Molecular Fields in QSAR

Eugene V. Radchenko
Department of Chemistry, Lomonosov Moscow State University

Structure–activity relationships

- **«Structure of a compound determines its properties, including biological activity»**

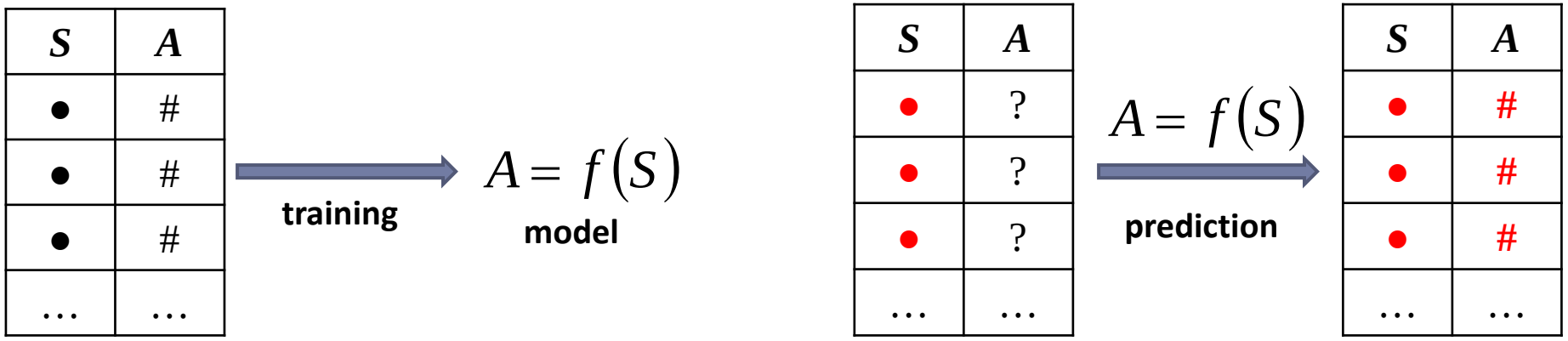
$$A = f(S)$$

  - Follows from structural theory of organic chemistry
  - **Quantitative Structure-Activity Relationships (QSAR)**
  - **Quantitative Structure-Property Relationships (QSPR)**

- Mutually complementing approaches
  - **Structure-based**: models of the **biotarget structure and its interactions with ligands**
  - **Ligand-based**: data on **known ligands and their activities** usually analyzed using statistical learning techniques

Quantitative Structure-Activity Relationships

- **Model** is derived from analysis of available experimental activity data for compounds comprising a **training set**
- Allows to **predict** (estimate) activity for new compounds
- Needs sufficiently broad **applicability domain**
- Desirable: **interpretation** / explanation of a model *wrt* mechanism of action and structural features significant for activity



training set

Structure representation: descriptors

$$A = f(S) = f(D_1, D_2, D_3, \ldots)$$

- Statistical learning usually needs numerical data
- Molecular (structural) **descriptors** – numerical parameters describing *certain* features and facets of the structure of a compound
- Generally, the full diversity of chemical space would require infinite number of descriptors
- Only need description sufficient for a **specific problem**

D_1	D_2	D_3	…	A
#	#	#	…	#
#	#	#	…	#
#	#	#	…	#
…	…	…	…	…

training set

training

$$A = f(D_1, D_2, D_3, \ldots)$$

model

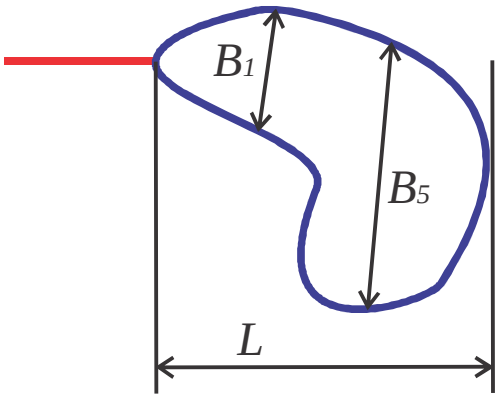
Molecular (structural) descriptors

$$A = f(S) = f(D_1, D_2, D_3, \ldots)$$

- Numerical parameters representing certain features and facets of the structure of a compound
- Almost unlimited diversity of potentially available descriptors
- Desirable: connection to mechanism of action, interpretability
- Classification of descriptors
  - Topological
  - Physico-chemical
  - Substructural
  - **Superstructural**
  - **3D structural**

Physico-chemical descriptors

- **1D Steric descriptors**
  - Molecular mass
  - Taft's substituent constant E_s
  - Molecular dimensions
  - Moments of inertia
  - Molecular volume
  - Molecular surface area (true 3D or topology-based estimate)
  - Polar surface area PSA
  - Molecular shape
  - Substituent STERIMOL parameters
    - L – substituent length along connecting bond
    - B1 – minimum width perpendicular to bond
    - B5 – maximum width perpendicular to bond

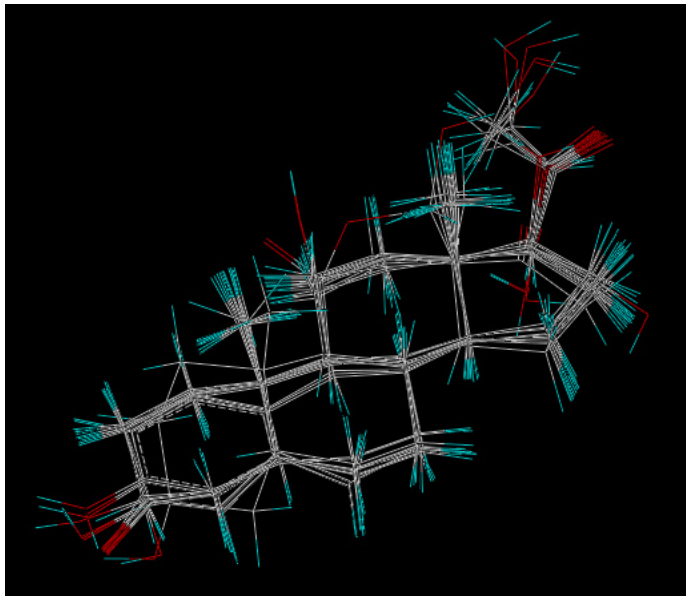


3D QSAR

- Desire for direct QSAR analysis of the 3D molecular structures and interactions
- Especially for activities mediated by specific ligand-biotarget binding
- **Comparative Molecular Field Analysis (CoMFA)**
  - *De facto* standard 3D QSAR method
  - R. Cramer et al., 1988
  - Implemented in SYBYL software, patent now expired
  - **Molecular mechanics force fields** adequately model intermolecular interactions
  - Uniform descriptor matrix can be obtained by **sampling molecular interaction fields over rectangular 3D grid**

CoMFA: alignment

- Alignment of ligand 3D structures
- Requires **chemical / mechanistic consistency**
- The **most problematic step** in CoMFA study
- **Manual alignment by common substructure** – requires consistent conformations
- Automatic 'field fit' alignment – proposed but never actually used
- External alignments
  - X-ray data for ligand-target complexes
  - Pharmacophore-based alignment
  - Docking-based alignment – pose and conformation uncertainty

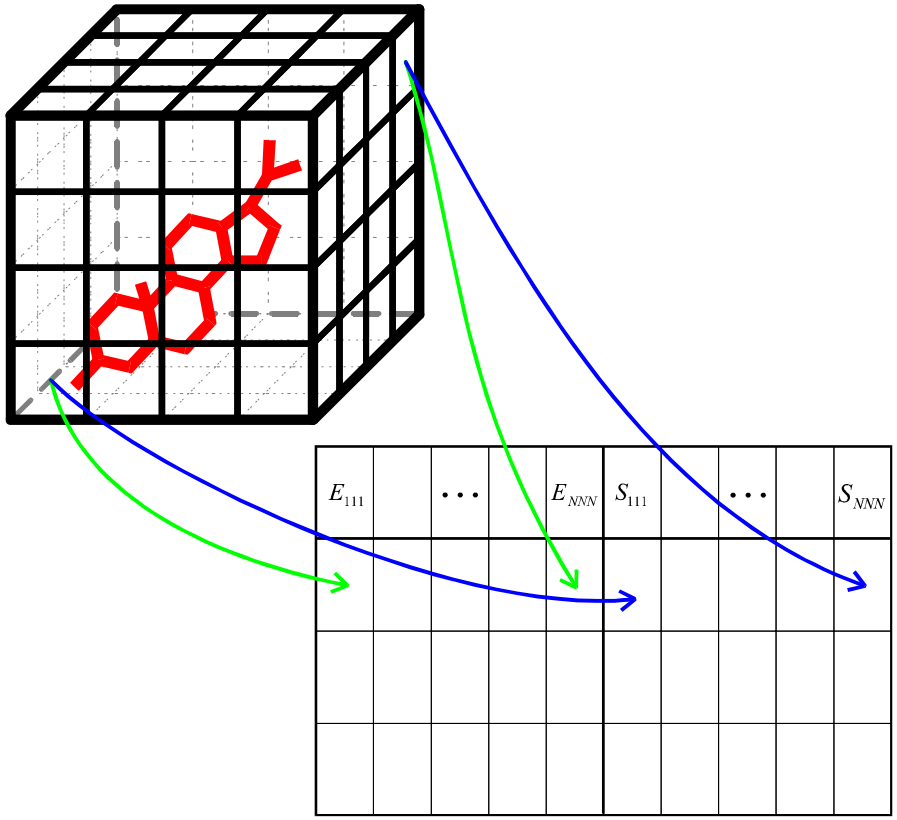


CoMFA: descriptors

- **Electrostatic and steric intermolecular interaction energies** (molecular interaction fields) sampled in nodes of a **rectangular 3D grid**
- Probe atom (commonly CH_3^+)

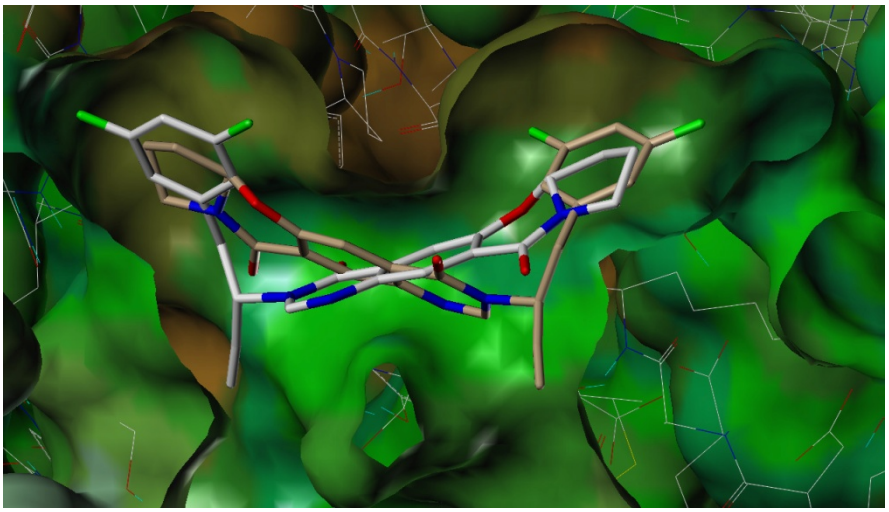
$$E = \frac{1}{4\pi\varepsilon_0\varepsilon}\sum_i \frac{q_i\, q_p}{d_{ip}}$$

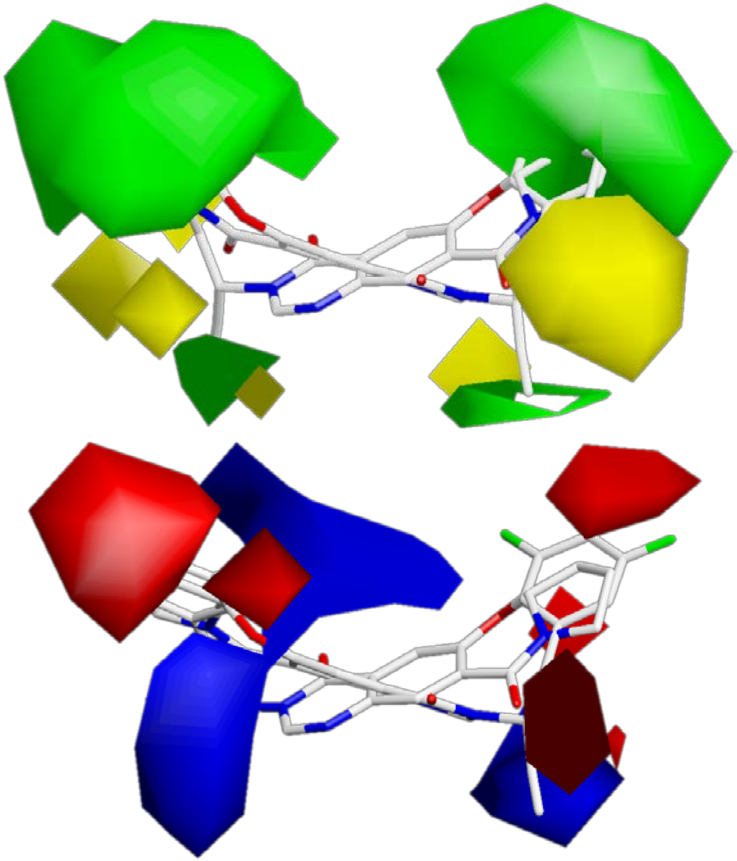
$$S = \sum_i \left[-\frac{A_{ip}}{d_{ip}^6} + \frac{B_{ip}}{d_{ip}^{12}} \right]$$



CoMFA: analysis

- Statistical modeling: partial least squares regression (PLSR)
- Predictive models
- Activity maps
- Design of better structures
- Interpretation, comparison to target structure





Steric: green – favorable, yellow – unfavorable interaction
Electrostatic: favorable red – negative, blue – positive charge

CoMFA: alignment problem

- Difficult and tedious
- Especially for flexible structures
- **Formal rules/procedures**
- Quasi-topological ("topological in 3D") models

- Topomer CoMFA
  - Standard conformations/rules for various groups
  - E.g., fully extended alkyl chains
- Template CoMFA
  - Conformation templates based on X-ray data
  - + Standard rules
  - Promising preliminary results but no broad application

CoMFA extensions

- Additional molecular fields
  - Local lipophilicity: **Molecular Lipophilic Potential (MLP)**

$$MLP = \sum f_i \, e^{-d_{ip}/2}$$

  - Orbital densities (HOMO, LUMO)

3D QSAR redefinitions

- **Comparative Molecular Similarity Indices Analysis (CoMSIA)**
  - G. Klebe et al., 1994
  - Softer field descriptors, no singularities, less sensitive to misalignment
  - **Gaussian functions** for similarity to probe index

$$A_F = -\sum f_i \, f_p \, e^{-\alpha d_{ip}^2}$$

- Additional fields
  - Electrostatic
  - Steric
  - Hydrophobic
  - Hydrogen bond donor
  - Hydrogen bond acceptor

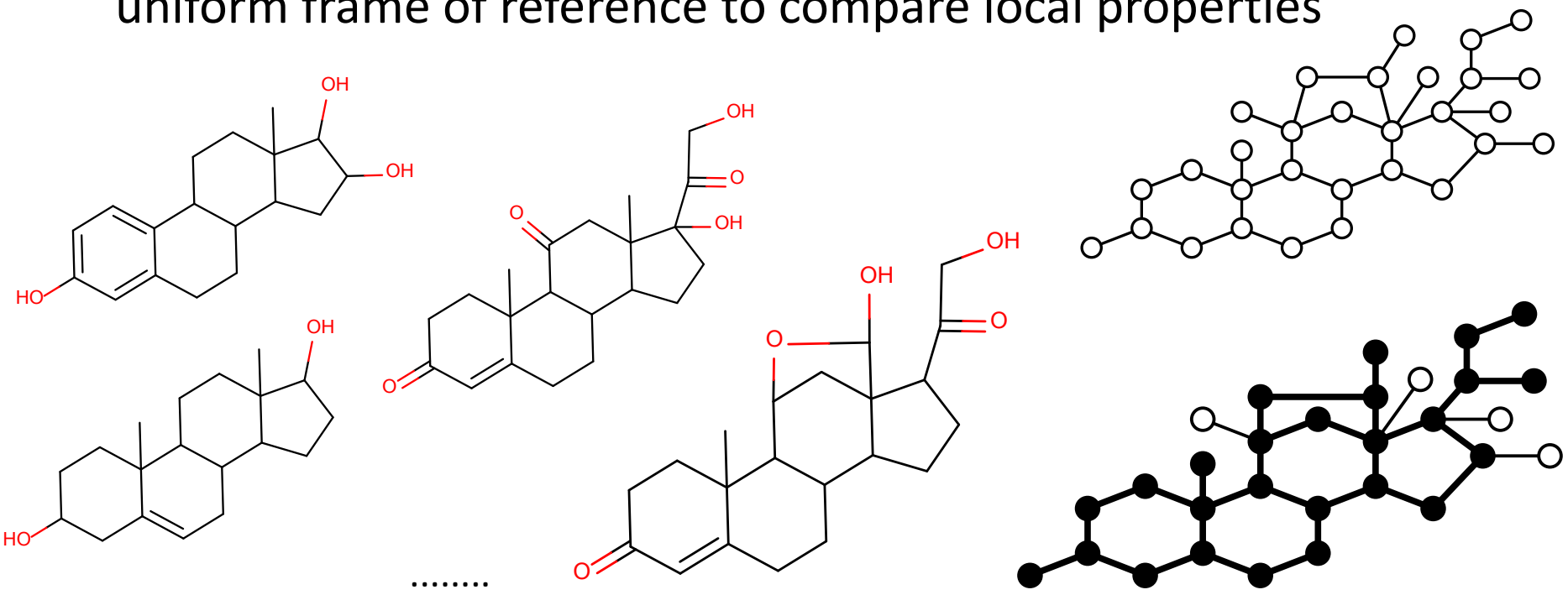
3D QSAR future

- CoMFA patent expired
- Development halted, product discontinued by Certara

- Basic CoMFA workflow *(no topomer/template features)* can be performed in other software (with varying usability)
  - Open3DQSAR
  - Schrödinger
  - Cresset

Superstructural approaches in QSAR

- Topological methods, free from 3D alignment problems
- Account for mutual arrangement of fragments and local structural features
- **Molecular Field Topology Analysis (MFTA)**
- Structural (2D) formulas alignment → **molecular supergraph** – uniform frame of reference to compare local properties



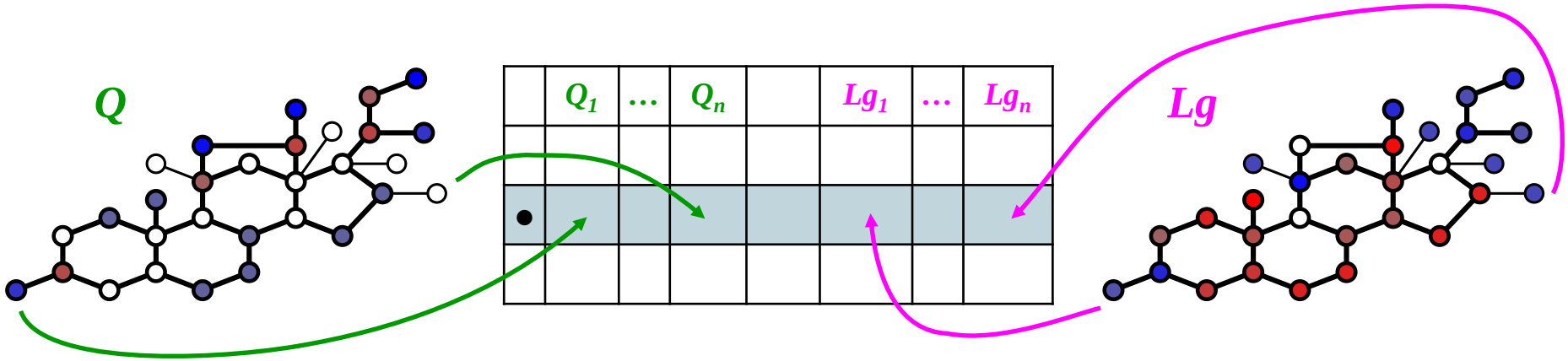
.....

MFTA: descriptors

- Local physico-chemical properties reflect various ligand-target interactions
  - Electrostatic: effective atomic charge $\boldsymbol{Q}$
  - Steric: atom/environment van der Waals radius $\boldsymbol{R}$, $\boldsymbol{Re}$
  - Group lipophilicity $\boldsymbol{Lg}$
  - Hydrogen bond donor/acceptor ability $\boldsymbol{Hd}$, $\boldsymbol{Ha}$

MFTA: descriptor matrix

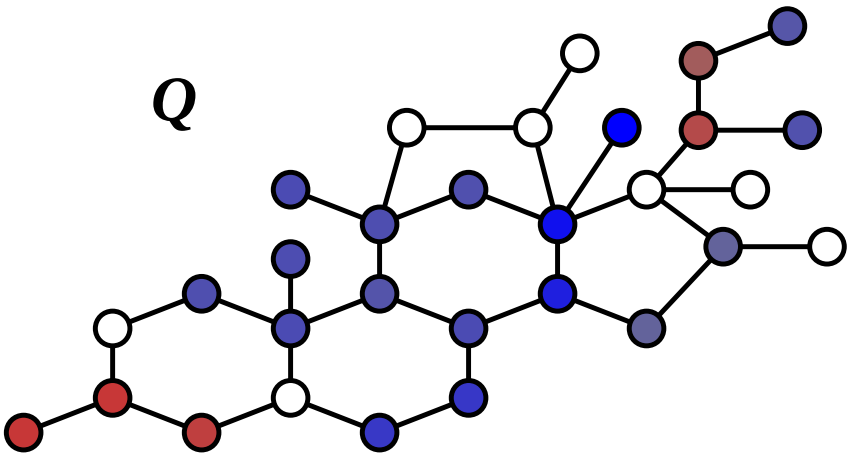
- Descriptor matrix
  - Occupied supergraph positions – atomic descriptors
  - Unoccupied positions – neutral values

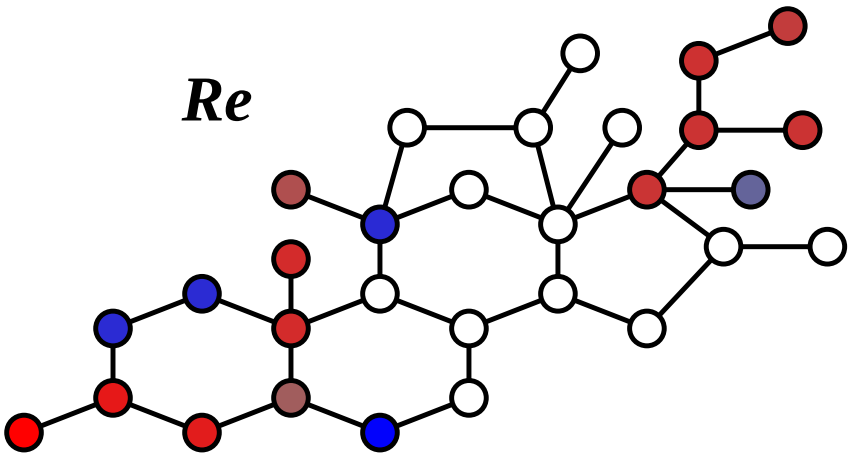


Red: positive Blue: negative

MFTA: analysis

- Statistical modeling: partial least squares regression (PLSR)
- Predictive models
- Activity maps – descriptor influence on activity
- Comparison to biotarget structure
- Design of better structures





Red: increase Blue: decrease

Some references

1. Cramer R.D., Patterson D.E., Bunce J.D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, *J. Am. Chem. Soc.*, 1988, **110** (18), 5959–5967.
2. Cramer R.D. Topomer CoMFA: A design methodology for rapid lead optimization, *J. Med. Chem.*, 2003, **46** (3), 374–388.
3. Cramer R.D., Wendt B. Template CoMFA: The 3D-QSAR grail?, *J. Chem. Inf. Model.*, 2014, **54** (2), 660–671.
4. Cramer R.D. Template CoMFA generates single 3D-QSAR models that, for twelve of twelve biological targets, predict all ChEMBL-tabulated affinities, *PLoS One*, 2015, **10** (6), e0129307.
5. Klebe G., Abraham U., Mietzner T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity, *J. Med. Chem.*, 1994, **37** (24), 4130–4146.
6. Palyulin V.A., Radchenko E.V., Zefirov N.S. Molecular Field Topology Analysis method in QSAR studies of organic compounds, *J. Chem. Inf. Comp. Sci.*, 2000, **40** (3), 659–667.
7. Radchenko E.V., Palyulin V.A., Zefirov N.S. Molecular Field Topology Analysis in drug design and virtual screening, *Chemoinformatics Approaches to Virtual Screening*, ed. A. Varnek and A. Tropsha, RSC, 2008, 150–181.
8. Radchenko E.V., Makhaeva G.F., Palyulin V.A., Zefirov N.S. Chemical similarity, shape matching and QSAR, *Computational Systems Pharmacology and Toxicology*, RSC, Cambridge, 2017, 120–173.
9. Radchenko E.V., Palyulin V.A., Zefirov N.S. Molecular Field Topology Analysis (MFTA) in the design of neuroprotective compounds, *Computational Modeling of Drugs Against Alzheimer's Disease*, Springer, 2018, 139–159.