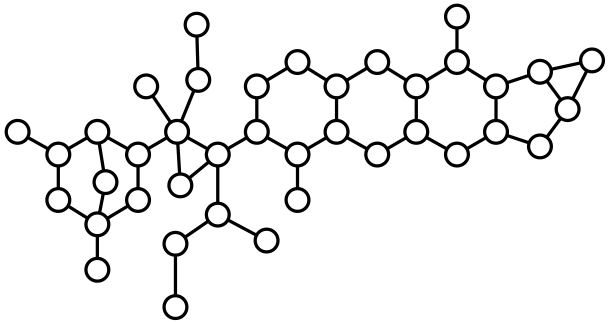
Fragmental supergraph – compact representation of all possible structures



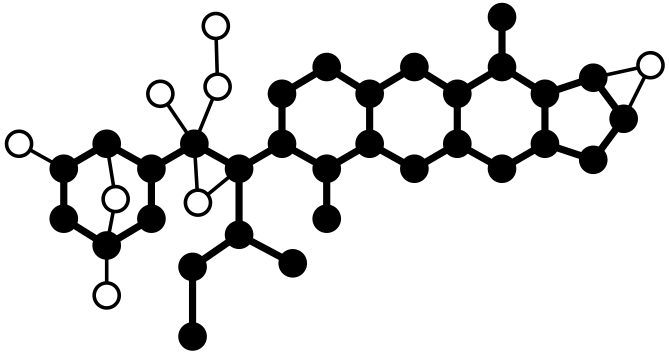
Positive AMPA receptor modulators. Benzotriazinone and benzopyrimidinone derivatives

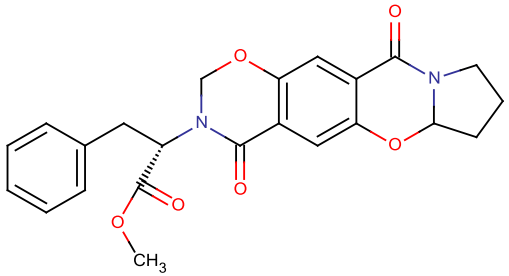
Total 37 compounds
Excluded 3 compounds with no or very low activity
Training set: 34 compounds of consistent stereochemistry
Endpoint: $pEC_{2x} = \log(1 / EC_{2x})$

Molecular Field Topology Analysis (MFTA)



Molecular supergraph





Example mapping

Muller R. et al., *Bioorg. Med. Chem. Lett.*, 2011, **21**, 6170-6175.
Palyulin V.A., Radchenko E.V., Zefirov N.S., *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 659-667.

MFTA models of ampakine activity

Local molecular descriptors:

Q – **atomic charge** (electronegativity equalization approach)

R_e – **effective group van der Waals radius** (steric requirements of central atom and attached non-hydrogen atoms)

L_g – **group lipophilicity** (contributions of central atom and attached hydrogens)

H_d – **hydrogen bond donor ability**

H_a – **hydrogen bond acceptor ability**

Descriptors	N_F	R^2	*RMSE*	Q^2	*RMSEcv*
Q, R_e	2	0.73	0.54	0.44	0.74
Q, R_e, L_g	2	0.80	0.47	0.47	0.72
Q, R_e, H_d, H_a	2	0.73	0.54	0.45	0.73
Q, R_e, L_g, H_d, H_a	2	0.78	0.49	0.49	0.71

N = number of compounds, N_F = number of factors in the PLSR model,
R^2 = squared correlation coefficient, ***RMSE*** = root-mean-square error,
Q^2 = cross-validation parameter, ***RMSEcv*** = root-mean-square error of cross validation

Optimal MFTA model

Local molecular descriptors:

Q **– atomic charge**

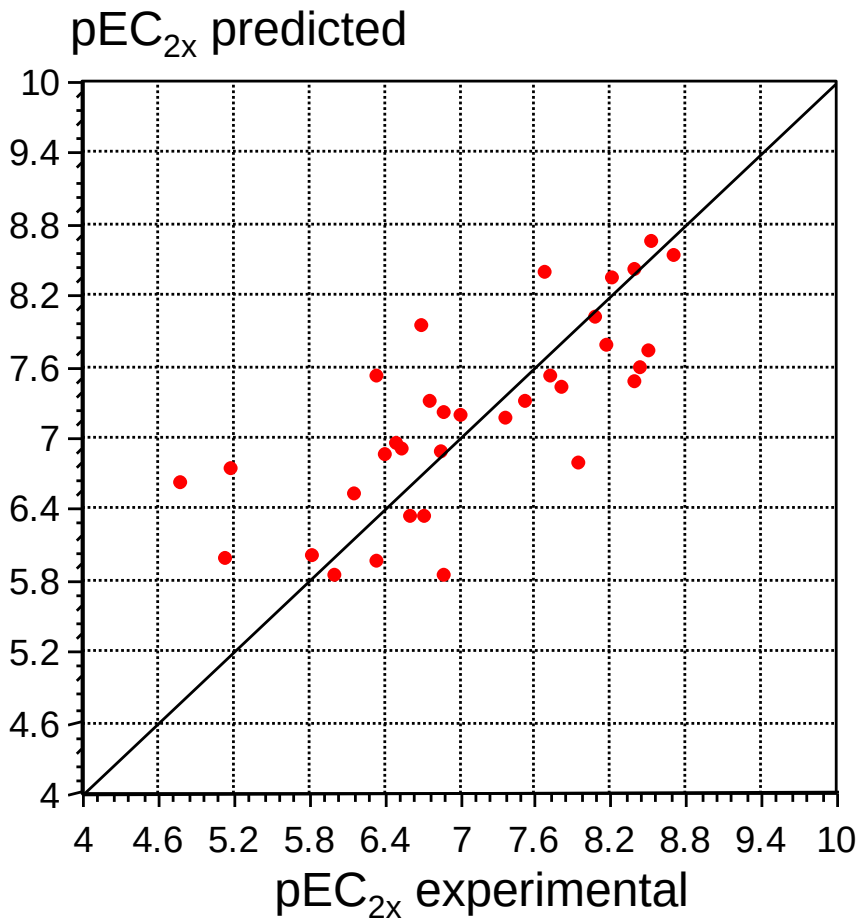
R_e **– effective group van der Waals radius**

L_g **– group lipophilicity**

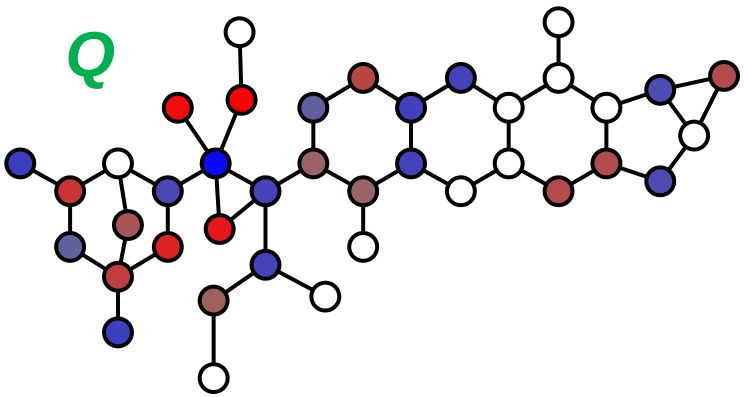
R^2 = 0.80

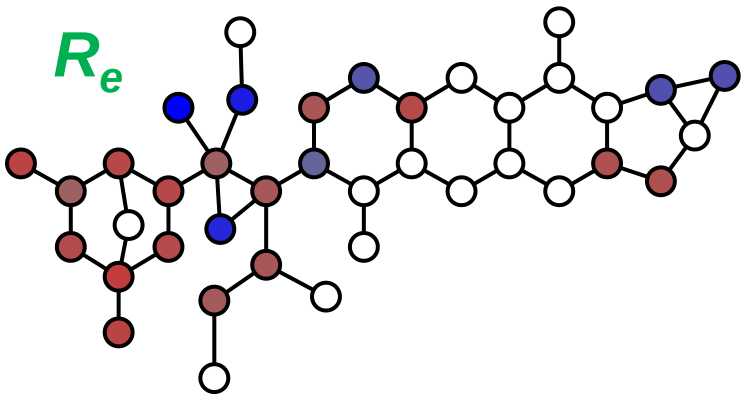
Q^2 = 0.47

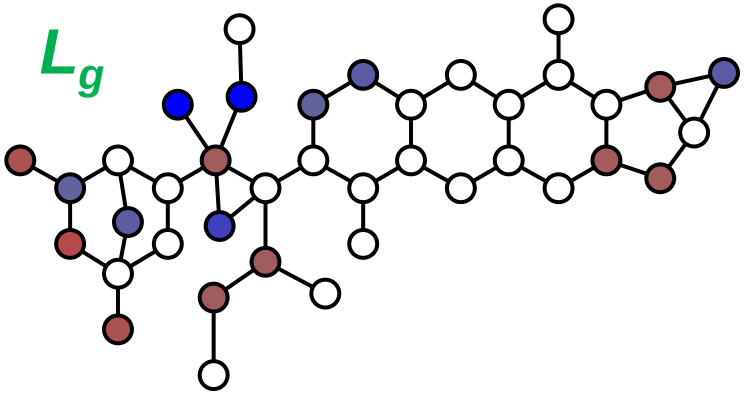
RMSEcv = 0.72



MFTA activity maps







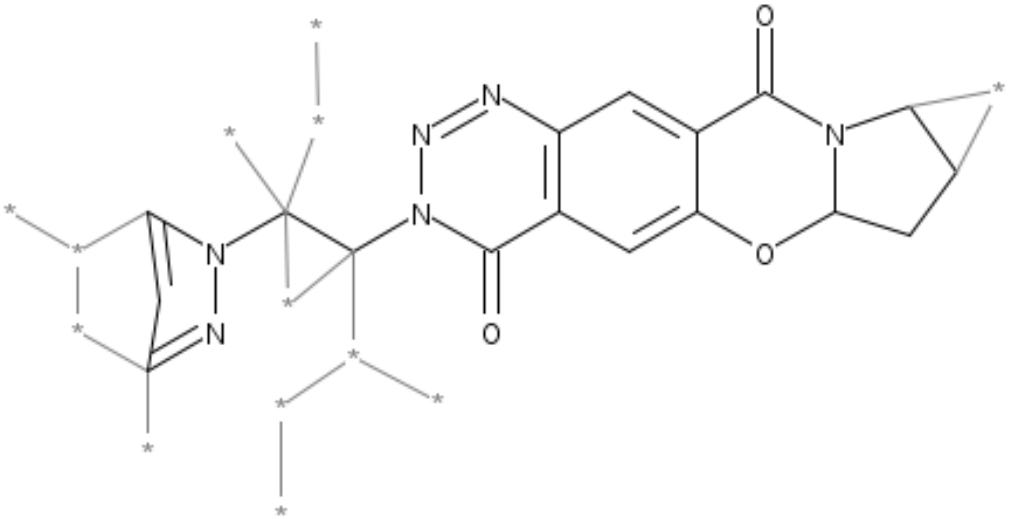
Activity is increased by F and other acceptor groups in phenyl group and α-position but not in β-position

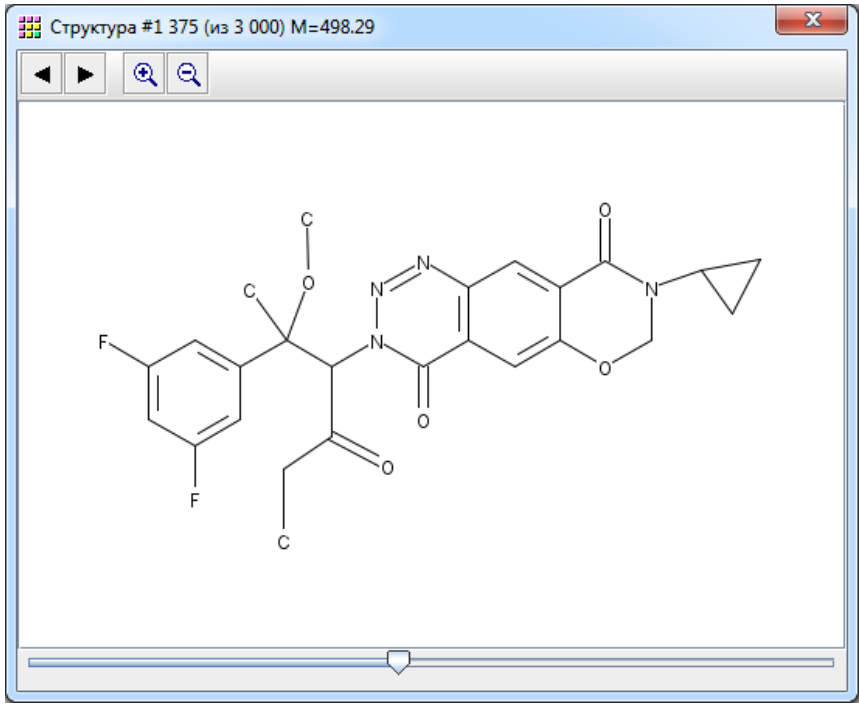
Cyclopropyl group is not desirable

Red: activity increases **Blue: activity decreases**

Molecular Design and Virtual Screening of Promising Structures

Generated structure library was built using specialized MFTA-oriented structure generator in stochastic mode – 3000 out of 14708 structures





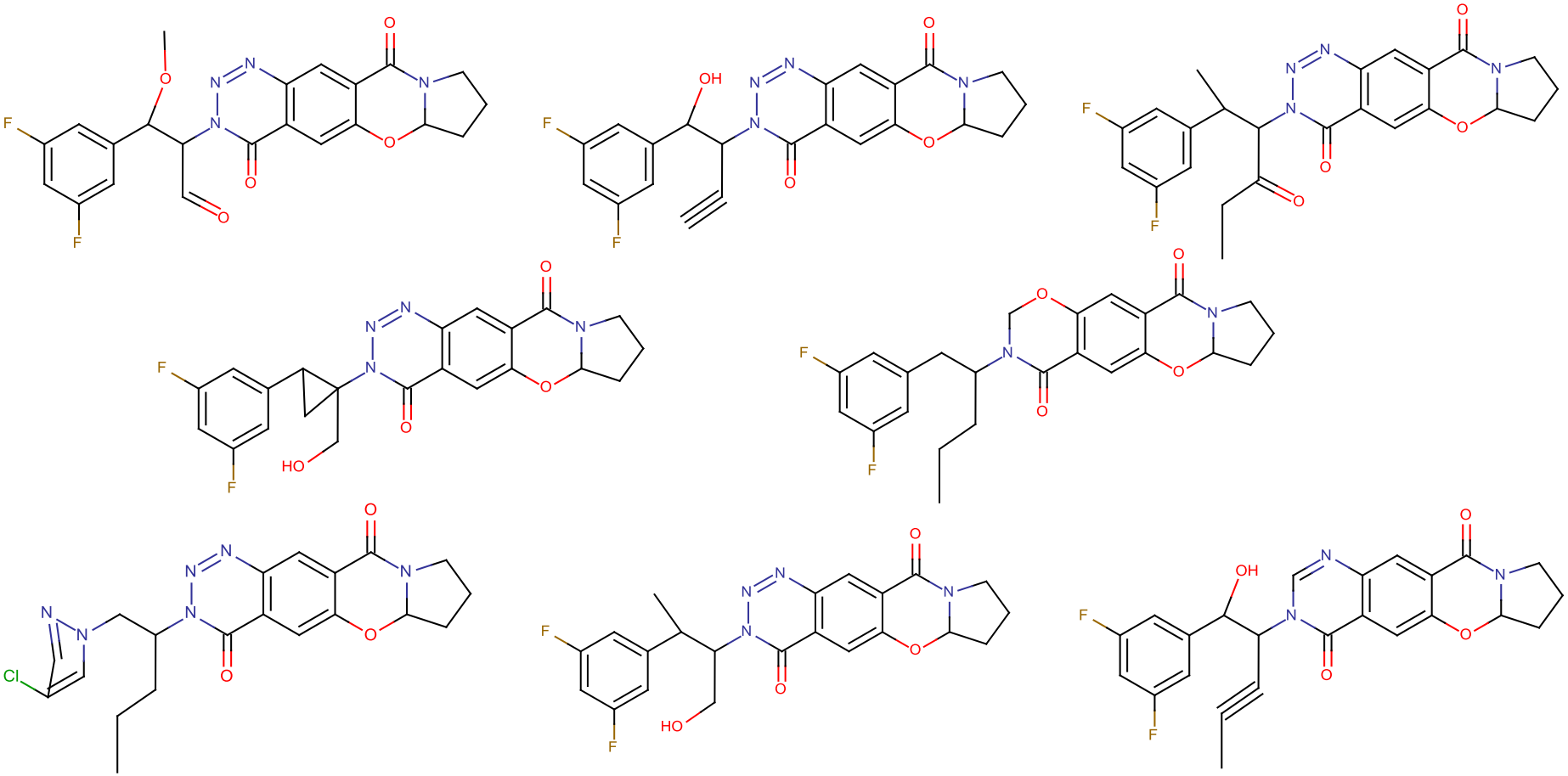
Melnikov A.A., Palyulin V.A., Radchenko E.V., Zefirov N.S. *Doklady Chemistry*, 2007, *415*, 196-199.
Radchenko E.V., Palyulin V.A., Zefirov N.S., in *Chemoinformatics Approaches to Virtual Screening*, ed. by A.Varnek, A.Tropsha, RSC, 2008, 150-181.

Molecular Design and Virtual Screening of Promising Structures

Selection of desired activity profile

– 49 structures have predicted $pEC_{2x} > 8$

Examples of proposed structures



Application Example

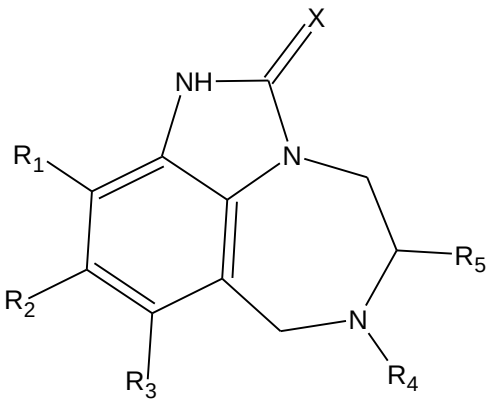
Anti-HIV Activity of Tetrahydroimidazobenzodiazepinone (TIBO) Derivatives: Reverse Transcriptase Inhibition

73 compounds

Kukla M.J. et al. *J. Med. Chem.* 1991, 34, 746-751

Ho W. et al. *J. Med. Chem.* 1995, 38, 794-802

Breslin H.J. et al. *J. Med. Chem.* 1995, 38, 771-793.



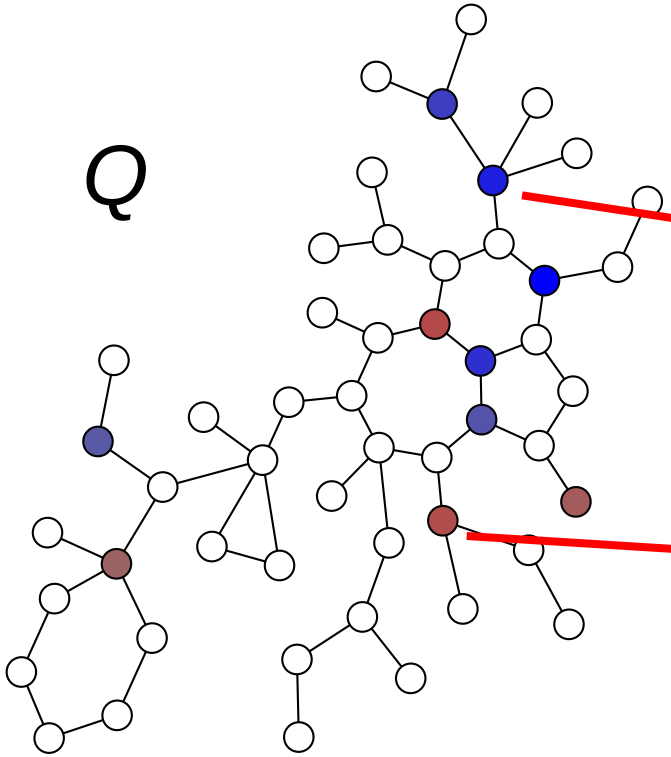
Activity: log (1/IC_{50})

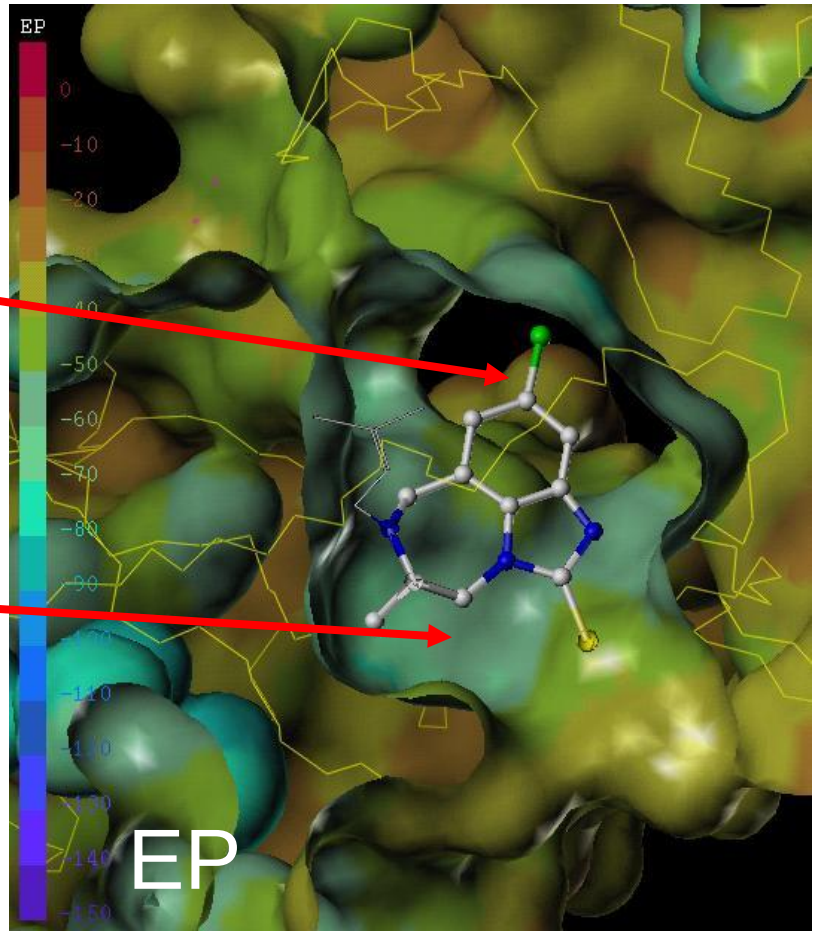
6 MFTA models for different local descriptors

Descriptor set	F	R	Q^2
Q	5	0.927	0.597
R	6	0.925	0.695
$Q\ R$	5	0.932	0.626
$Q\ R\ L_g$	**5**	**0.942**	**0.686**
$Q\ R\ H_d\ H_a$	6	0.943	0.672
$Q\ R\ H_d\ H_a\ L_g$	6	0.951	0.626

Anti-HIV Activity of TIBO Derivatives (Reverse Transcriptase Inhibition)

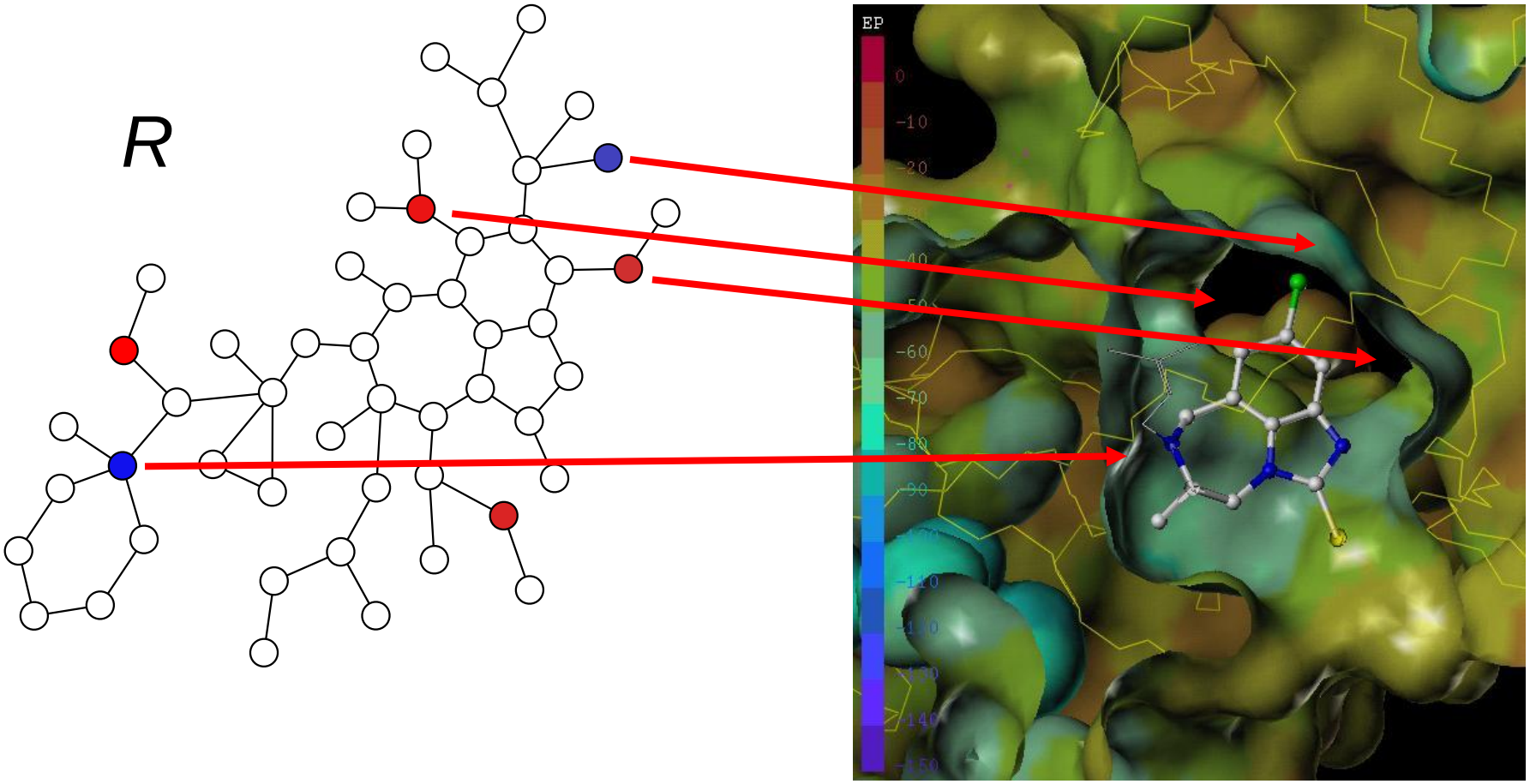
Complementarity of MFTA descriptor contributions and biotarget properties: 1tvr + 9-Cl-TIBO





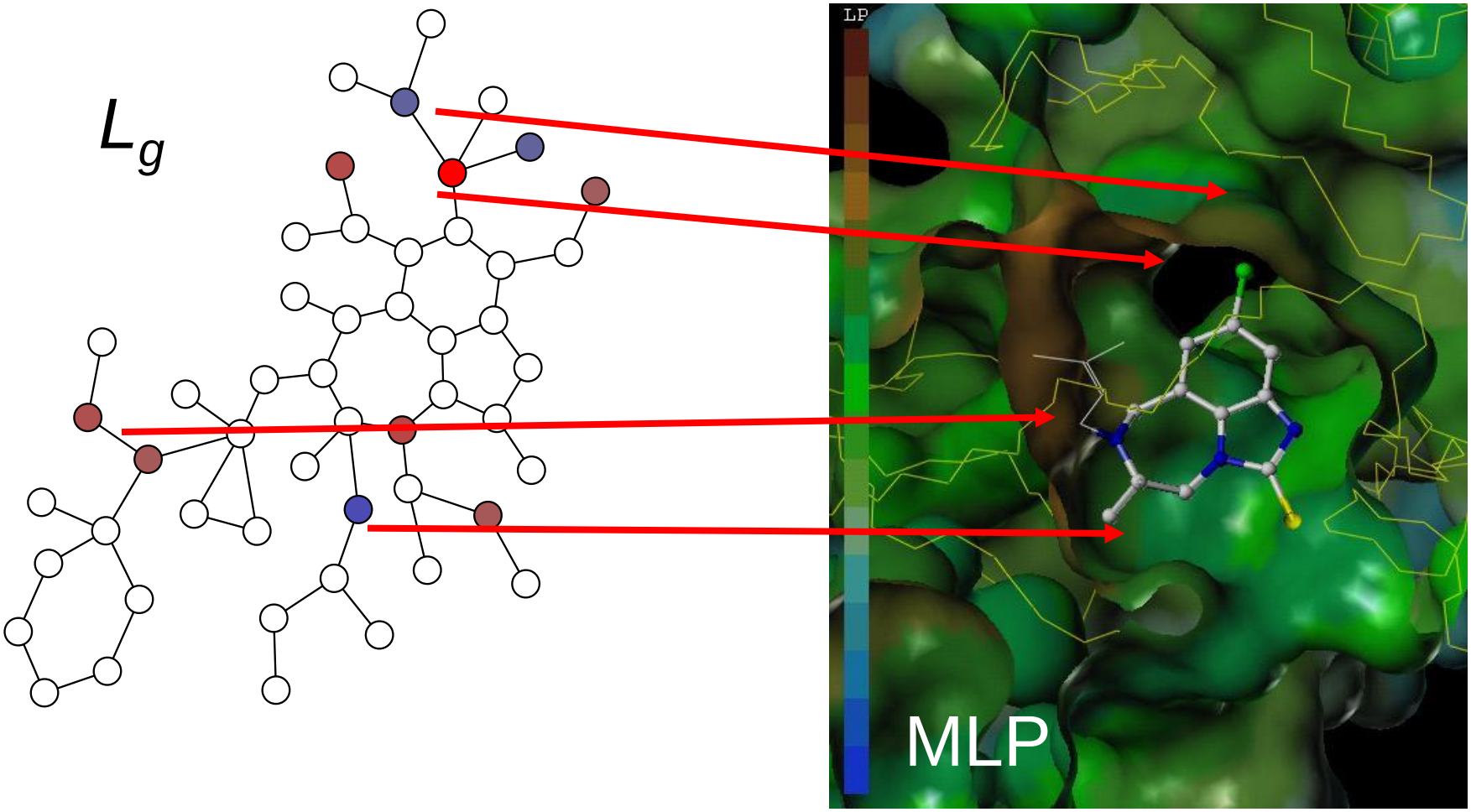
Anti-HIV Activity of TIBO Derivatives (Reverse Transcriptase Inhibition)

Complementarity of MFTA descriptor contributions and biotarget properties: 1tvr + 9-Cl-TIBO



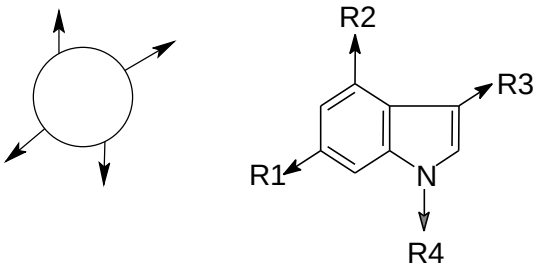
Anti-HIV Activity of TIBO Derivatives (Reverse Transcriptase Inhibition)

Complementarity of MFTA descriptor contributions and biotarget properties: 1tvr + 9-Cl-TIBO



Structure generation

Central fragment

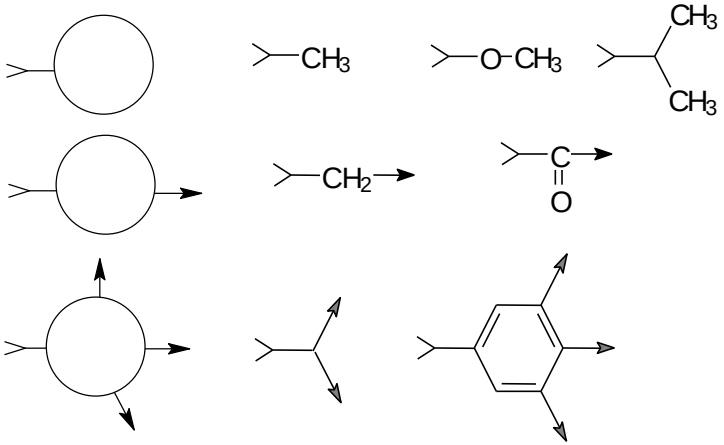


Tree-structured substituents constructed from three types of micro fragments:

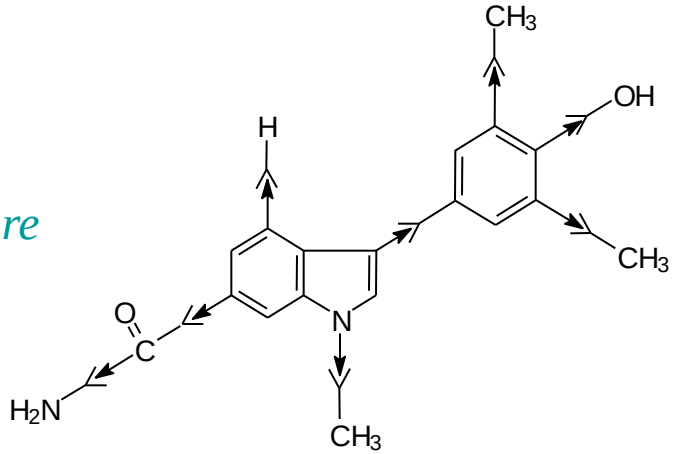
Terminal

Linear

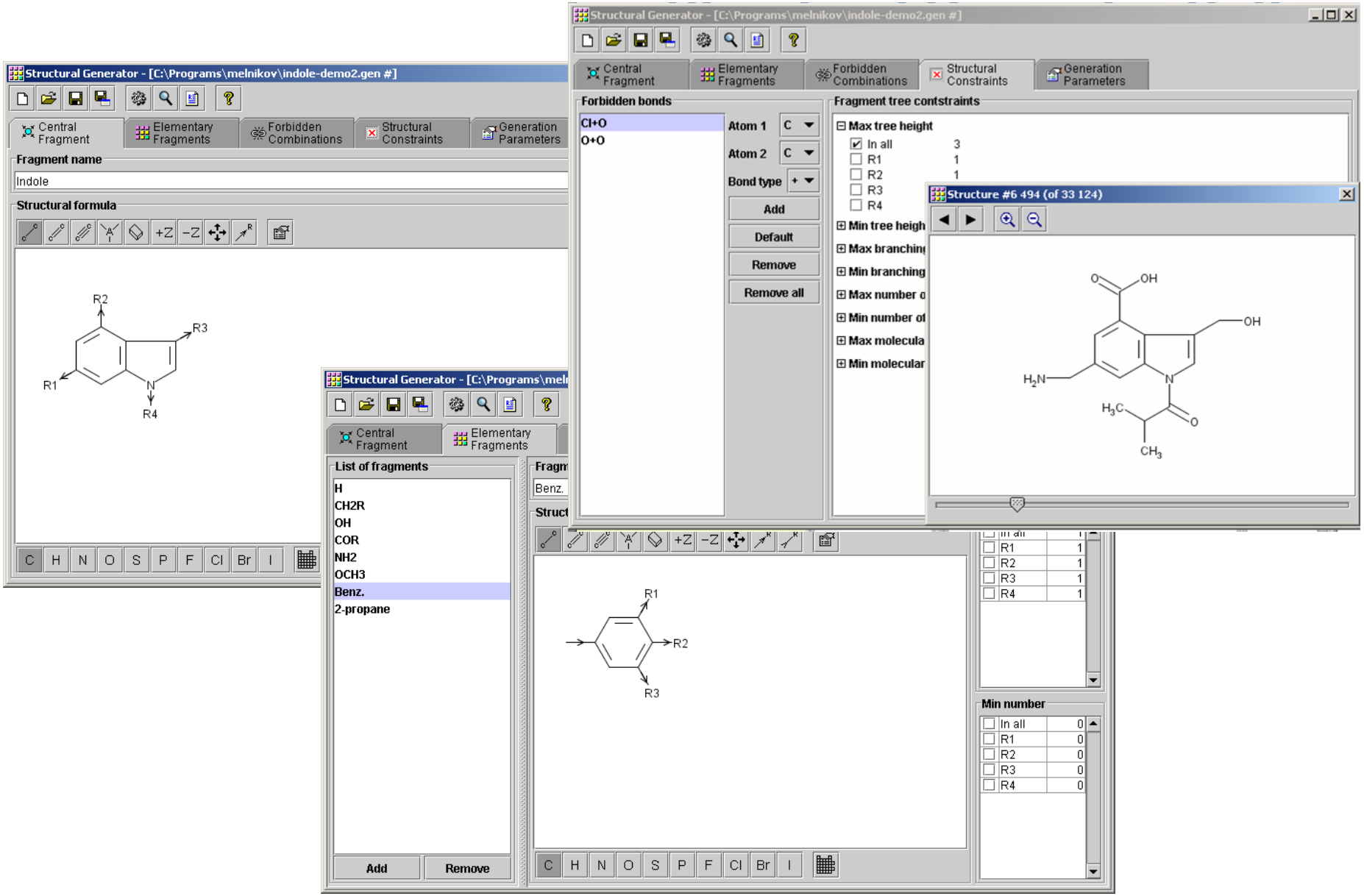
Branched



Example of a generated structure



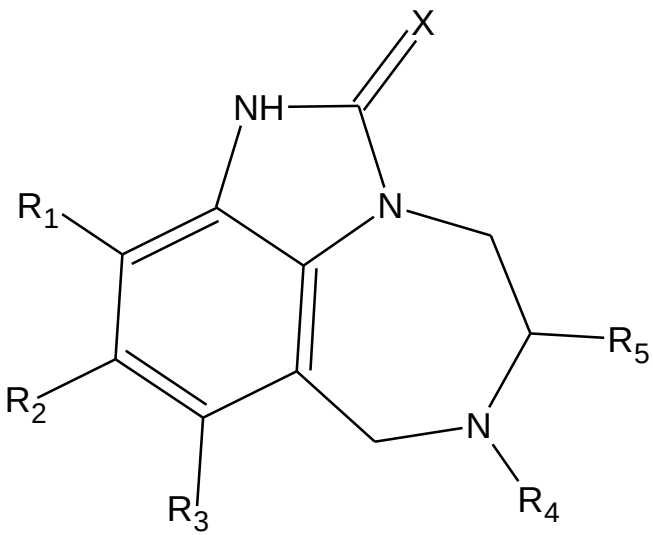
Structural Generator



Anti-HIV Activity of TIBO Derivatives (Reverse Transcriptase Inhibition)

Design of more potent structures

Structure generation



X = S

R_1-R_3 = H, CH_3, OH, NH_2, Cl, OCH_3, $N(CH_3)_2$, $CONH_2$

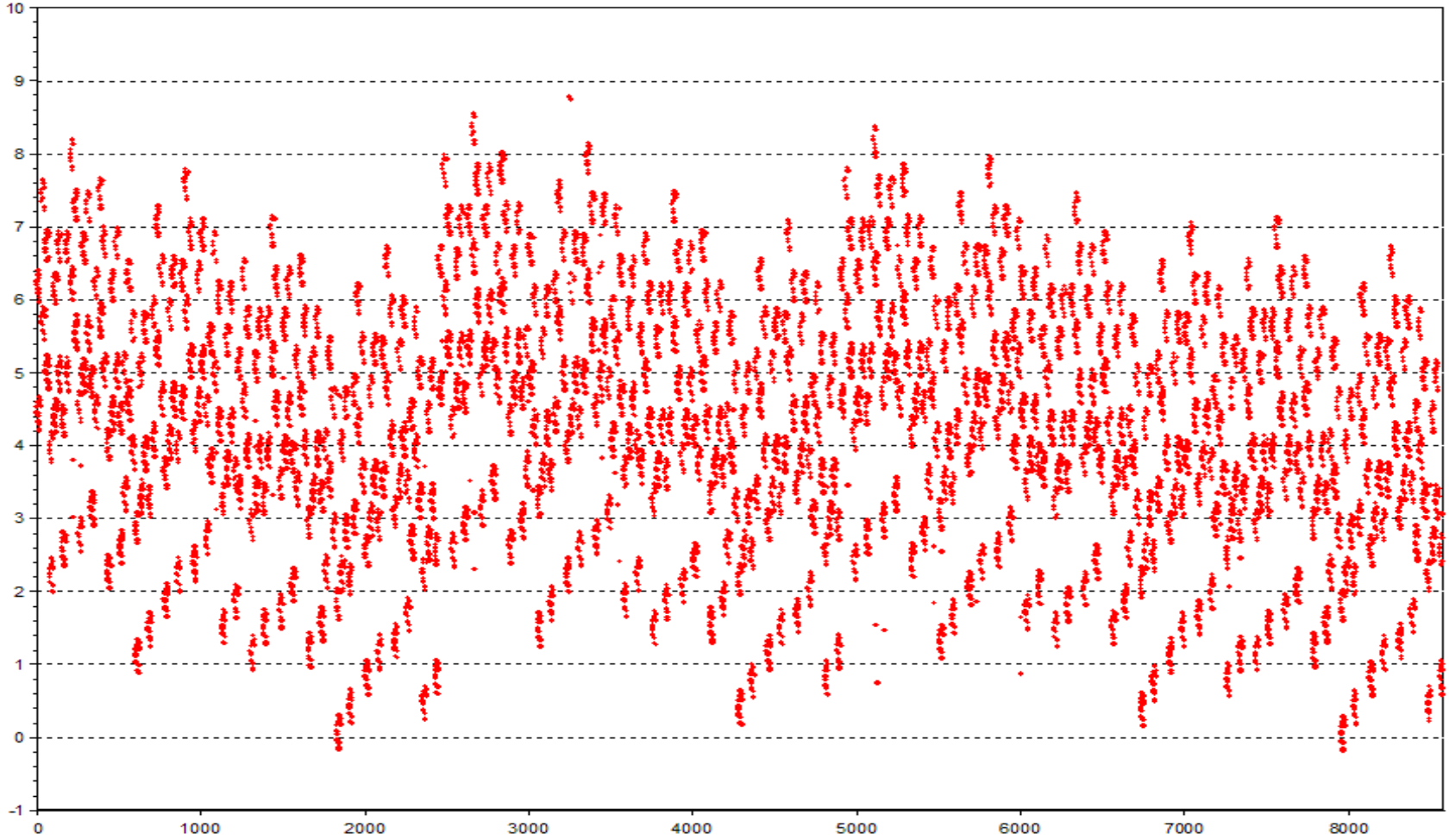
R_4-R_5 = H, CH_3, Et, Pr, ^{i}Pr

Total number of generated structures: 8575

Anti-HIV Activity of TIBO Derivatives (Reverse Transcriptase Inhibition)

Design of more potent structures

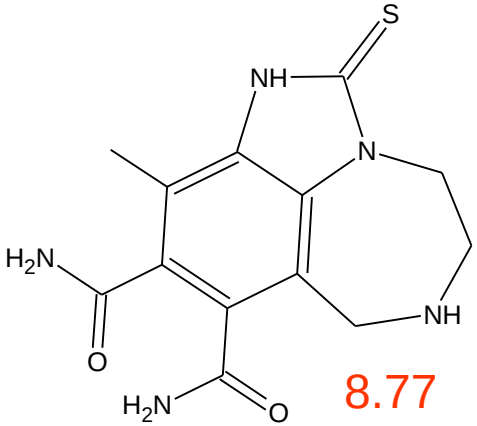
Prediction of activity using MFTA model



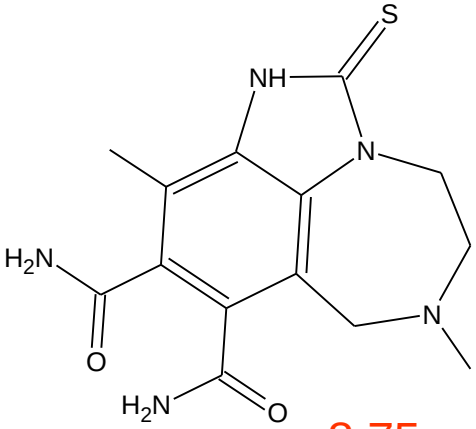
Anti-HIV Activity of TIBO Derivatives (Reverse Transcriptase Inhibition)

Design of more potent structures

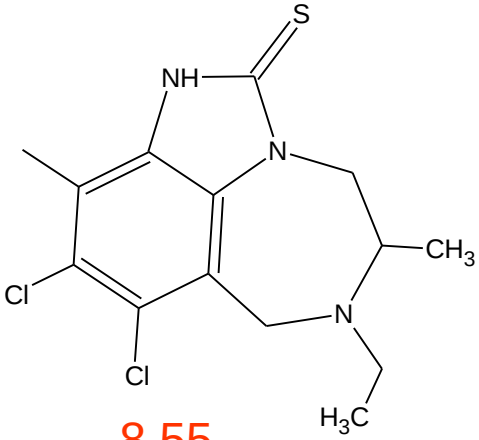
5 best predicted structures



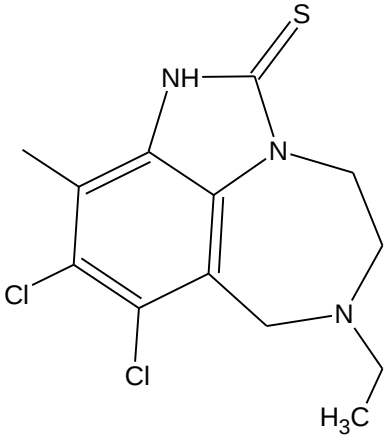
8.77



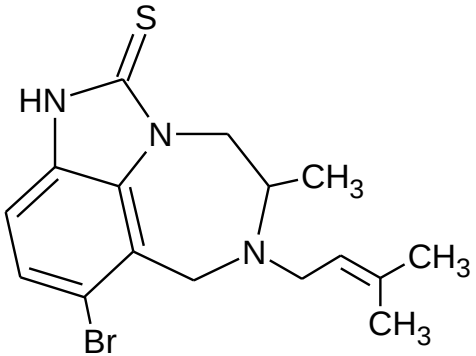
8.75



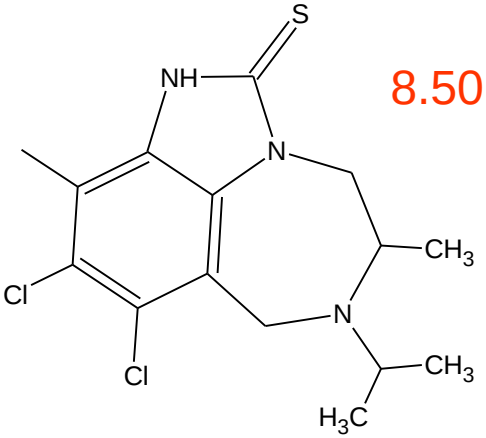
8.55



8.50



Best experimental
8.52

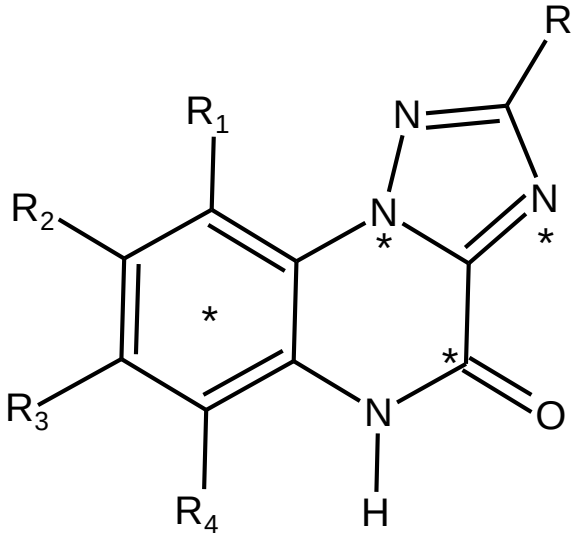


8.50

Ligand Selectivity

MFTA Study of the AMPA and Glycine/NMDA Receptor Antagonists

The difference between -logK_i for AMPA and glycine/NMDA was used as a target property in CoMFA for selectivity modelling.



-logK_i(AMPA) + logK_i(glycine/NMDA)

⇓

MFTA

⇓

SELECTIVITY MAPS

MFTA Study of the Antagonists of AMPA and glycine/NMDA Receptors: Modelling the Selectivity

$A_1=\log(1/K_i(\text{AMPA}))$

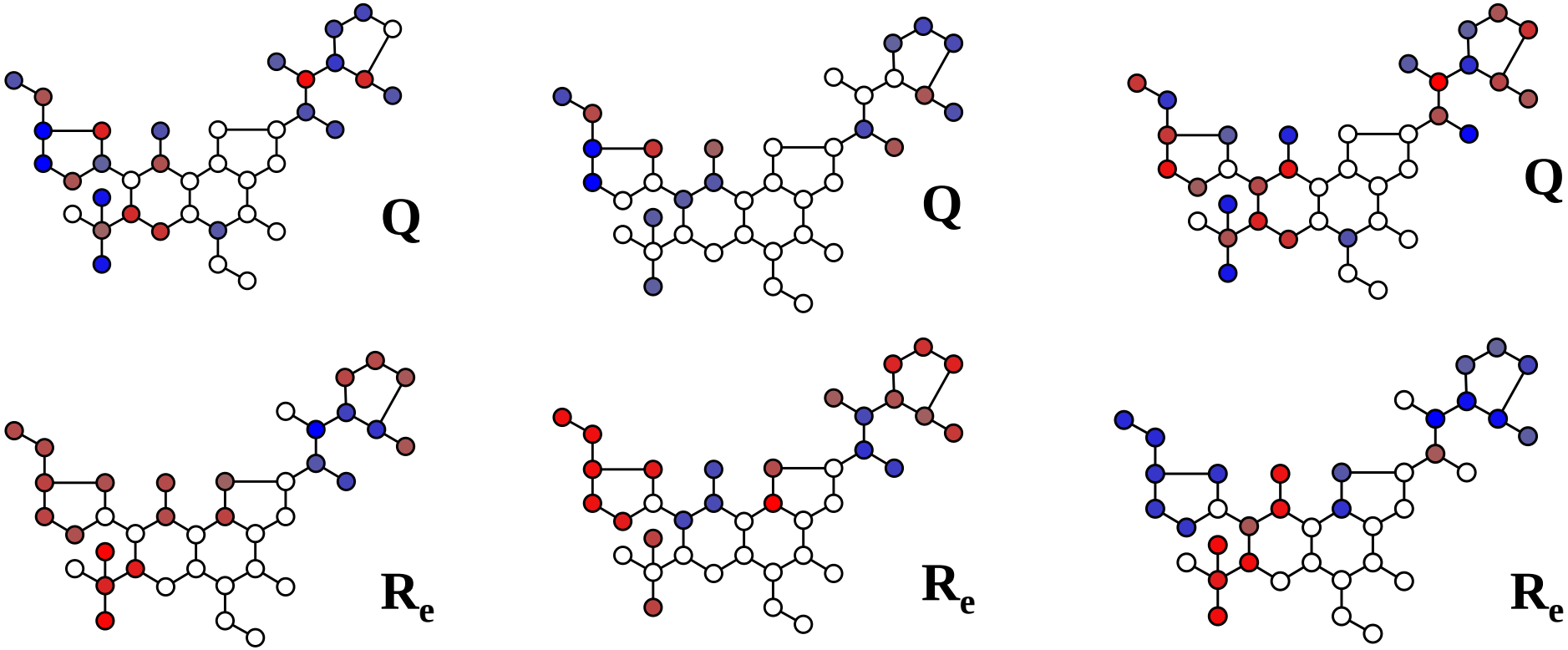
Training set: N=50, N_F=2, R^2=0.71, RMSE=0.38, Q^2=0.53, $RMSE_{cv}$=0.49

Selectivity $S=A_1-A_2$

Training set, N=50, N_F=2, R^2=0.69, RMSE=0.49, Q^2=0.45, $RMSE_{cv}$=0.66

$A_2=\log(1/K_i(\text{glycine/NMDA}))$

Training set, N=50, N_F=2, R^2=0.55, RMSE=0.54, Q^2=0.30, $RMSE_{cv}$=0.69



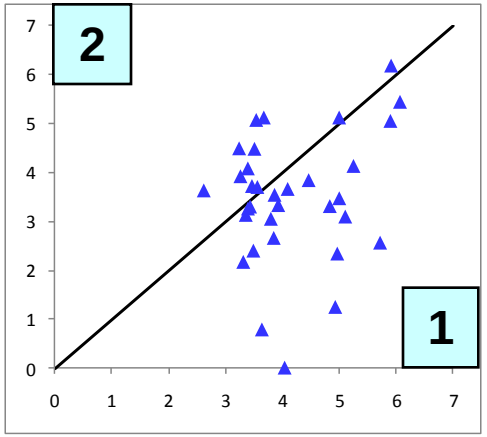
Red circles mean that the increase of a local descriptor value corresponds to the increase in activity, blue circles mean the opposite.

Multi-Target Selectivity: Options

Separate activity models

– selectivity calculated based on predicted activities

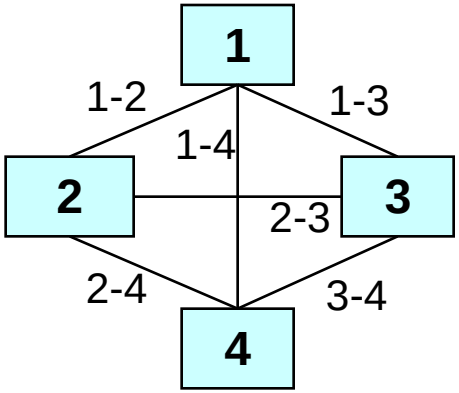
– factors controlling selectivity are difficult to identify



Models of pairwise selectivity (selectivity fields/maps in CoMFA/MFTA)

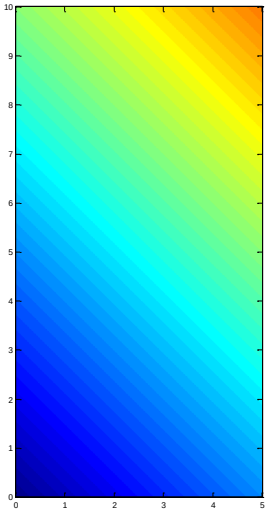
– more focused picture of factors controlling it

– rapidly increasing number of parameters
e.g. 4 activities → 6 selectivities → 10 endpoints total



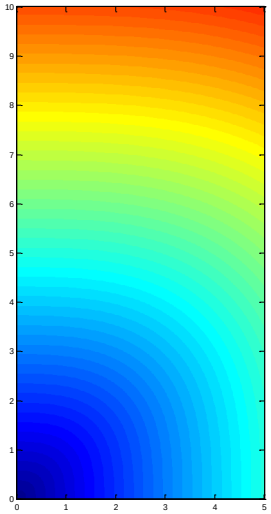
Multi-Target Activity.
Integral Activity Measures: Generalized Means.
Numerical Simulation

Arithmetic mean (M_1) (balances large and small values)



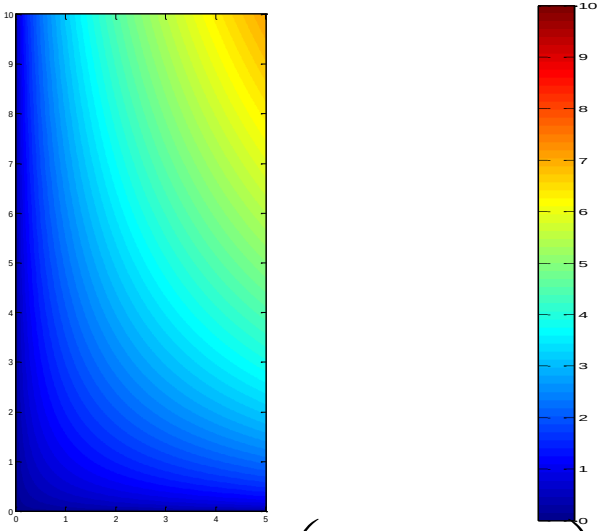
$$M_1 = \frac{1}{n}\sum_i x_i$$

Cubic mean (M_3) (dominated by larger values)



$$M_3 = \left(\frac{1}{n}\sum_i x_i^3\right)^{1/3}$$

Geometric mean (M_G) (simultaneous effect)

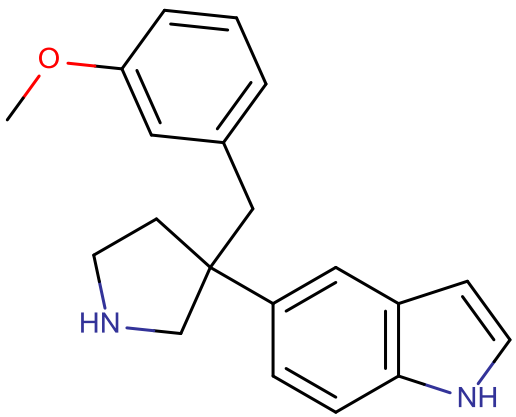


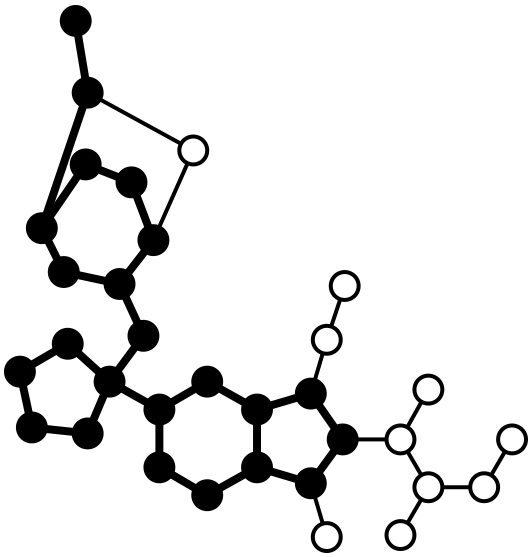
$$M_G = \sqrt[n]{\prod_i x_i} = \exp\left(\frac{1}{n}\sum_i \log x_i\right)$$

Triple Reuptake Inhibitors (TRI)

Inhibitors of serotonin transporter (SERT), norepinephrine (noradrenaline) transporter (NET) and dopamine transporter (DAT) as potential antidepressants

3,3-Disubstituted pyrrolidines





Bannwart L.M. et al., Novel 3,3-disubstituted pyrrolidines as selective triple serotonin / norepinephrine / dopamine reuptake inhibitors, *Bioorg. Med. Chem. Lett.*, 2008, 18 (23), 6062-6066

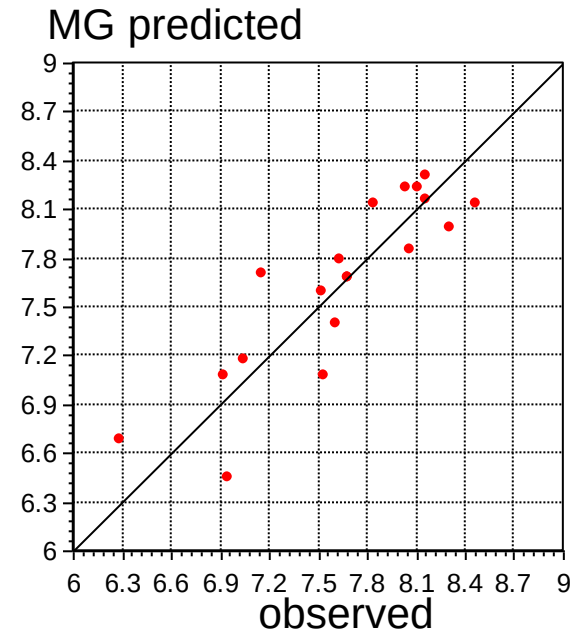
Triple Reuptake Inhibitors: MFTA

pKi SERT (Descriptors: Q, Re)
N = 19 N_F = 5 R^2 = 0.979 RMSE = 0.106 **Q^2 = 0.806** RMSEcv = 0.314

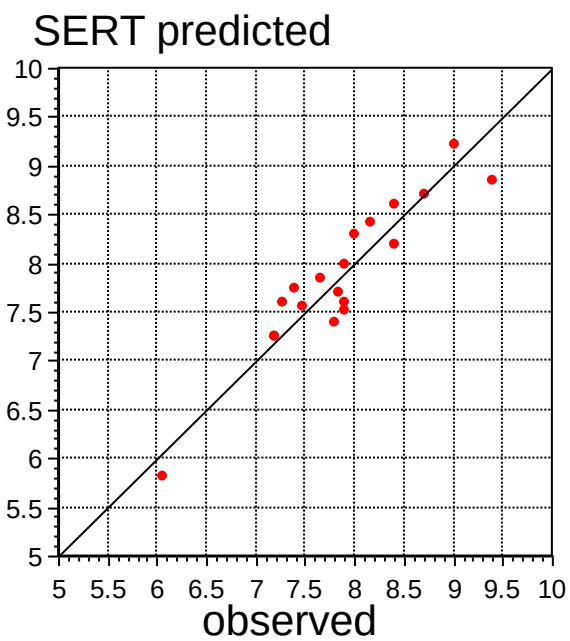
pKi NET (Descriptors: Q, Re)
N = 19 N_F = 3 R^2 = 0.827 RMSE = 0.274 **Q^2 = 0.643** RMSEcv = 0.392

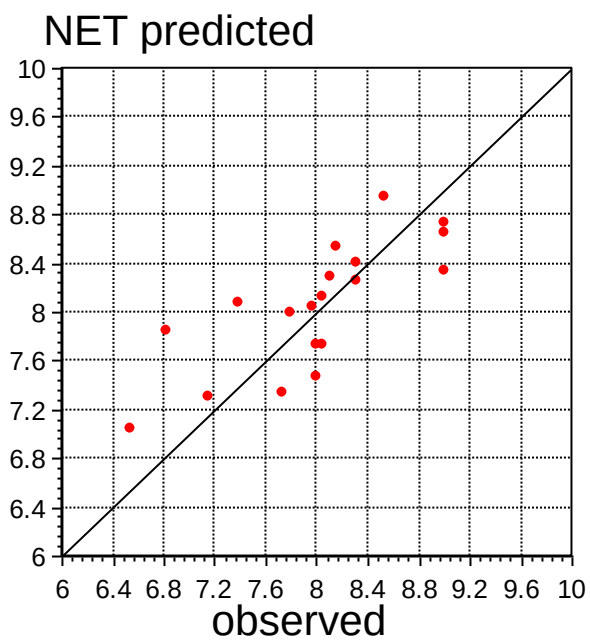
pKi DAT (Descriptors: Q, Re)
N = 19 N_F = 3 R^2 = 0.816 RMSE = 0.245 **Q^2 = 0.478** RMSEcv = 0.411

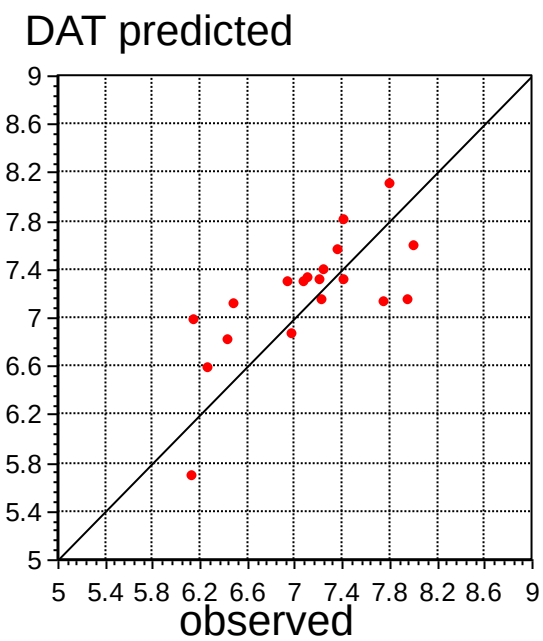
pKi MG (Descriptors: Q, Re)
N = 19 N_F = 5 R^2 = 0.985 RMSE = 0.067 **Q^2 = 0.766** RMSEcv = 0.263



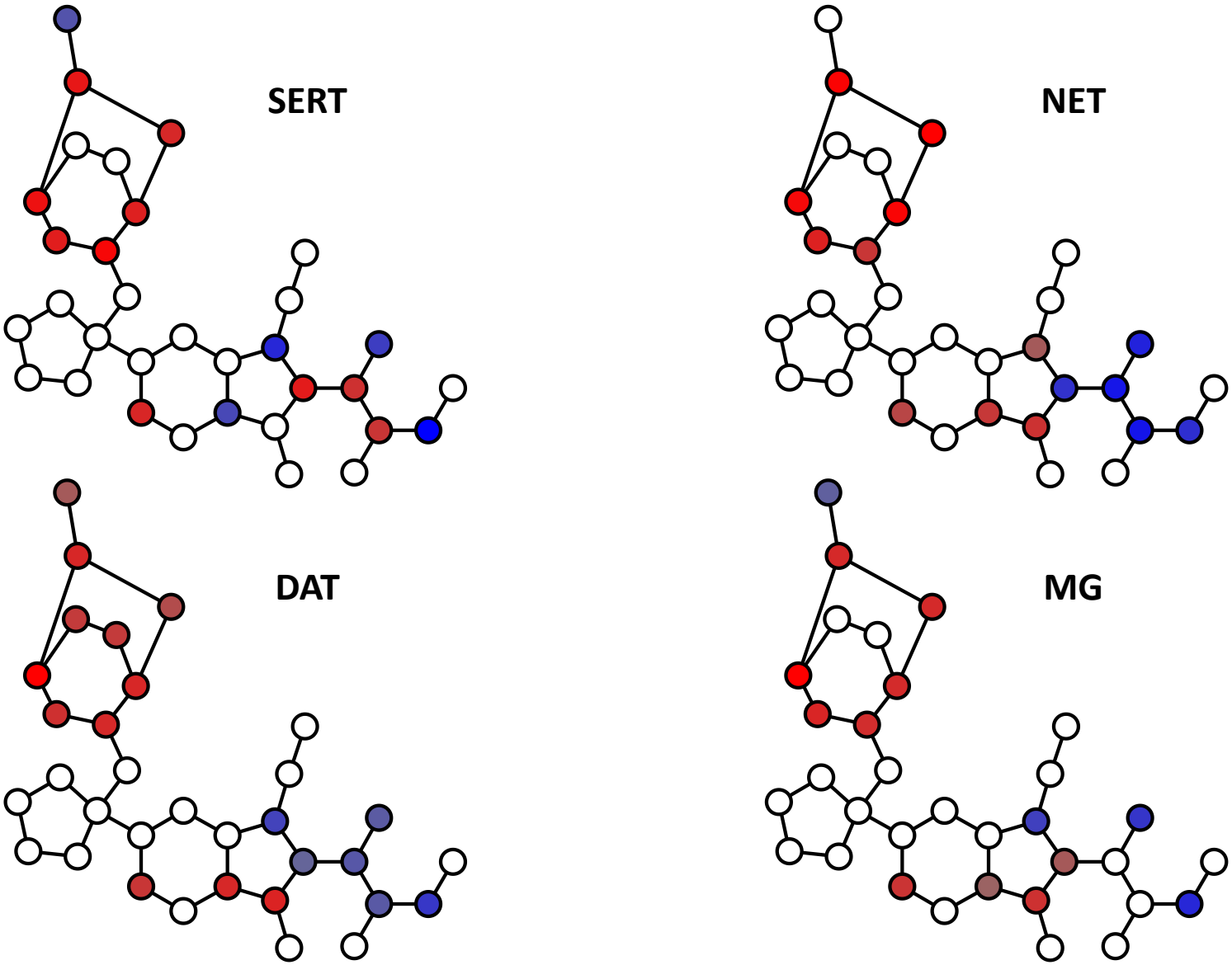
Plots of CV-predicted vs observed endpoints



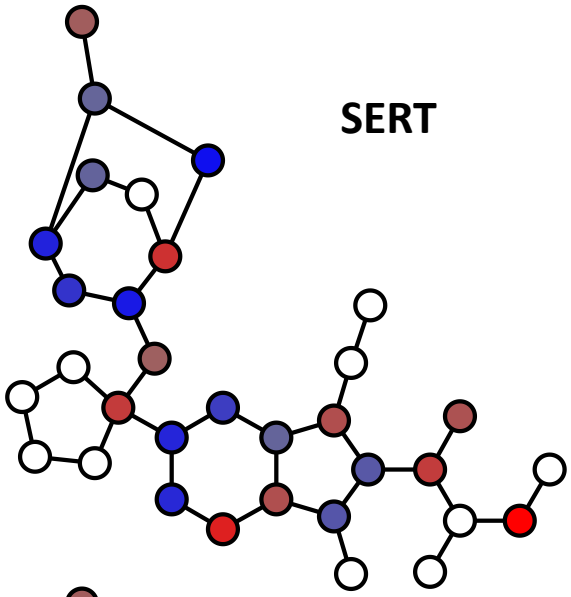


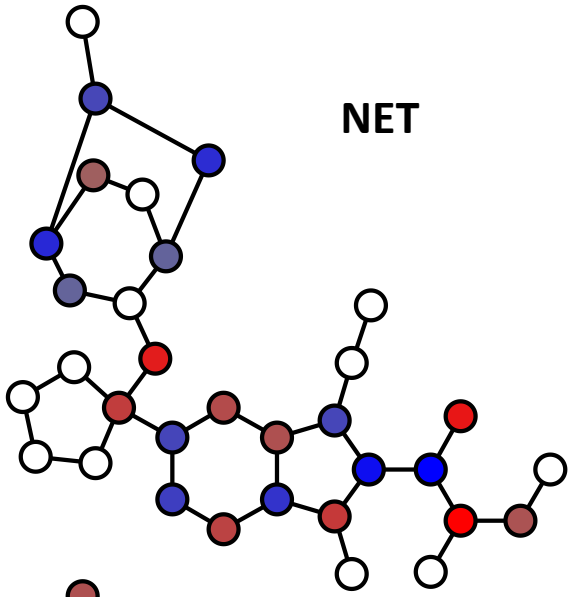


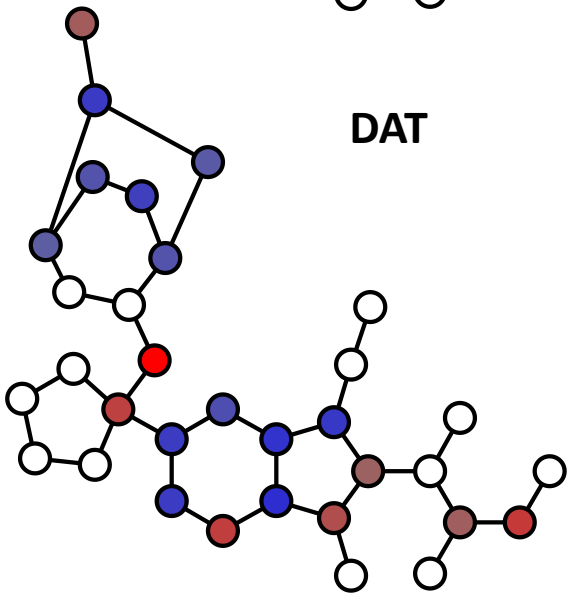
TRI Activity Maps: Steric Descriptors (Re)

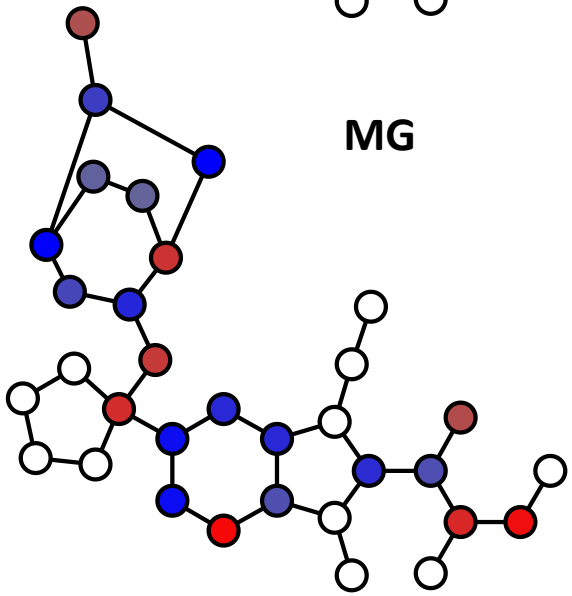


TRI Activity Maps: Partial Atomic Charges (Q)



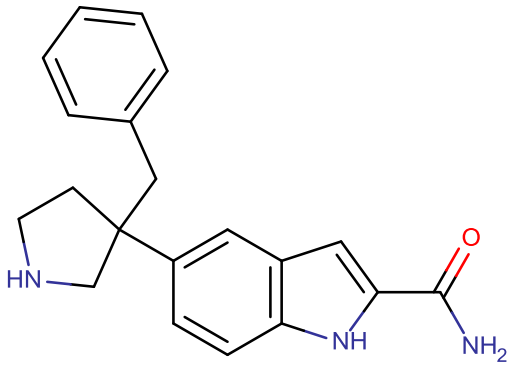


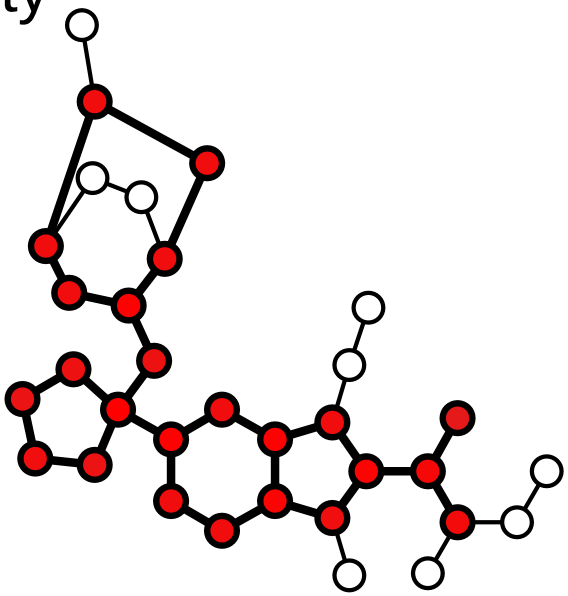


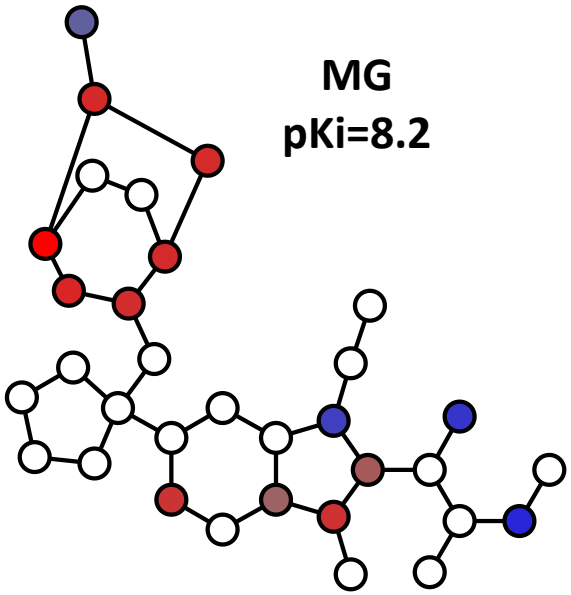


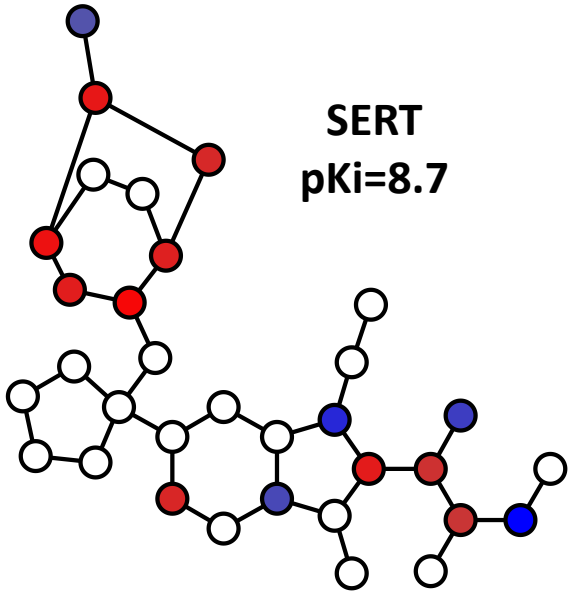
TRI Activity Maps: Steric Descriptors (Re)

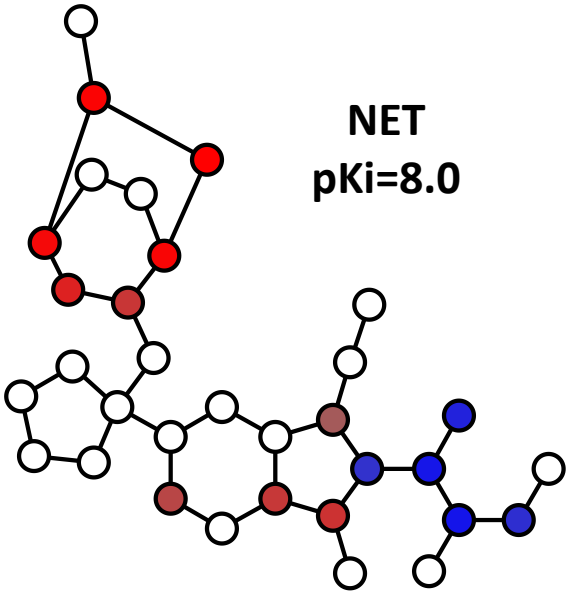
Balanced high TRI activity

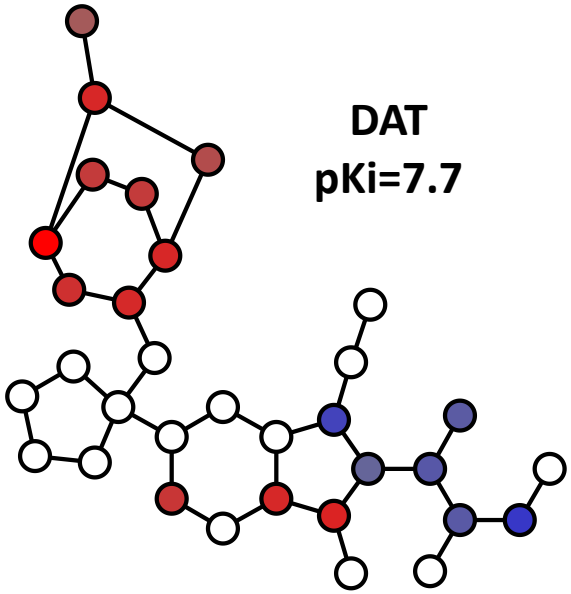






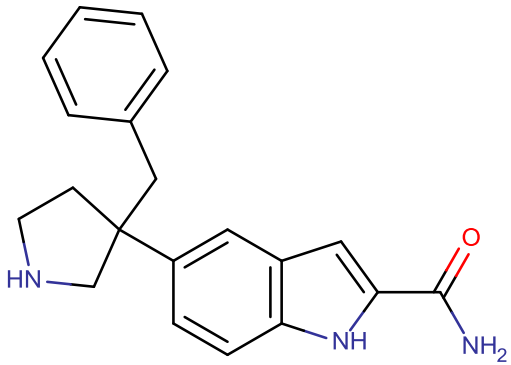


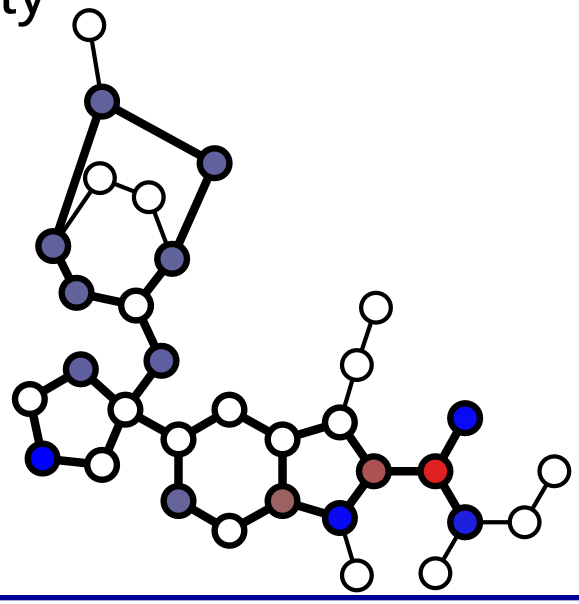


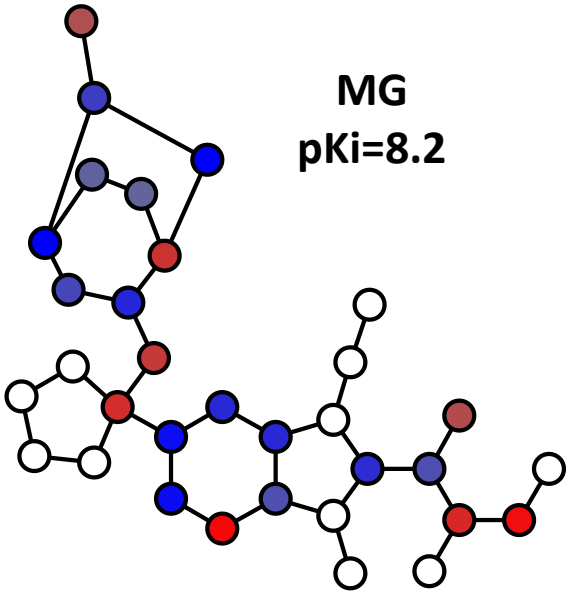


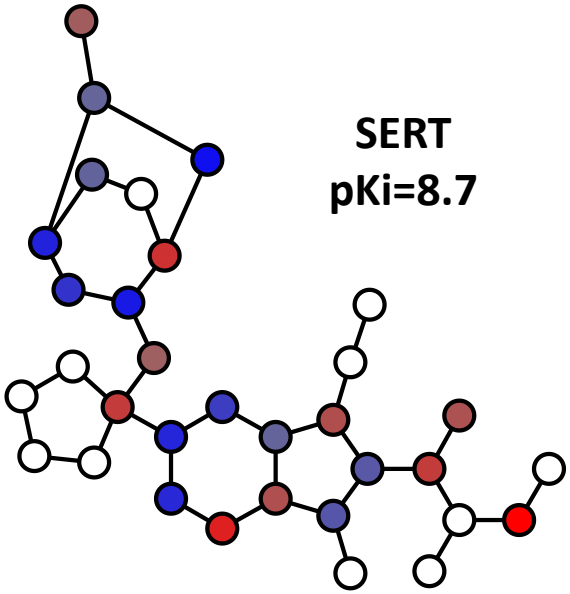
TRI Activity Maps: Partial Atomic Charges (Q)

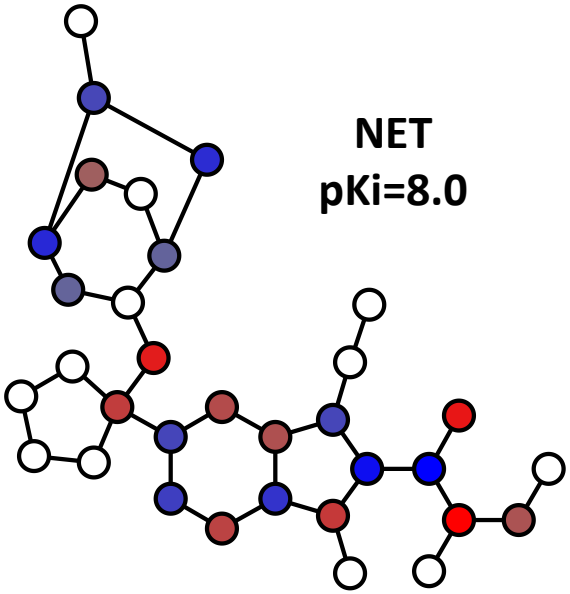
Balanced high TRI activity

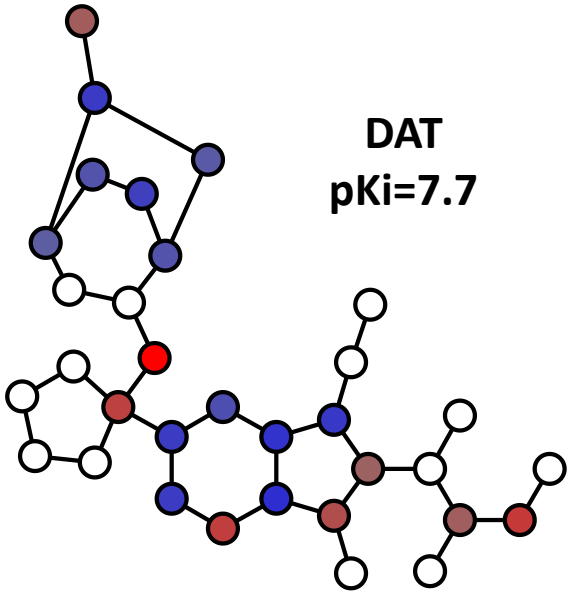






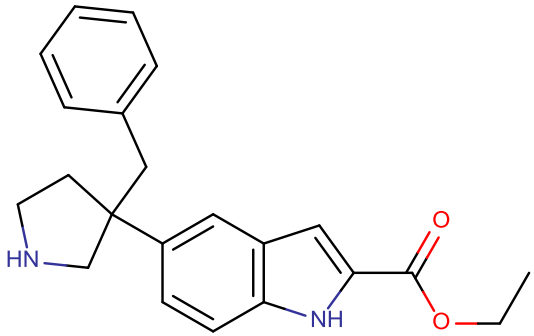


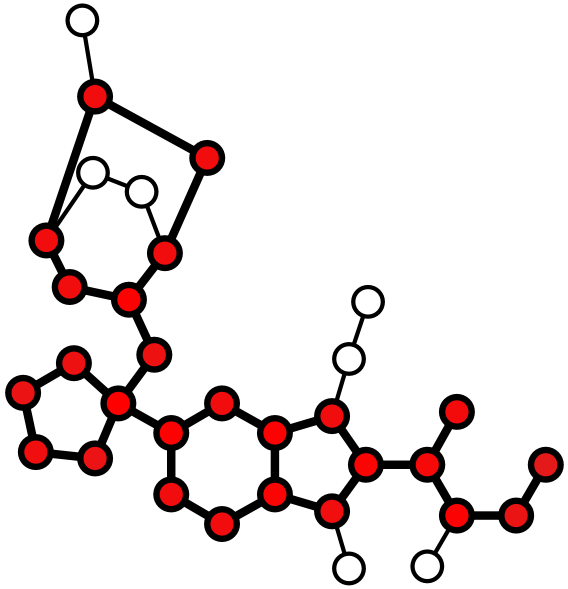


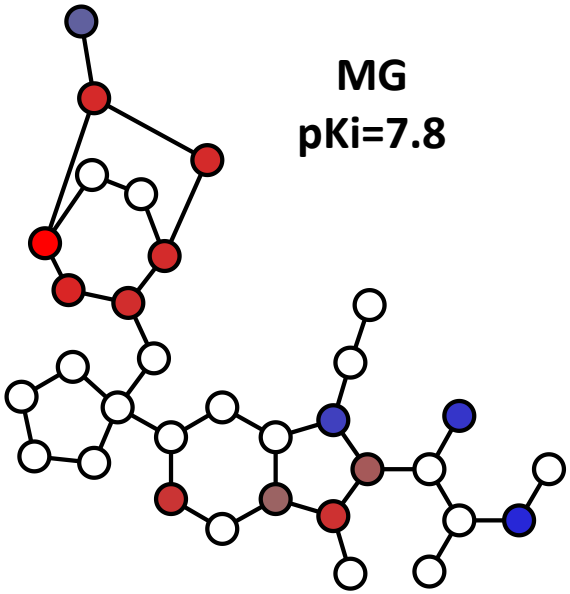


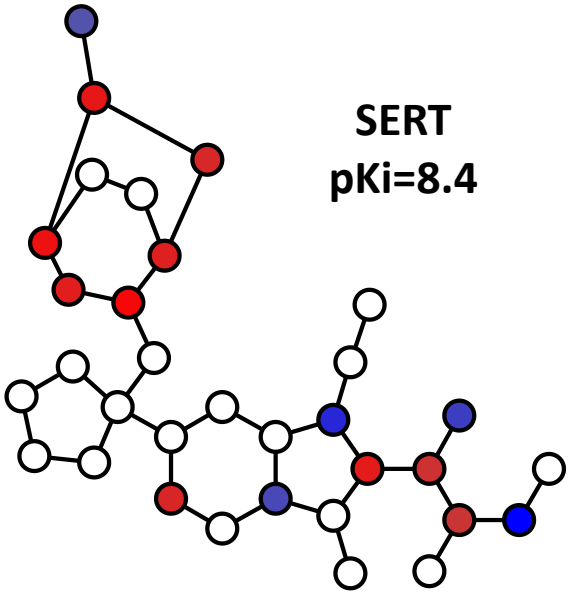
TRI Activity Maps: Steric Descriptors (Re)

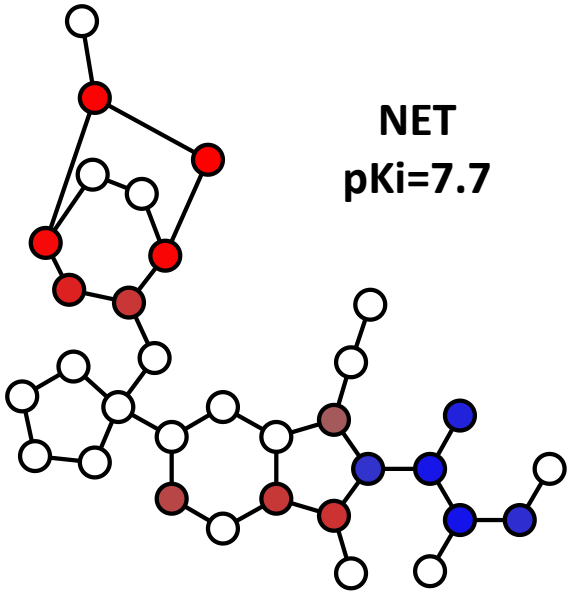
SERT-selective: SERT > NET > DAT

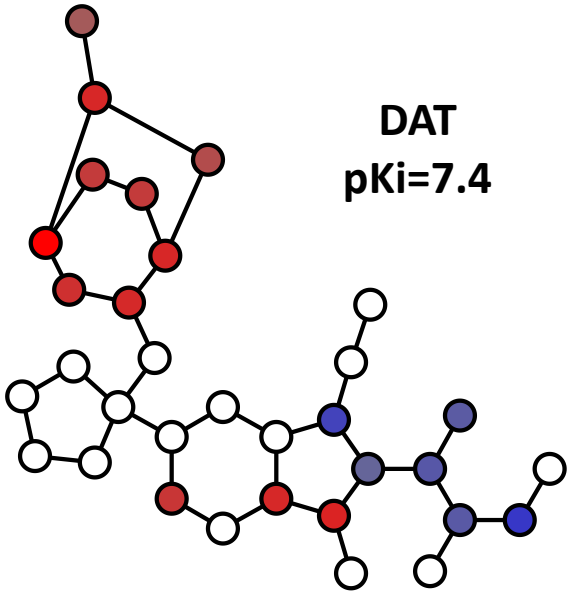






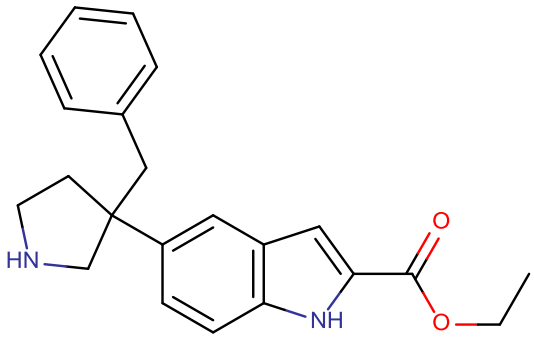


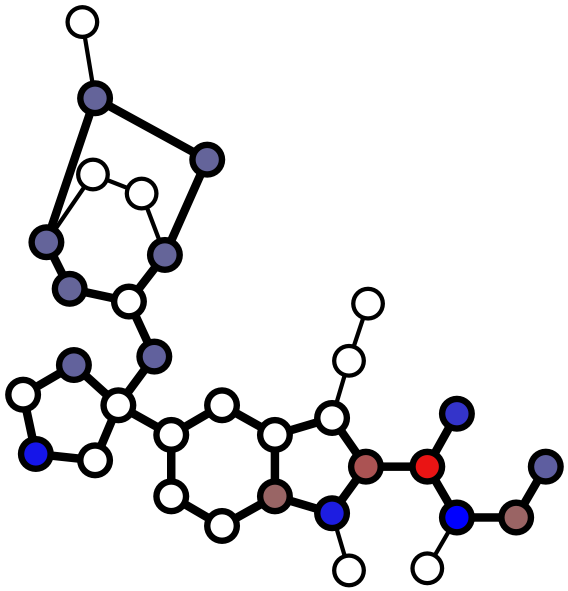


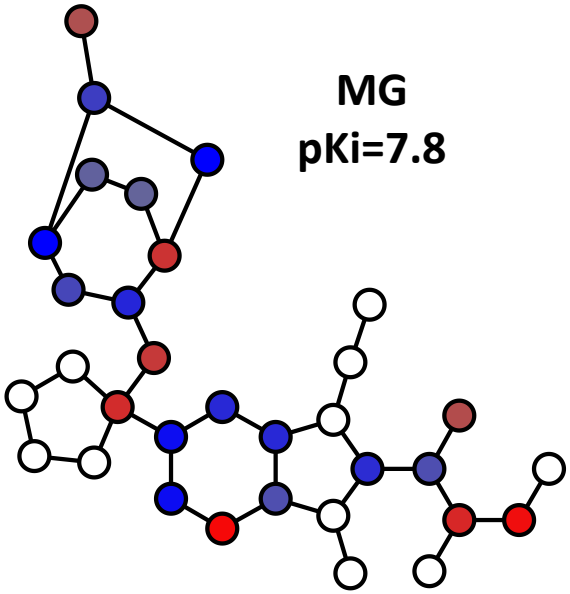


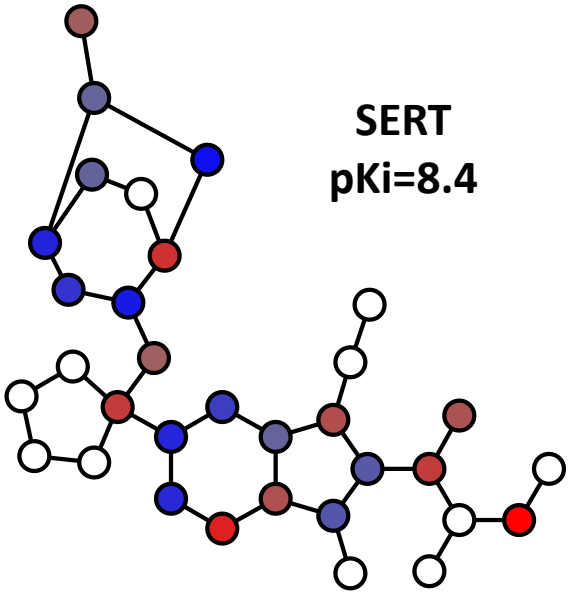
TRI Activity Maps: Partial Atomic Charges (Q)

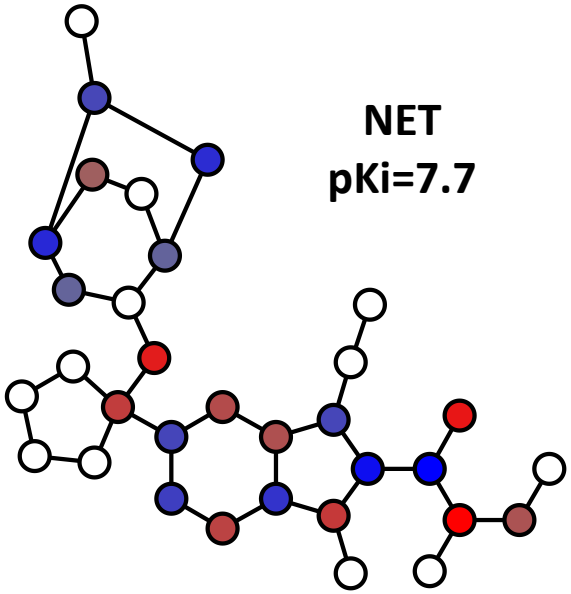
SERT-selective: SERT > NET > DAT

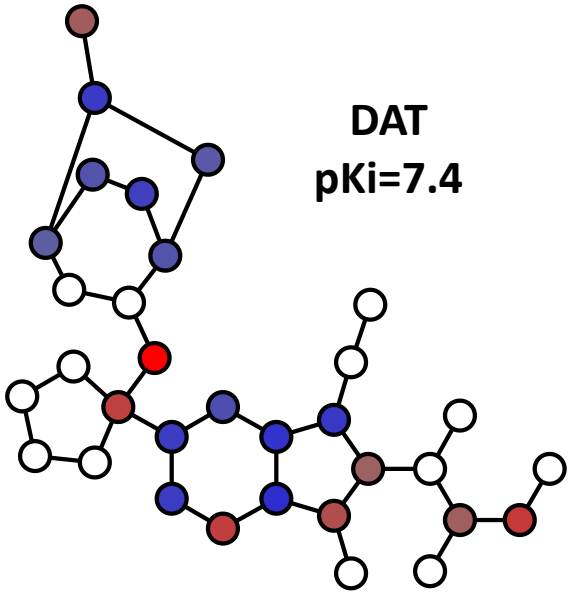












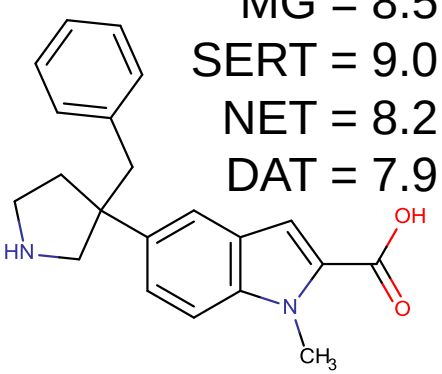
Triple Reuptake Inhibitors: Virtual Screening

- *Representative set of 5000 structures of the same class was generated from the MFTA supergraph in stochastic mode*

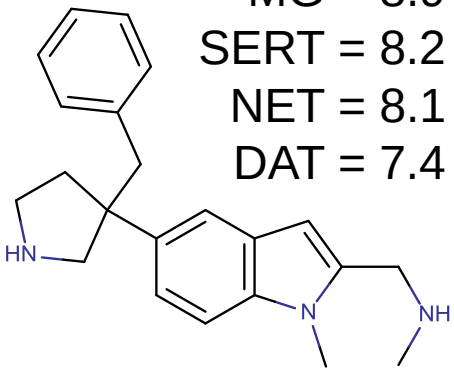
- *Endpoints were predicted using MFTA MG models and then individual models*

- *14 compounds have predicted pKiMG value > 8.5*

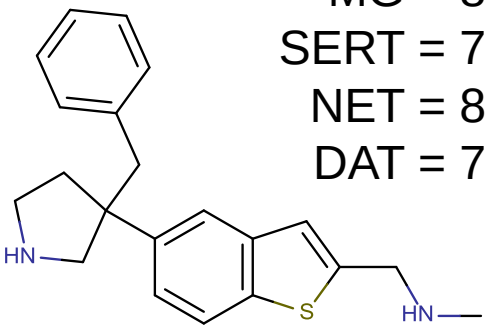
Pred pKi
MG = 8.5
SERT = 9.0
NET = 8.2
DAT = 7.9



Pred pKi
MG = 8.9
SERT = 8.2
NET = 8.1
DAT = 7.4



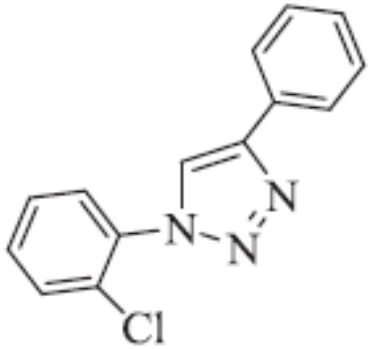
Pred pKi
MG = 8.7
SERT = 7.9
NET = 8.1
DAT = 7.6



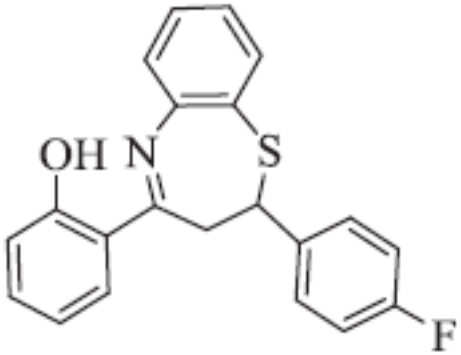
Dual inhibition of the α-glucosidase and butyrylcholinesterase

Common pharmacophore for large structurally diverse set of compounds

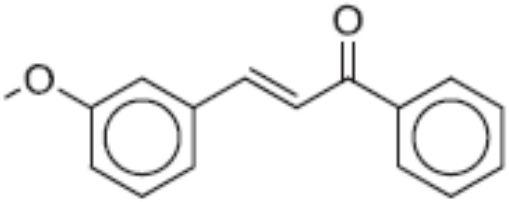
1,2,3-Triazoles



Benzothiazepines



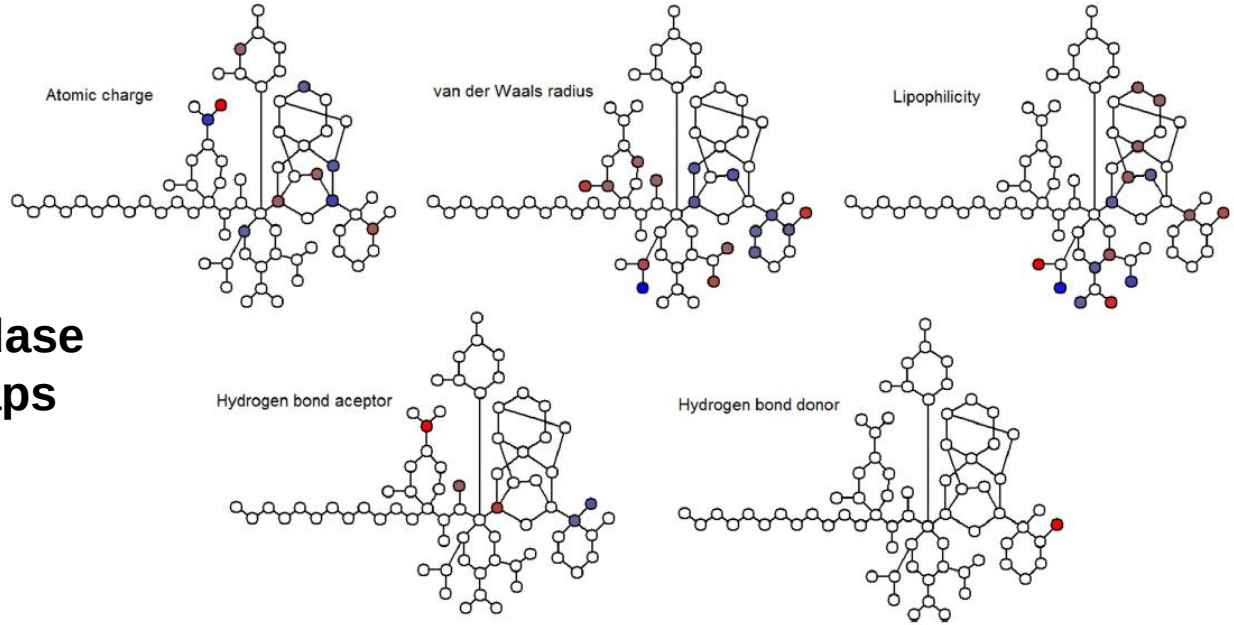
Chalcones



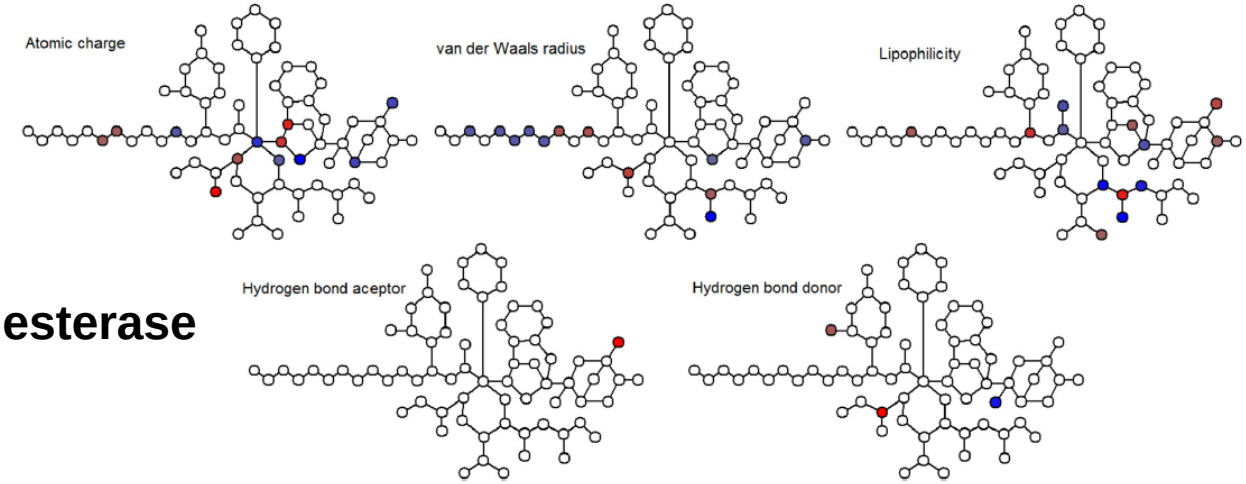
Jabeen F., Oliferenko P.V., Oliferenko A.A. et al., Dual inhibition of the α-glucosidase and butyrylcholinesterase studied by Molecular Field Topology Analysis, Eur. J. Med. Chem., 2014, 80, 228-242

Dual inhibition of the α-glucosidase and butyrylcholinesterase

α-Glucosidase activity maps



Butyrylcholinesterase activity maps



Acknowledgements

This work was supported by grants of

Russian Foundation for Basic Research,

Russian Academy of Sciences,

Ministry of Education and Science of Russian Federation

THANK YOU

FOR YOUR ATTENTION!