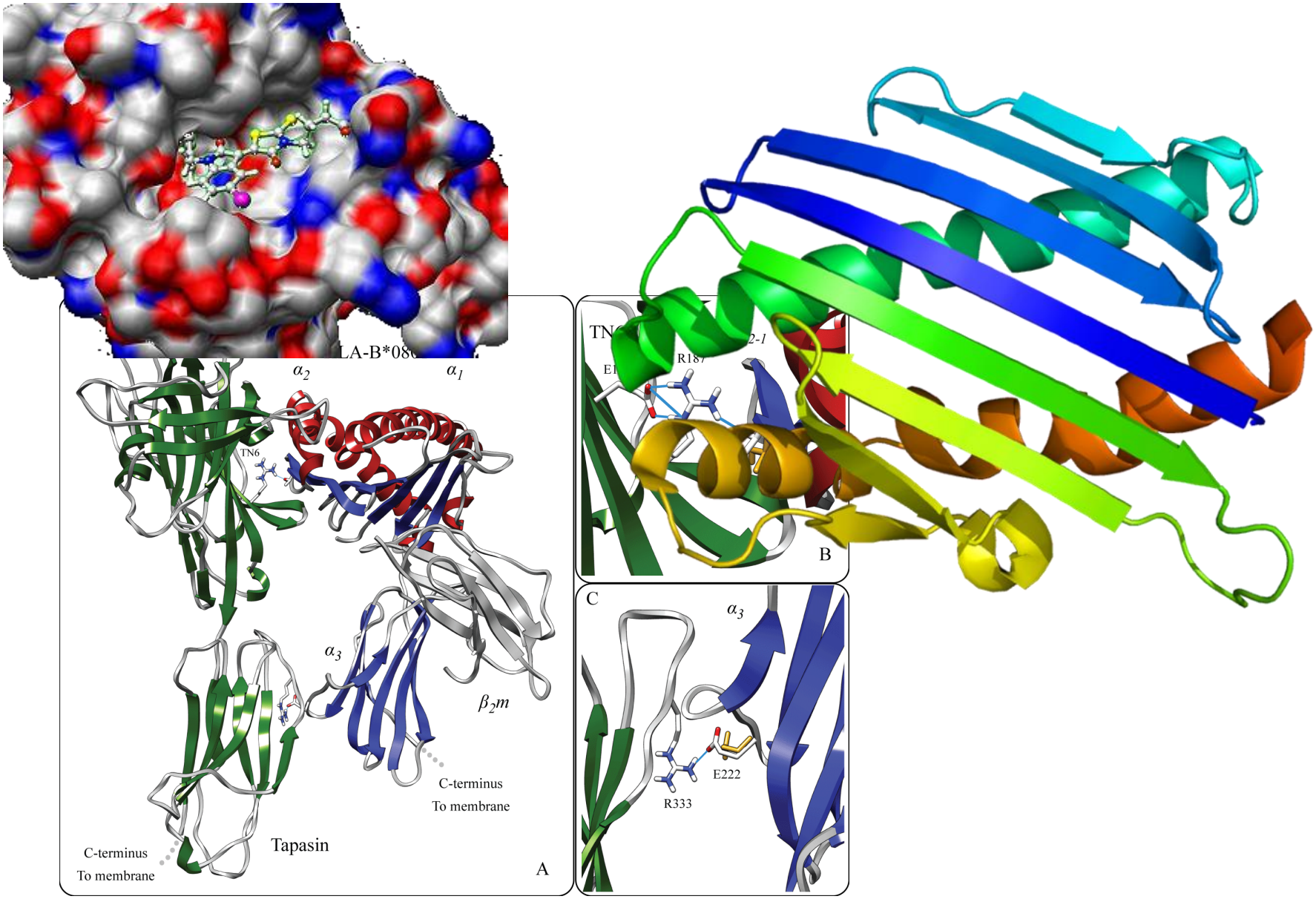
University of Silesia, Katowice, Poland
11 – 22 March 2013

Molecular docking

Dr. Pavel Polishchuk

A.V. Bogatsky Physico-Chemical Institute
of National Academy of Sciences of Ukraine
Odessa, Ukraine
pavel_polishchuk@ukr.net

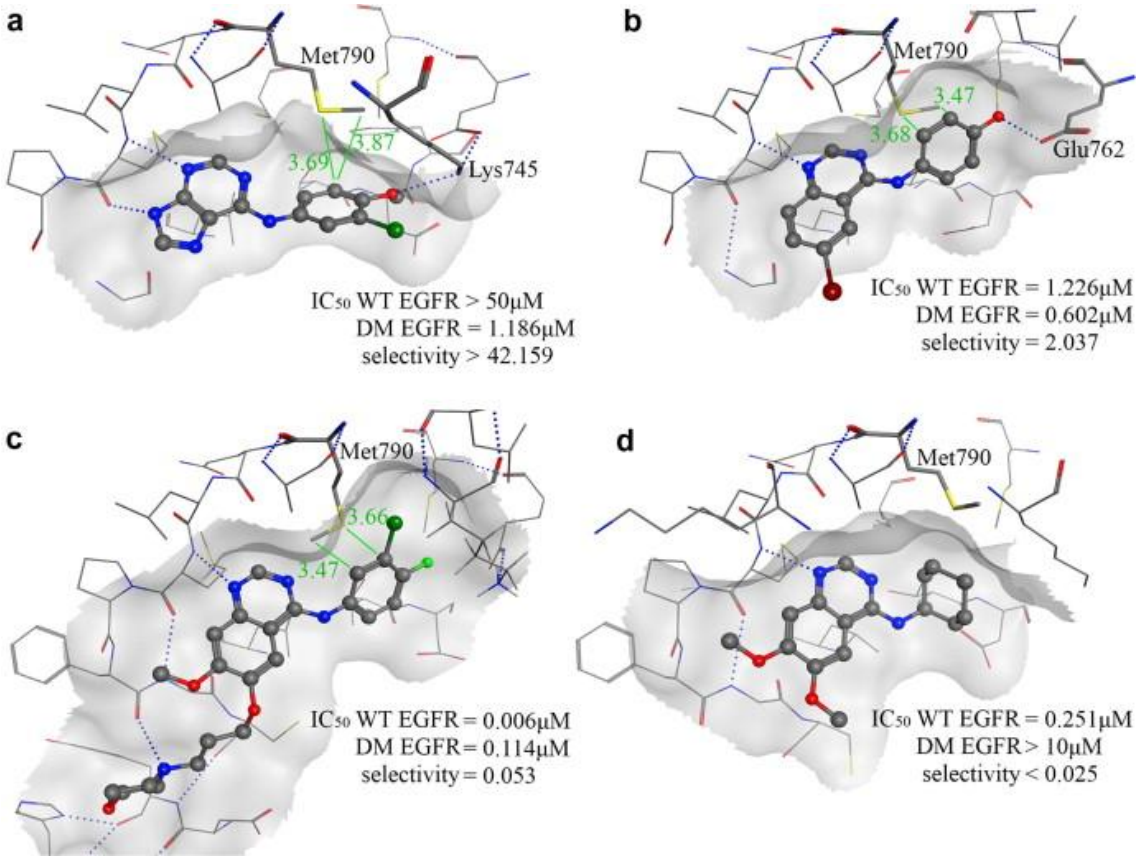
What is molecular docking?



Docking is the *in silico* tool which predicts

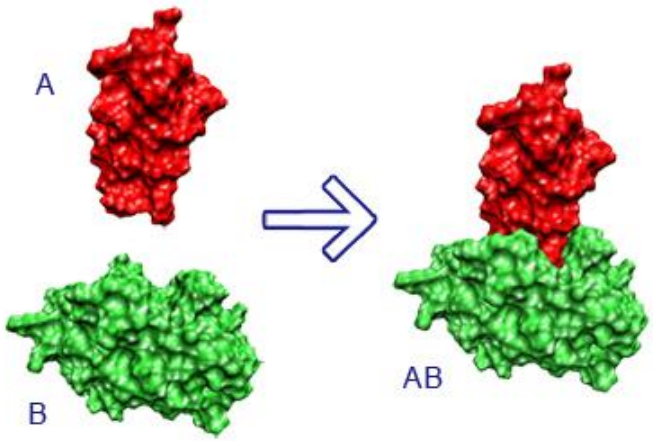
Pose – a possible relative orientation of a ligand and a receptor as well as conformation of a ligand and a receptor when they are form complex

Score – the strength of binding of the ligand and the receptor.



Bioorganic & Medicinal Chemistry, 20 (12), 2012, 3756–3767.

Why docking is a “problem”?



Complex 3D jigsaw puzzle

Conformational flexibility

Mutual adaptation (“induced fit”)

Solvation in aqueous media

Complexity of thermodynamic contribution

No easy route to evaluation of ΔG

Simplification and heuristic approaches are necessary

“At its simplest level, this is a problem of subtraction of large numbers, inaccurately calculated, to arrive at a small number.”

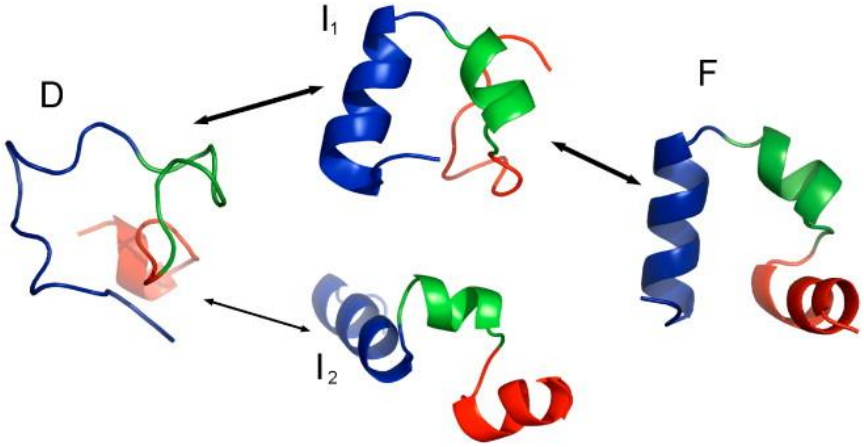
(Leach A.R., Shoichet B.K., Peishoff C.E.. *J. Med. Chem.* 2006, *49*, 5851-5855)

Sampling and scoring

Protein-ligand docking software consists of two main components which work together:

1. **Search algorithm (sampling)** - generates a large number of poses of a molecule in the binding site.

2. **Scoring function** - calculates a score or binding affinity for a particular pose



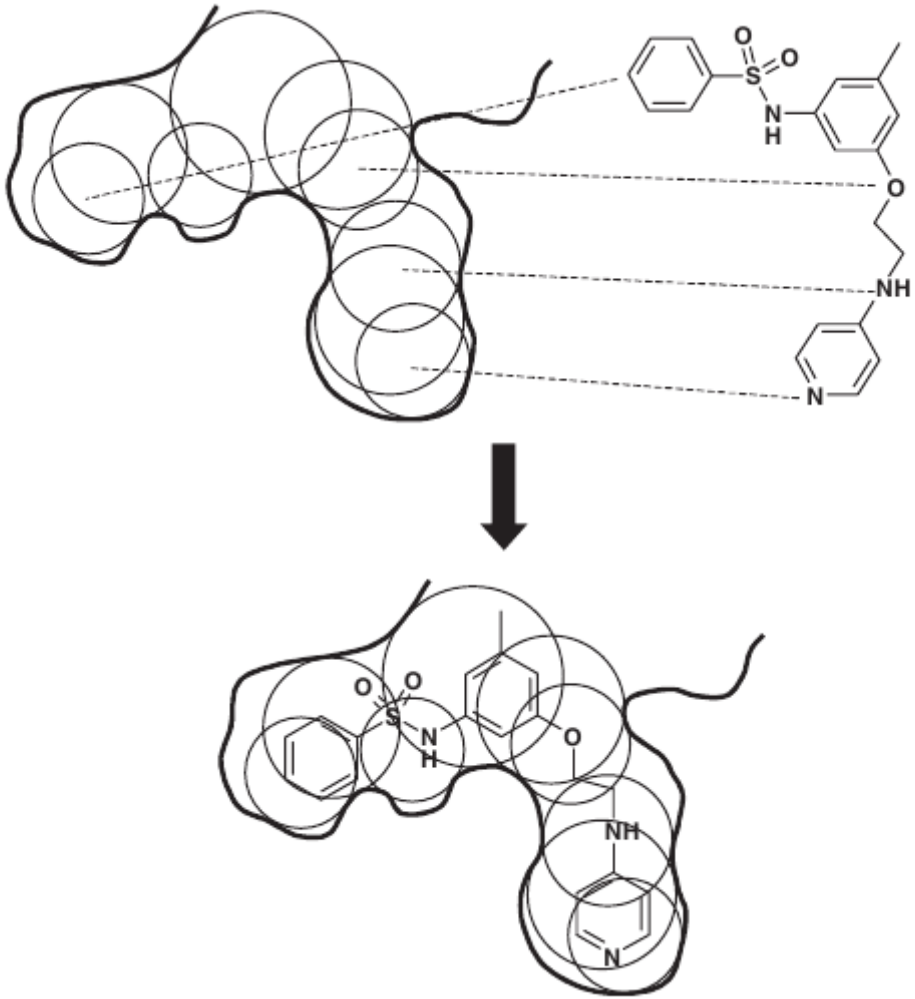
Search algorithms

Ligand	Receptor
Rigid	Rigid
Flexible	Rigid
Flexible	Flexible

Fast & Simple

Slow & Complex

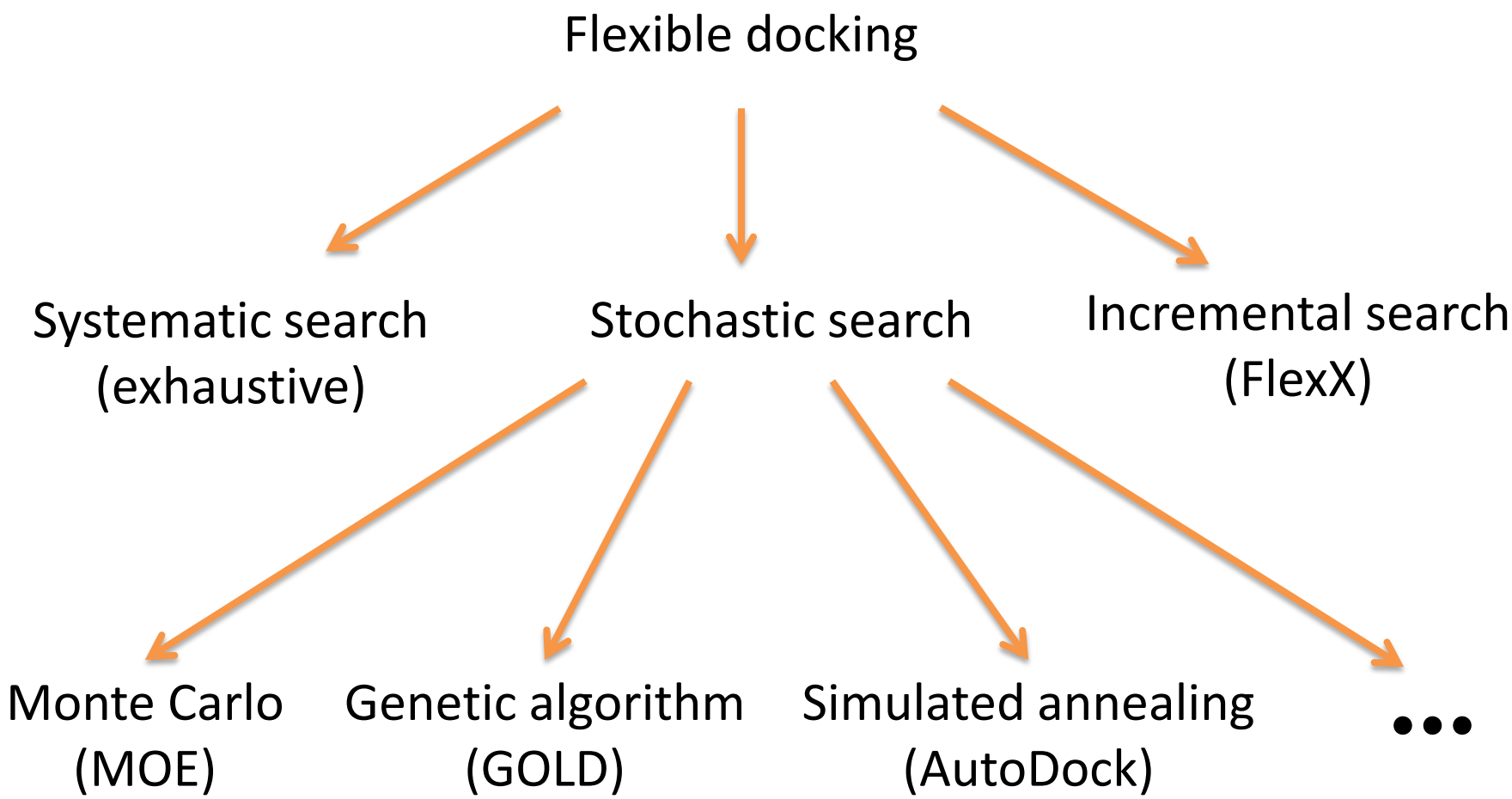
Rigid docking: DOCK



Binding site is filled with spheres (“negative image”) and the spheres centers are matched to ligand atoms

Leach A. Introduction to chemoinformatics

Ligand flexibility



Ligand flexibility: Monte Carlo

Start with conformation A (energy E_A)

Make random move to conformation B (energy E_B)

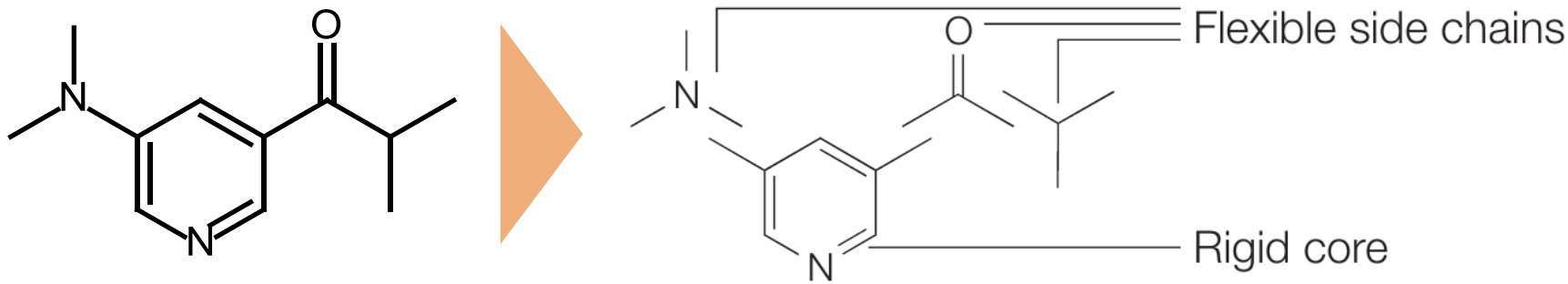
Accept move when:

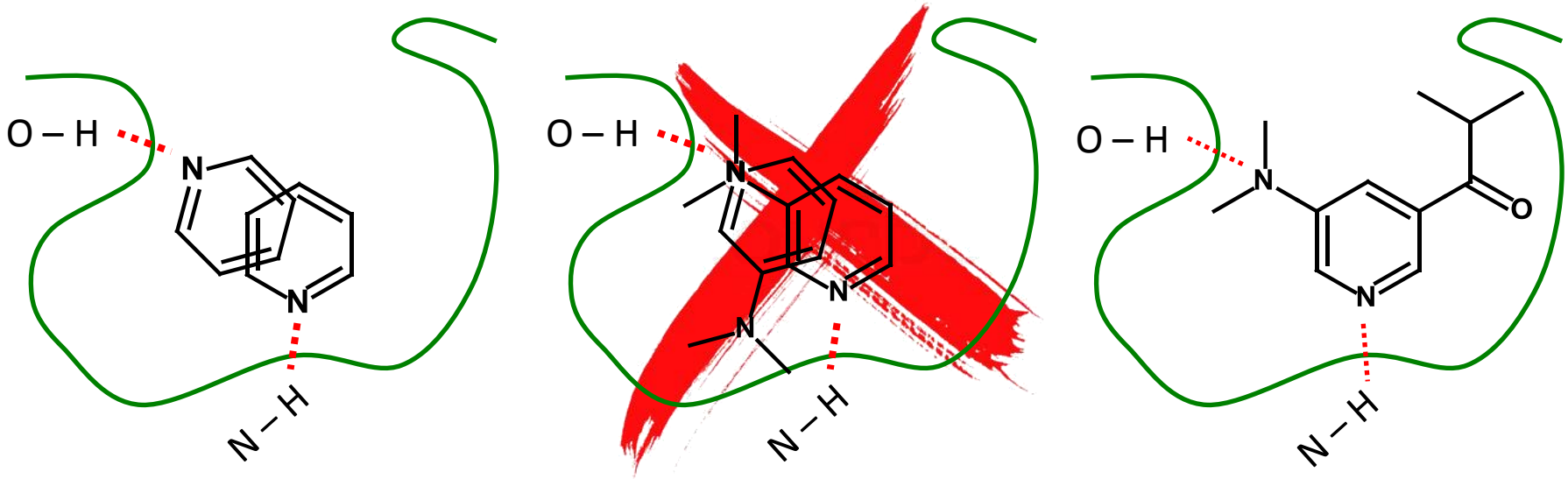
$E_B < E_A$ or if

$E_B > E_A$ except with probability P:

$$P = e^{\frac{E_B - E_A}{kT}}$$

Ligand flexibility: Incremental search

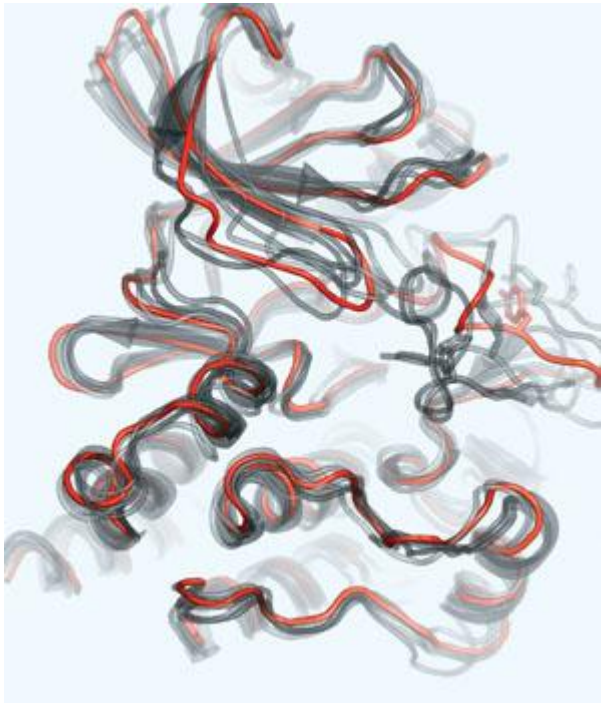




Protein flexibility

“Induced fit”
the protein may deform slightly to accommodate different ligands

Expand protein conformational space
generation of possible conformation and dock to them



Scoring functions

The ultimate goals of an ideal scoring function:

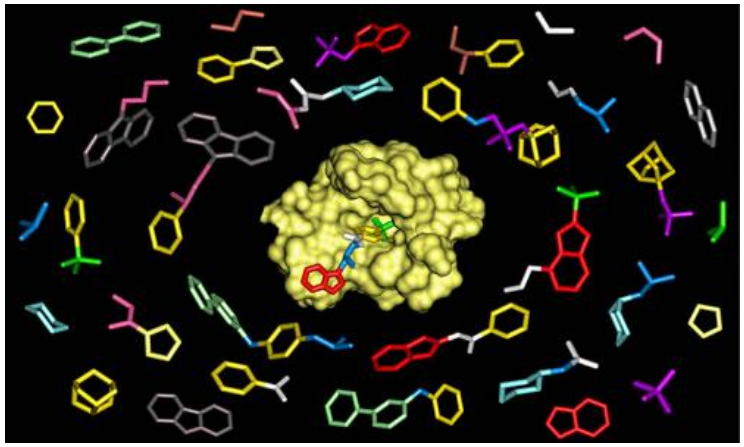
accurate within less that 1 pK_D unit (<1.4 kcal/mol)

generally valid (non system specific, large affinity range)

robust (tolerant with respect to the structural uncertainties)

physically meaningful (interpretable)

fast and easy to compute



Classes of scoring function

Forcefield-based

Based on terms from molecular mechanics forcefields

GoldScore, DOCK, AutoDock

Empirical

Parameterised against experimental binding affinities

ChemScore, PLP, Glide SP/XP

Knowledge-based potentials

Based on statistical analysis of observed pairwise distributions

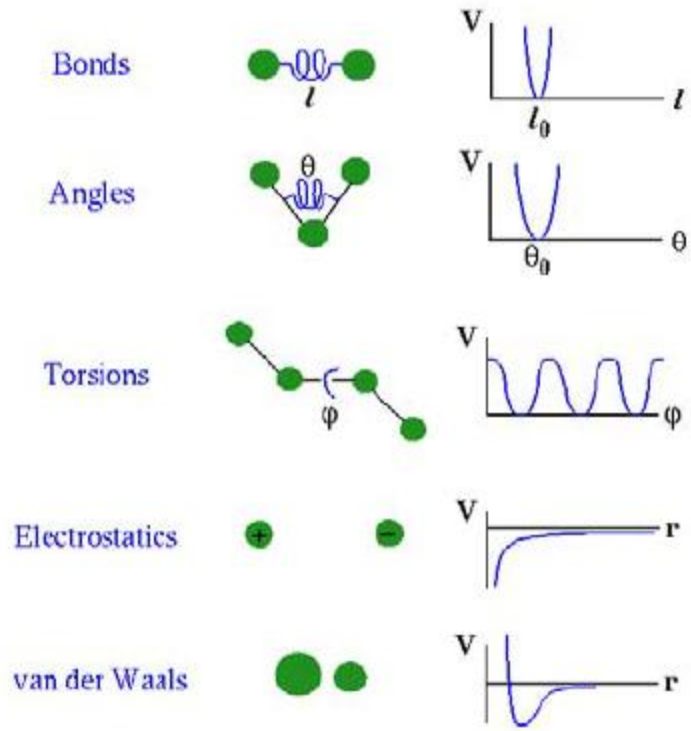
PMF, DrugScore, ASP

Force field-based scoring functions

Force field-based methods

Molecular Mechanics (MM):

- atoms → charged spheres
- bonds → springs
- classical potentials
- no electrons → no bond formation / cleavage
- typically parameterized to reproduce molecular potential energy surface (→ conformational ΔH in the gas phase!)



➡ Scoring protein-ligand complexes:

- \+ for pose prediction in docking
- − for ligand ranking by affinity

➡ Terms accounting for (de)solvation & entropic factors required (cf. MM-PBSA)

Typical force-field scoring function

electrostatic

Van der Waals

$$E = \sum_{i}^{lig} \sum_{j}^{prot} \left(\frac{q_i q_j}{\varepsilon_{ij} r_{ij}} + \frac{A_{ij}}{r_{ij}^{6}} - \frac{B_{ij}}{r_{ij}^{12}} \right) +$$

$$+ \Delta G_{HB} \sum \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{6}} \right) +$$

H-bonding

$$+ \Delta G_{tor} N_{rot} +$$

entropy term

$$+ \Delta G_{sol}$$

solvation term

Empirical scoring function

Empirical scoring functions

Regression-based:

$$pKi = \Sigma\, pKi_n\, f_n(\text{structure})$$

affinity

weighting factors

structure descriptors

determined via regression analysis (MLR, PLS)

Data:

Böhm's empirical scoring function

$$\Delta G_{bind} = \Delta G_0 + \Delta G_{hb} \sum_{h-bonds} f(\Delta R, \Delta\alpha) + \Delta G_{ionic} \sum_{\substack{\text{ionic} \\ \text{interactions}}} f(\Delta R, \Delta\alpha) + \Delta G_{lipo} \left| A_{lipo} \right| + \Delta G_{rot} NROT$$

H-bonding and **ionic terms** are dependent on geometry of interactions, large deviation in distance and angles are penalized

Lipophilic term is proportional to the contact surface area involving non-polar atoms

Conformational entropy term is proportional to the number of rotatable bonds

Böhm H.-J., Journal of Computer-Aided Molecular Design 1994, 8, 243-256

Knowledge-based scoring function

Knowledge-based scoring functions

Derivation from crystal-structure data

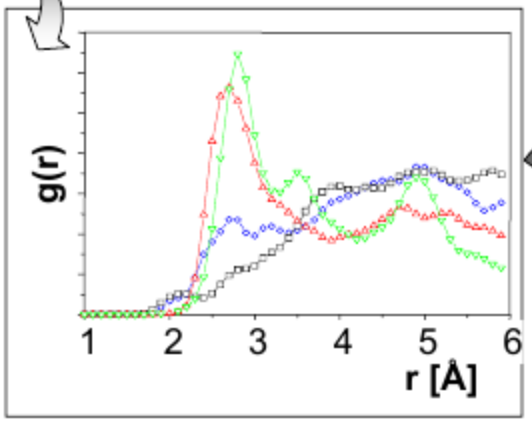
$$P_{ij}(r) = -\ln \frac{g_{ij}(r)}{g_{ref}}$$

P_{ij}: distance-dependent pair potential

g_{ij}: frequency distribution of atom-atom contacts

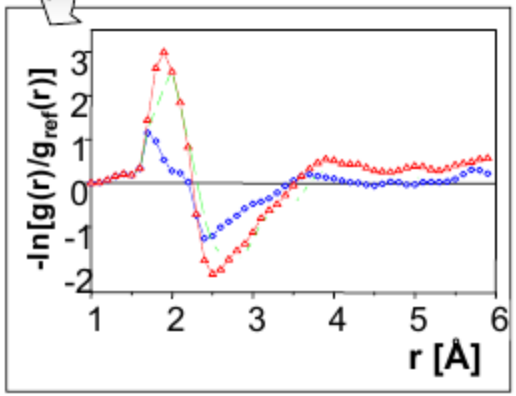
g_{ref}: reference distribution

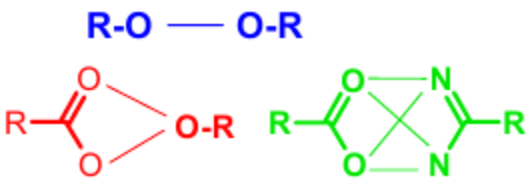
Frequency of occurrence



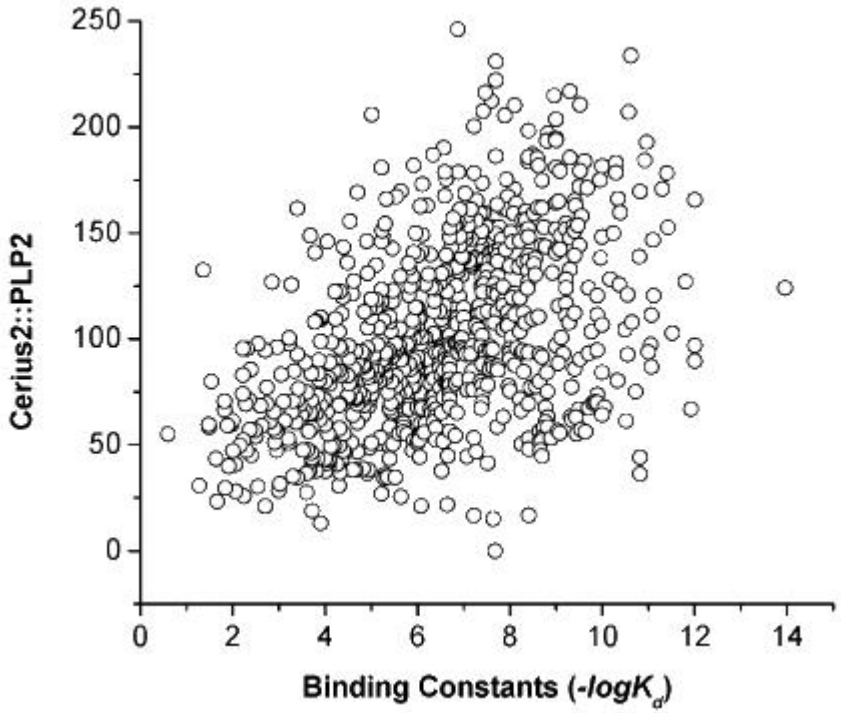
No experimental affinities used!

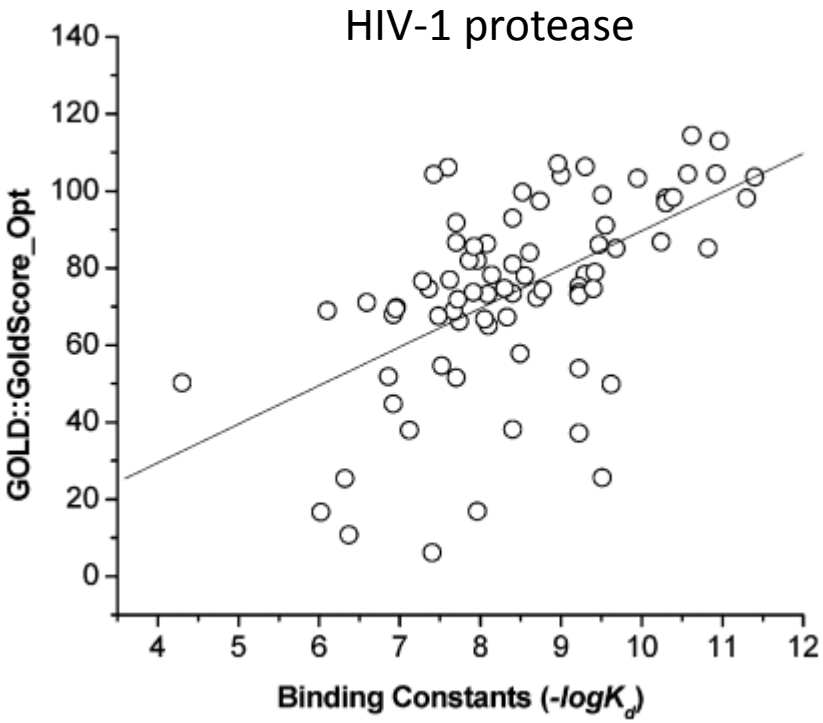
Statistical potential





Quality of the scoring predictions





Test set size = 800 molecules

r < 0.55

R. Wang *J. Chem. Inf. Comput. Sci.*, **2004**, *44* (6), pp 2114–2125

Consensus scoring

Rank-by-number

average score values obtained from each scoring procedure

Rank-by-rank

average ranks obtained from each scoring procedure

Rank-by-vote

sum the number of times the compound is within 2-5-10% top scored compound according by each scoring function

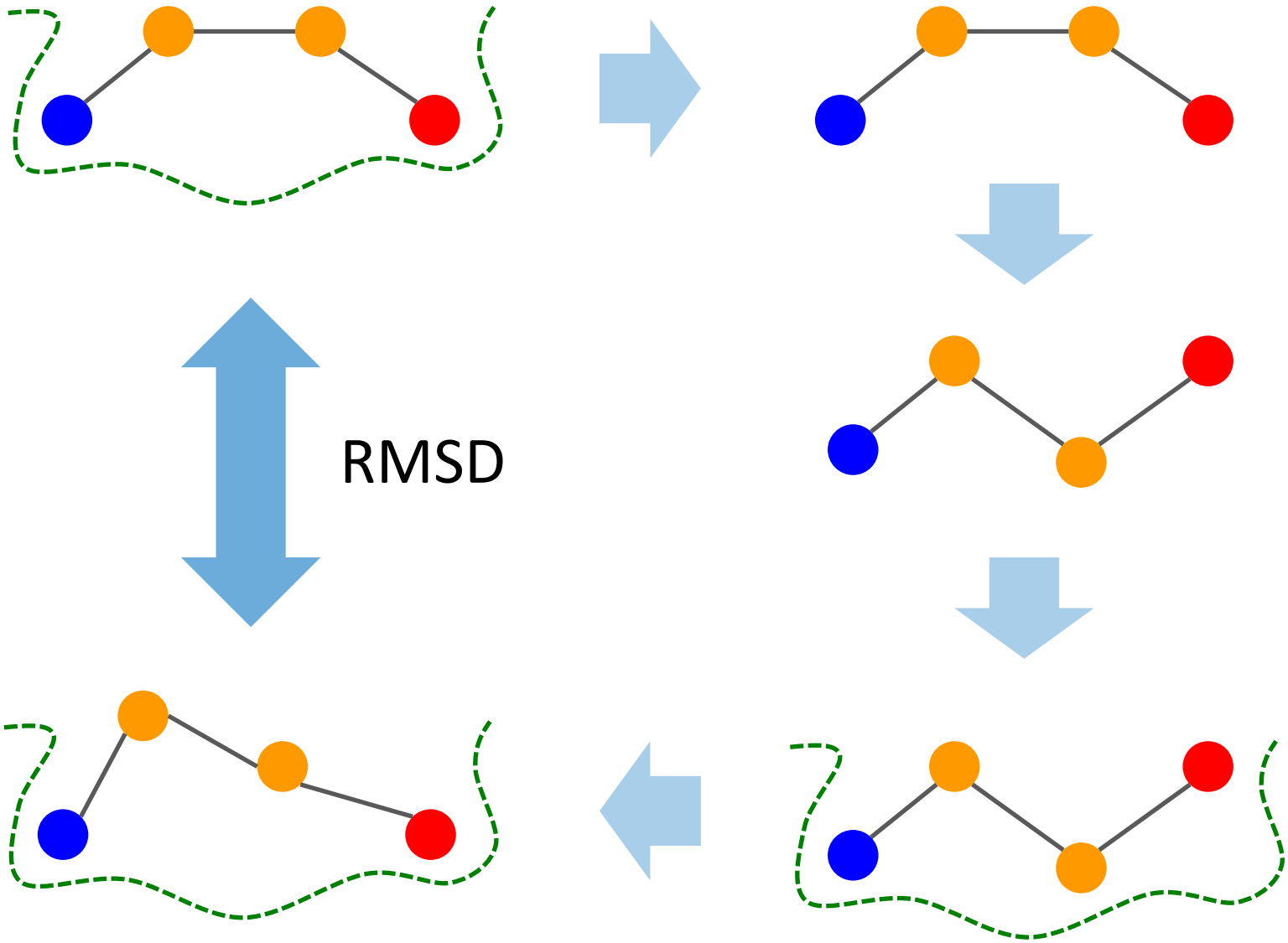
Scoring functions: Recommendations

Validate the scoring function for your system of interest

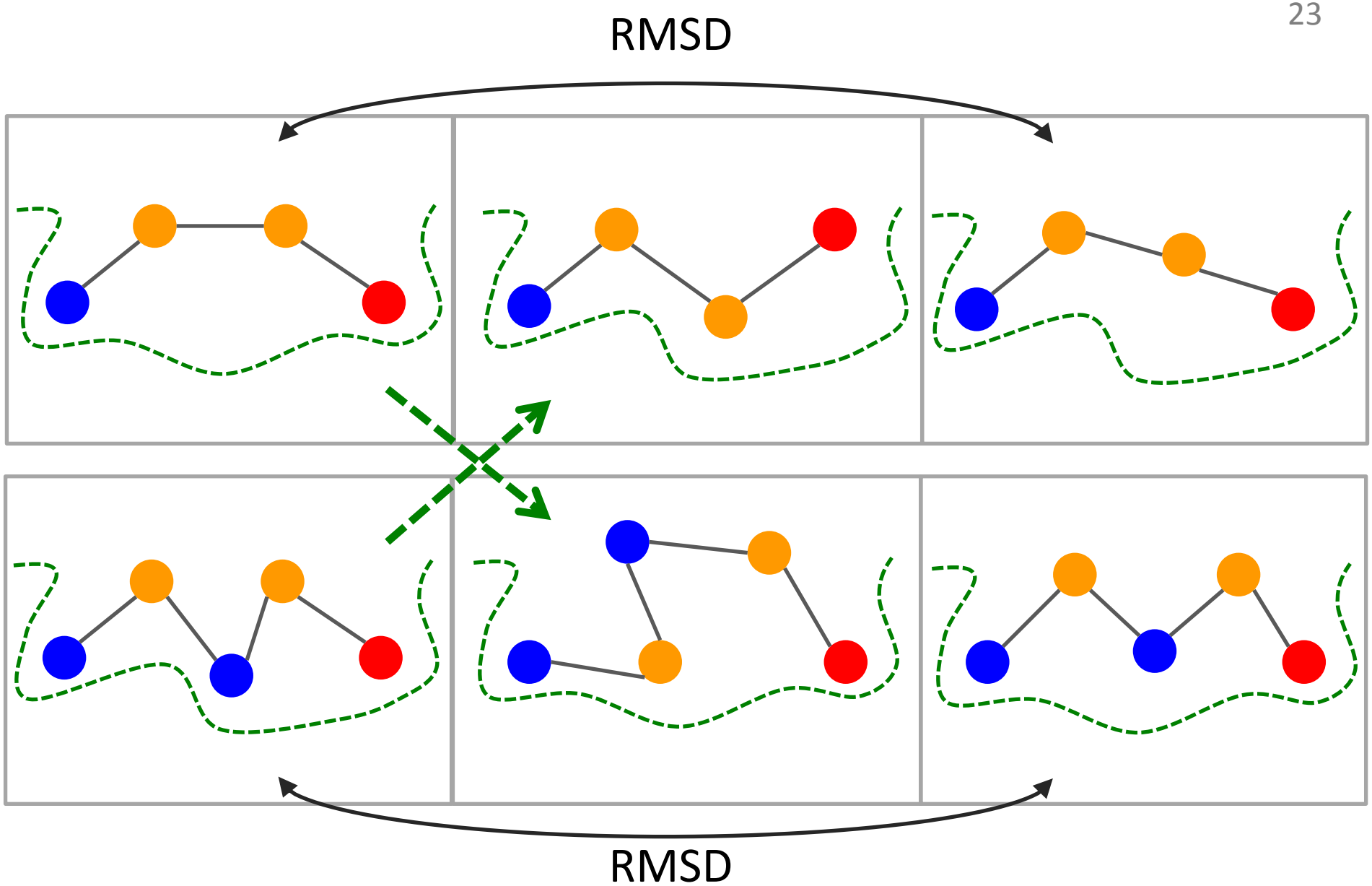
Apply several scoring functions
(“Consensus scoring”)

Train the scoring function for your system
(“Tailored scoring function”)

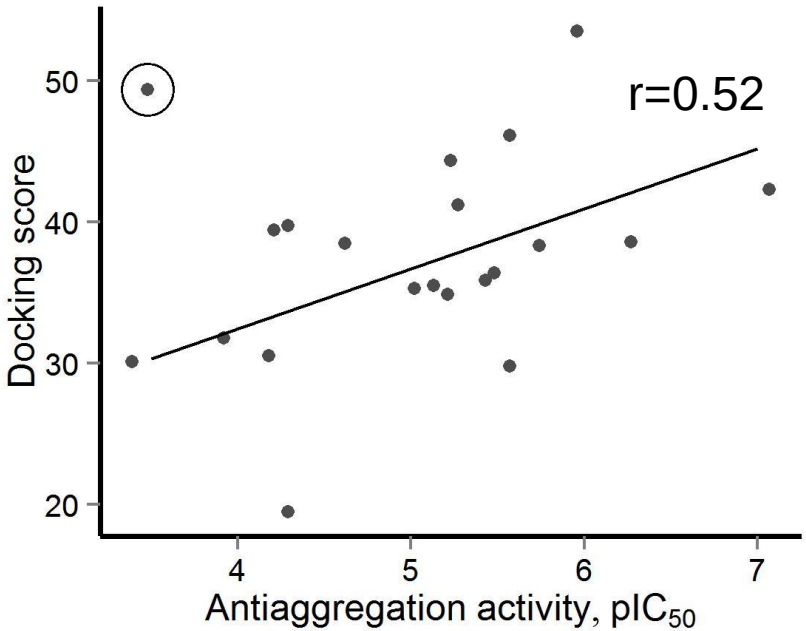
Docking validation: Re-docking (self-docking)

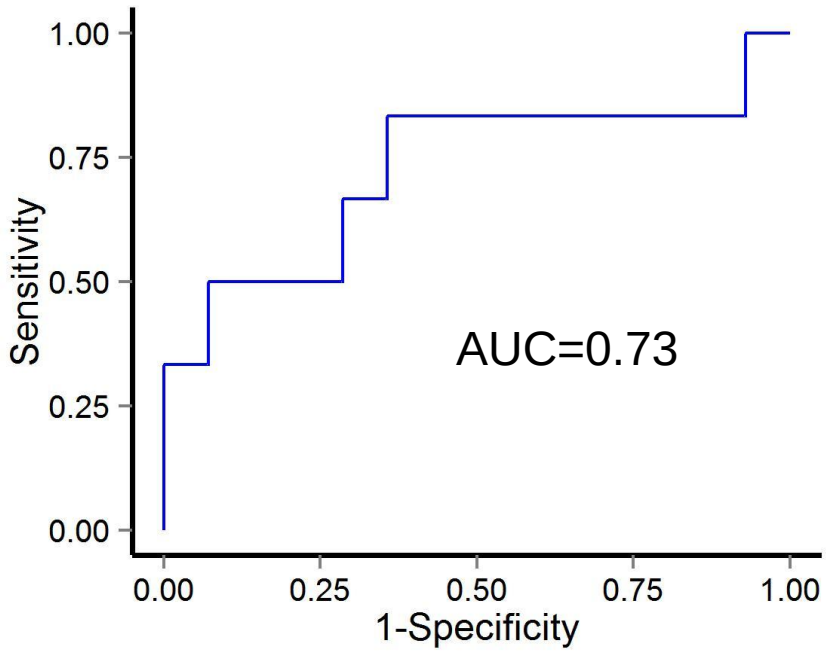


Docking validation: Cross-docking



Docking validation: Screening of test set compounds



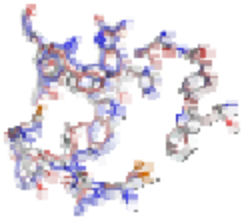


Recommendations

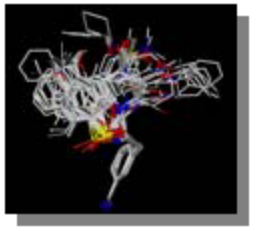
Before docking ...

... take care of the setup!

- Protein structures:
  - Protonation states and H-bonding networks
  - Quality and completeness of structural data
  - Location of binding site
  - Experimental data about water molecules and flexible regions



- Ligand structures:
  - Protonation states (influenced by protein!)
  - Tautomers
  - Conformers



- Docking program:
  - Choose suitable parameters
  - Validate, validate, validate (in particular for your system)!
- Know you program!
- Check structures and setup visually!
- Critically assess the quality of automated setup routines!

Recommended literature

Principles of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions
I. Halperin, B. Ma, H. Wolfson, and R. Nussinov
PROTEINS: Structure, Function, and Genetics 47:409–443 (2002)

A review of protein-small molecule docking methods
R. D. Taylor, P. J. Jewsbury & J. W. Essex
Journal of Computer-Aided Molecular Design, 16: 151–166, 2002

A Critical Assessment of Docking Programs and Scoring Functions
G. L. Warren, C. W. Andrews, Anna-Maria Capelli et al.
J. Med. Chem., 2006, 49 (20), pp 5912–5931