KATEDRA FYZIKÁLNÍ CHEMIE
UNIVERZITY PALACKÉHO V OLOMOUCI

INSTITUTE OF MOLECULAR AND TRANSLATIONAL MEDICINE

6th Advanced *in silico* Drug Design

KFC/ADD

Molecular Docking intro

Karel Berka

UP Olomouc, 30.1.-3.2. 2023

EMBL-EBI

INSTITUTE OF PHYSICS
National academy of Sciences of Ukraine

ÚOCHB AV ČR
IOCB PRAGUE

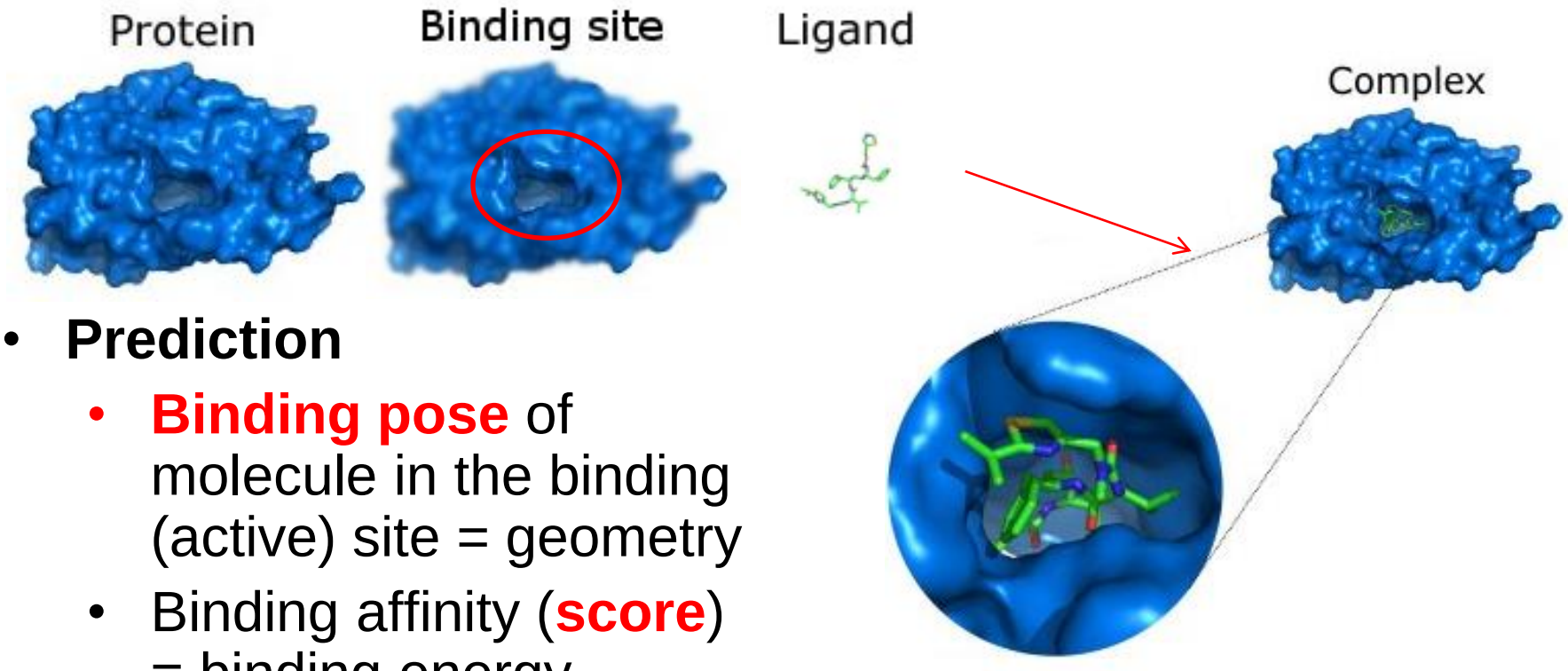
Molecular Docking Idea

- Finding the best ”fit“ of ligand to receptor

Bill Watterson: Calvin a Hobbes

Molecular Docking

Computational method mimicking binding of ligand to receptor

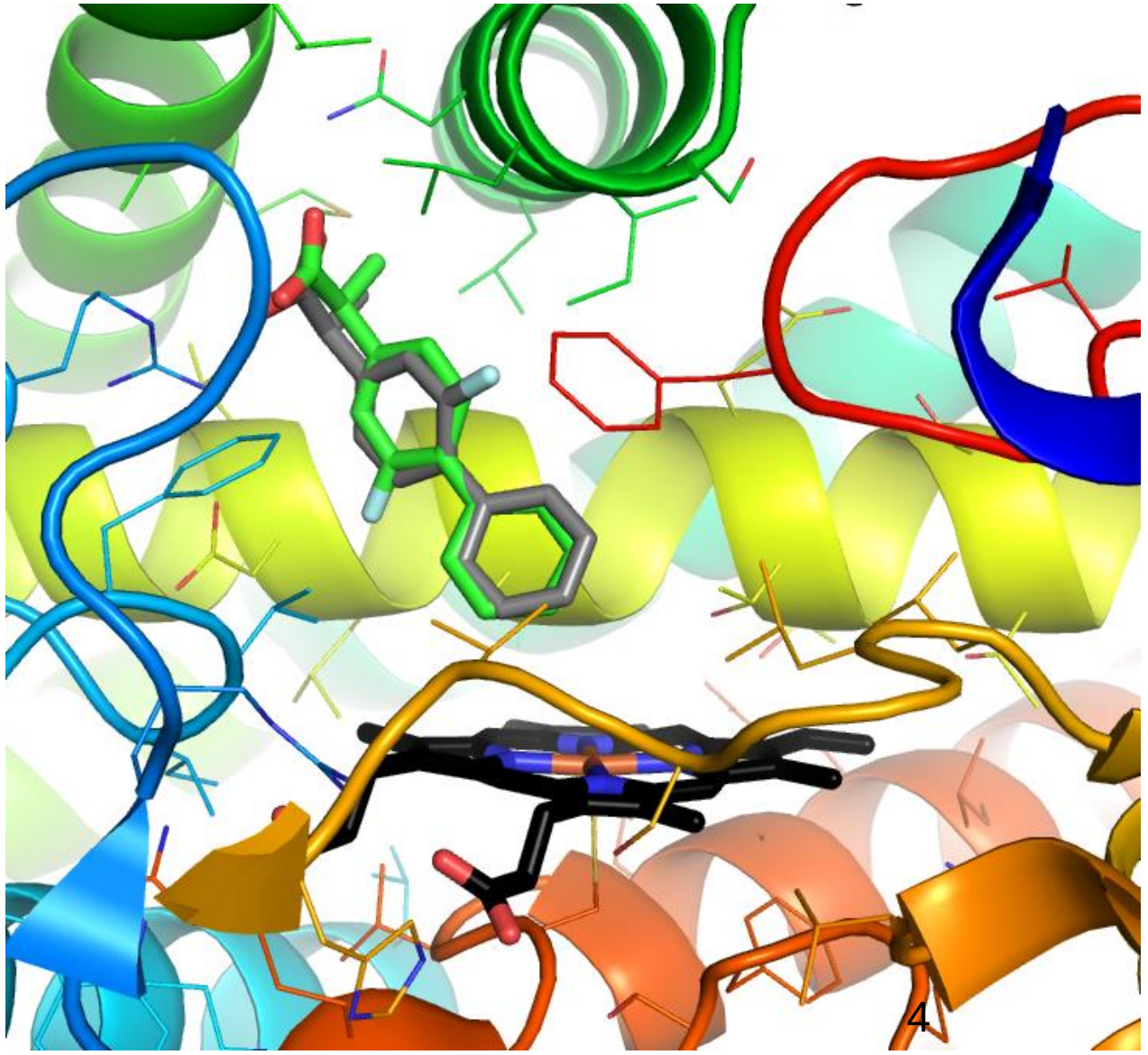


- **Prediction**
  - **Binding pose** of molecule in the binding (active) site = geometry
  - Binding affinity (**score**) = binding energy

Image credit: **Charaka Goonatilake, Glen Group, University of Cambridge.** http://www-ucc.ch.cam.ac.uk/research/cg369-research.html

Binding Pose

- Structural arrangement of ligand within receptor/enzyme
- Driven by molecular interactions



Energetics

- Equilibrium binding constant

$$K_d = [P...L] / [P][L]$$

- correspond to free energy of binding:

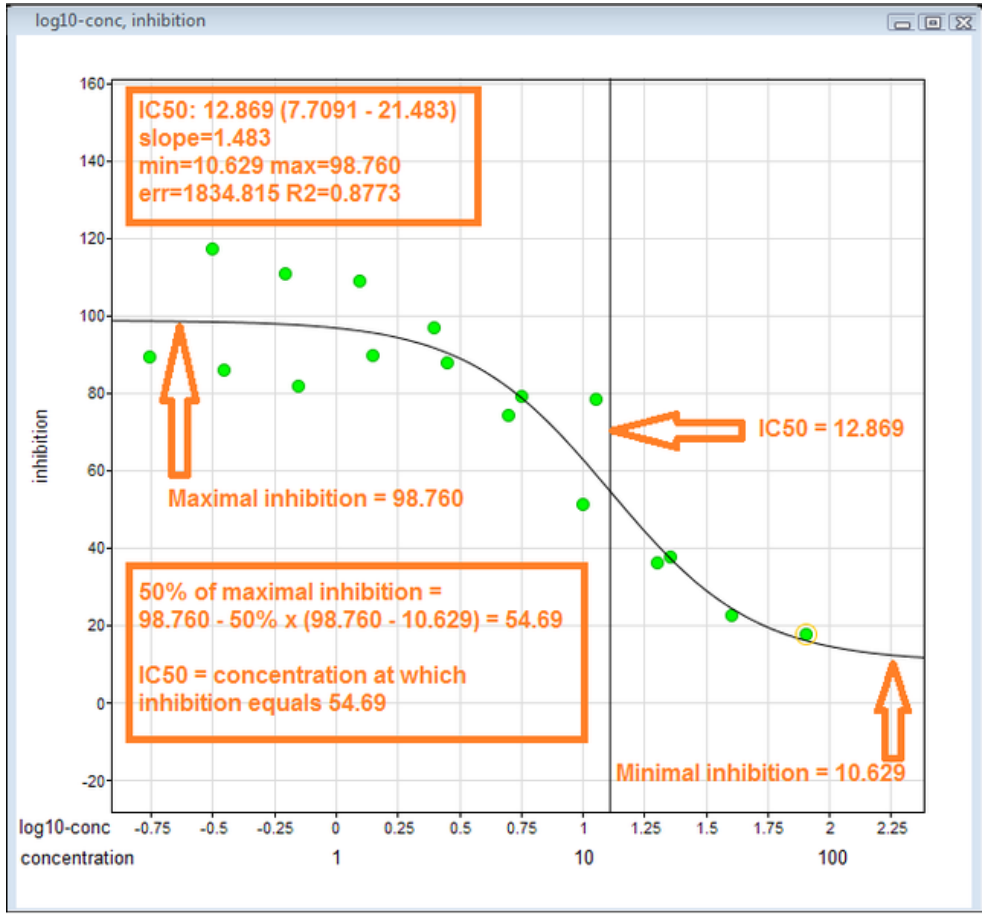
$$\Delta G_{bind} = -RT \ln K_d$$

Free energy – combination of enthalpy and entropy

$$\Delta G_{bind} = \Delta H_{bind} - T\Delta S_{bind}$$

- k_{cat}, K_i, IC_{50} , EC_{50}– other values used for characterization
  - depend on concentration and affinity of substrate and concentration of protein

IC_{50}



- Concentration with 50% of inhibition activity
  - Comparison of affinity between two compounds
  - Cheng-Prusoff equation

$$K_i = \frac{IC_{50}}{1 + \frac{[S]}{K_m}}$$

  - Often logarithmic (mol/L)

$$pIC_{50} = -log_{10}(IC_{50})$$

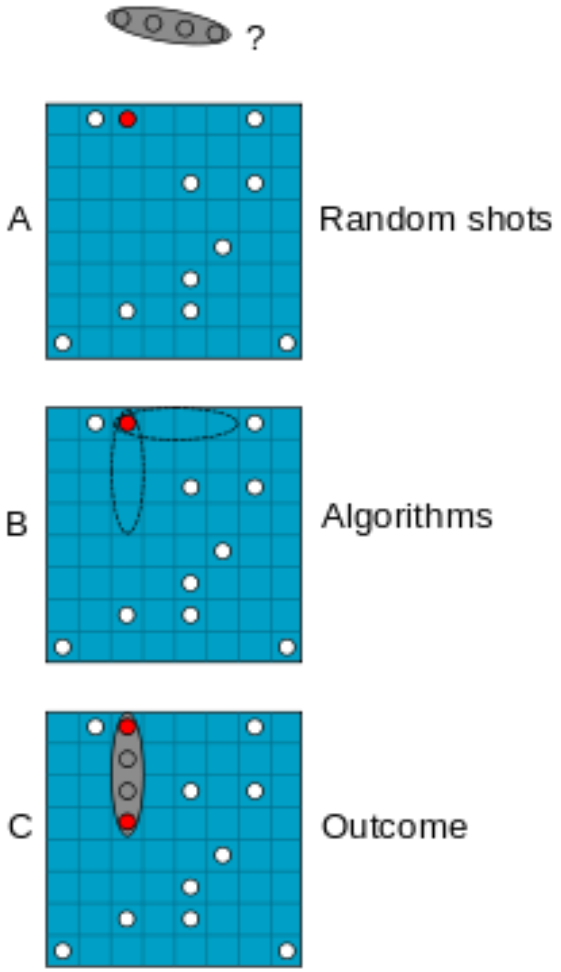
- Lower = better

pM (excelent) > nM (great) > µM (common) > mM (unusable)

Visual demonstration of how to derive IC50 value: Arrange data with inhibition on vertical axis and log(concentration) on horizontal axis; then identify max and min inhibition; then the IC50 is the concentration at which the curve passes through the 50% inhibition level. (wikipedia)

Search Algorithms

- Monte Carlo
  - Random selection
  - Metropolis condition
    - (if better energy -> accept new pose;
      else check depend on energy difference)
- Genetic algorithms
  - Poses described by “Genes”
  - Best poses “mate” to generate offspring
  - Converge faster than MC
- Simulated heating
  - Heating – more energy – barrier crossing
  - Cooling – minima search



NEEDS ENERGY FUNCTION!

Scoring Function

1. Score individual binding poses during search – **objective function**
2. Identification of lowest (best) binding energy
3. Sort **binding free energies** between individual ligands – selection of the best ligand

– Not necessarily the same for all points
  - First part is most computationally intensive – needs to be quickest
  - Sorting should be the finest

Scoring Function Types

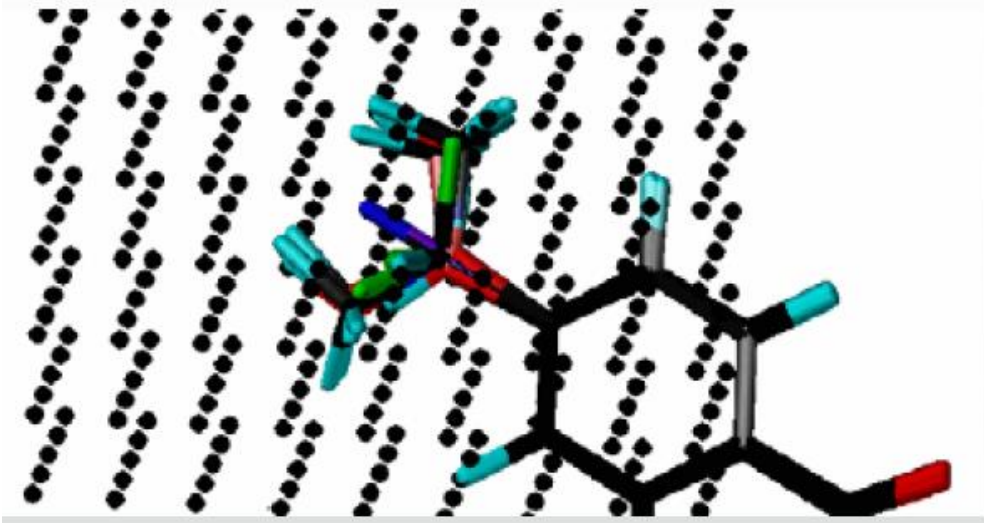
- **Force-field** – based on molecular mechanical force-fields
  - Physical model - Interaction terms (elstatic, vdW,…)
  - Goldscore, DOCK, Autodock
- **QM-based** – based on quantum chemical calculations
  - PM6-DH
- **Empirical** – parameterized against exp. binding affinities (K_d,IC_{50})
  - Arbitrary terms (H-bonds, hydrophobic contacts)
  - ChemScore, PLP, Glide SP/XP
- **Knowledge-based** – based on protein-ligand complexes
  - Boltzmann hypothesis
    - typical binding motives -> stronger binding
  - PMF, DrugScore, ASP

Force-field Scoring Functions

- **Physical interaction terms**

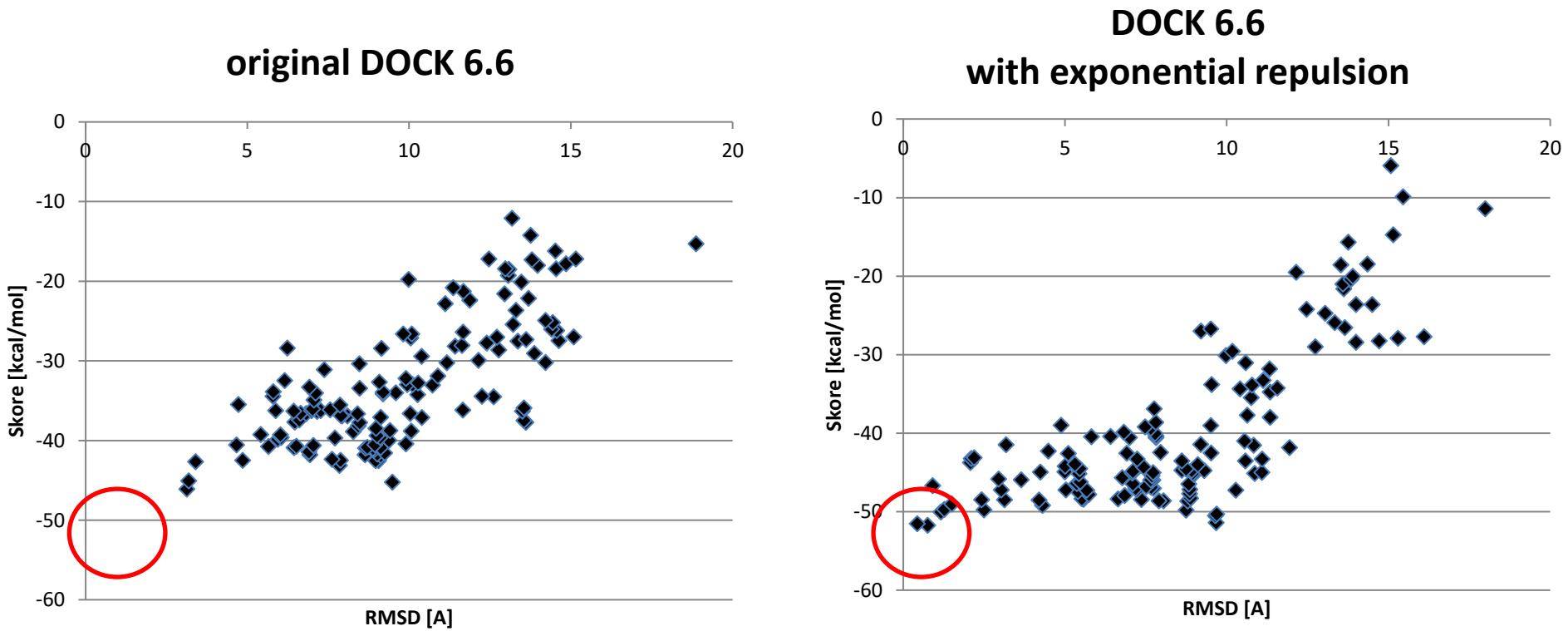
  $E = E_{bond} + E_{angle} + E_{dih} + E_{coulomb} + E_{vdw} + E_{solv}$

- Often only **intermolecular** terms ($E_{coul} + E_{vdw} + E_{solv}$)
- **Intramolecular** are usually changed to rigid (bonds, angles) or screened by some value (dihedrals by 5 deg)
- **Grid** – time-saving
  - Protein is divided into grid and interactions are pre-calculated at each point
  - Ligands interaction is evaluated by multiplication of grid potential with ligand atoms
  - Table search is quicker than full energy evaluation
  - Receptor is usually one, while there is a series of ligands



Scoring Function Problems Example

- Problems:
  - Repulsion – Exponential by nature, usually r^{12} in LJ potential
  - Electrostatics



Bazgier V, Berka K, Otyepka M, Banáš P: Exponential repulsion improves structural predictability of molecular docking. *J. Comput. Chem., 37(28), 2485–2494, 2016*

QM based Scoring Function

- Based on quantum chemical calculations
- PM6-DH2

$$\Delta G'_{\mathrm{w}} = \Delta H_{\mathrm{w}} - T\Delta S_{\mathrm{w}} + \Delta E_{\mathrm{def}}(\mathrm{I}) + \Delta\Delta G_{\mathrm{w}}(\mathrm{I}).$$

- ΔH_w - interaction enthalpy
- $-T\Delta S_w$ - interaction entropy
- ΔE_{def} - correction for inhibitor deformation
- $\Delta\Delta G_w$ - correction for inhibitor hydration

Dobeš et al. J Comput Aided Mol Des 2011, DOI: 10.1007/s10822-011-9413-5

Empirical scoring function

- Decomposition of binding energy into pre-defined “chemical terms”
- Specific interactions taken explicitly
  - H-bonding, π-π stacking, …

Linear form of terms is usually used (albeit unphysical)

$$\Delta G_{bind} = \Delta G_{solvent} + \Delta G_{conf} + \Delta G_{rot} + \Delta G_{t} + \Delta G_{r} + \Delta G_{vib}$$

Böhm's empirical scoring function

- linear summation of individual binding terms
- **Bohm's scoring function**
  - H-bonding, ion interaction, lipophilic interactions and conformational terms

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

- **Hydrogen bonding** a **ionic interactions**
  - Depend on na geometrical interaction – large deviations are penalized (ideal distance R, ideal angle α).
- **Lipophilic term**
  - Proportional to lipophilic surface contact between protein and ligand (A_{lipo})
- **Conformational entropic term**
  - penalization for freezing of internal rotations of ligand - entropy
  - Proportional to number of rotationable bonds of ligand (NROT)
- ΔG values of individual terms are constants obtained by linear regression on experimental binding data on 45 protein–ligand complexes

- Bohm, *J. Comput.-Aided Mol. Des.*, **1994**, *8*, 243

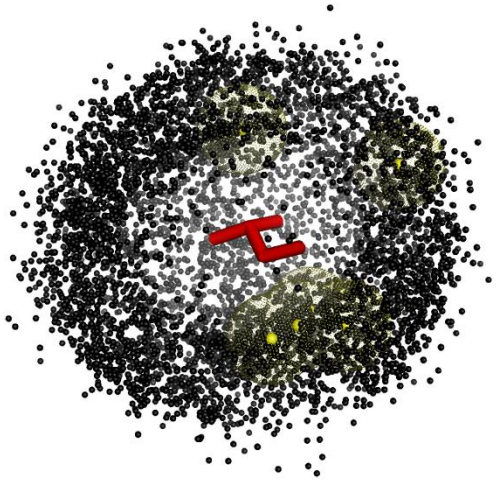
Knowledge-based Function

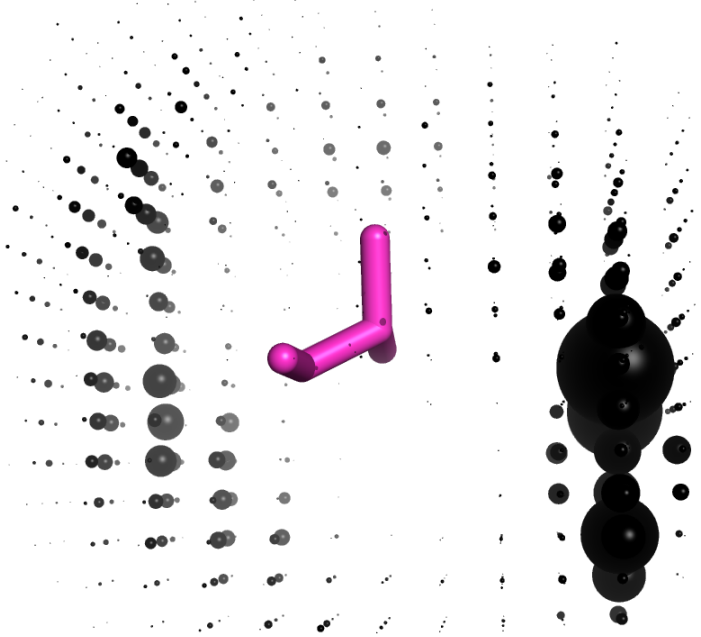
Correlation of structural data from ligand/protein complexes with free energy of binding

- Use a rigorous statistical mechanical result:

$$A = -kT \ln g(r)$$

  - This equation holds for an ensemble of particles at equilibrium (in gas)
  - not necessarily proteins





Drugscore

DRUGSCORE

$$\Delta W_{i,j}(r) = W_{i,j}(r) - W(r) = -\ln \frac{g_{i,j}(r)}{g(r)}$$

$$g(r) = \frac{\sum_i \sum_j g_{i,j}(r)}{i^* j}$$

Short-range (6 Å) contributions only – ignoring solvation

J. Mol. Biol. 295, 337-356, 2000

Docking Preparation

- Receptor
  - Identification of relevant structure
  - Structure preparation (missing atoms, hydrogen assignment)
- Ligand
  - Structure preparation
  - Isomers, conformations
- Other tasks
  - Water
  - Flexibility

H. Jhoti & A.R. Leach (eds) Springer 2007, chapter 8
Structure based drug discovery

Receptor Preparation

- Where
  - identification of binding site
- Good structure
  - Low R (accuracy)
  - Low B-factors (flexibility)
  - Low R-free (correctness)
- Flexibility
  - Rigid docking into several structures
    - Molecular Dynamics
    - more Xtals
    - AlphaFold models
  - Flexible docking

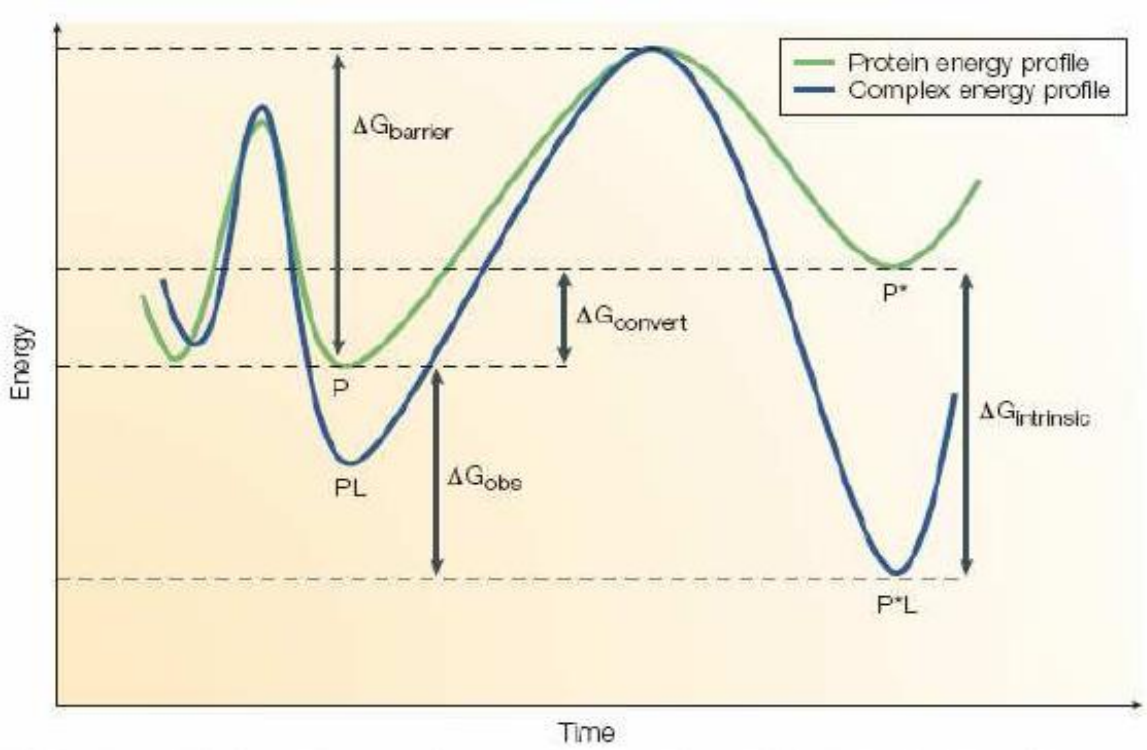
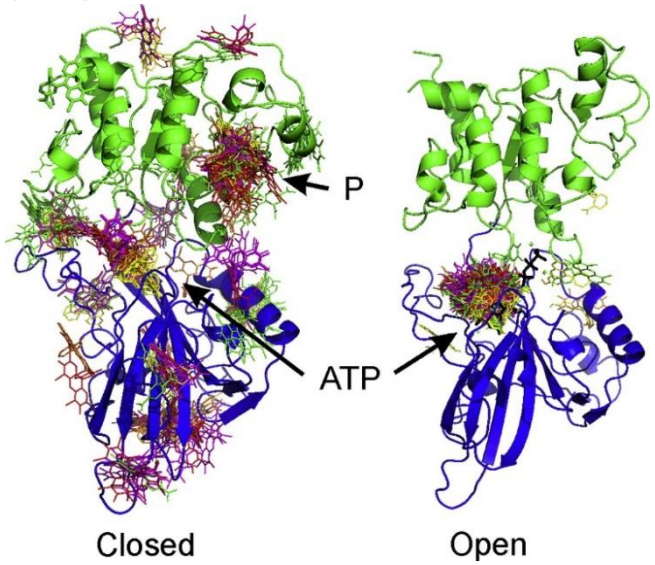


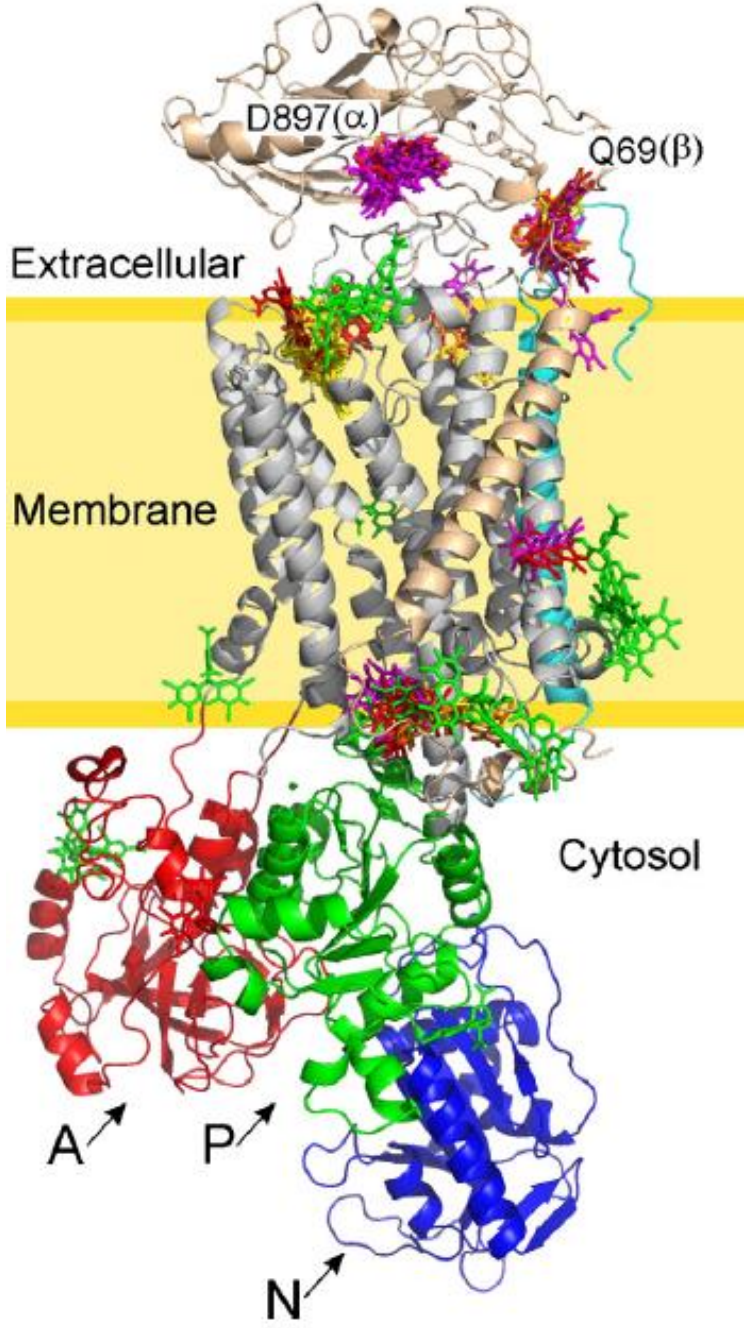
Figure 1 | **Protein mobility and ligand binding.** A protein is considered to exist in two conformations (P and P′) with an energy difference $\Delta G_{convert}$. The ligand (L) can bind the protein (P) to give a complex (PL), or bind to P* to give a complex (P*L). Although P* has a higher free energy, it might offer greater scope for interaction with L. For instance, P* might represent a conformer in which the binding site has opened and exposed hydrophobic patches. This is energetically unfavourable, but offers the potential for favourable interactions with the hydrophobic moiety of a suitable incoming L, thereby giving rise to a large, favourable interaction $\Delta G_{intrinsic}$. The resulting complex (P*L) has a lower energy than that of the complex PL. The observed affinity of L for the protein conformational ensemble is governed by ΔG_{obs}. Slow binding kinetics might well be observed, as P* is a higher-energy conformer than P and an energy barrier ($\Delta G_{barrier}$) must be surmounted before optimal binding to L can take place.

Teague, Nature Reviews Drug Discovery **2**, 527-541 (July 2003) | doi:10.1038/nrd1129

Example 1: Na^+/K^+–ATPase

- Ion pump
- Search for binding site
  - Fluorescent probes
  - RH241 probe
- Docking is highly sensitive to protein conformation





Havlikova M, … **Berka K**, … et al. *BBA*, 1828(2), 568, 2013
Huličiak M, … **Berka K**, … et al. BBA-Biomem*, 1859(10),2113-2122, 2014*

Protein Conformations

- **Rigid Receptor Approximation**
  - Most docking programs use rigid receptor for speed
- but…
  - Protein can deform in order to accept several ligands (**ligand-induced fit**)
  - Amino acids - several conformations
- **Flexible Receptor docking**
  - Increase of search size – higher computational demands
  1. Side chains only

     (docking selected sidechains together with ligands)
  2. Docking into several structures of protein

     Larger movements can be taken into account

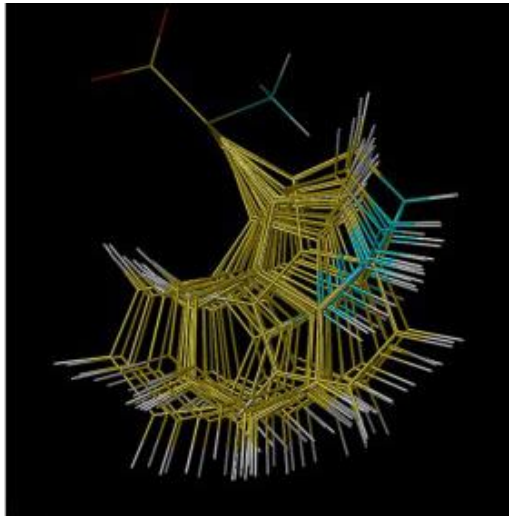
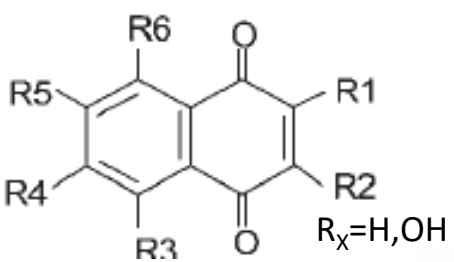


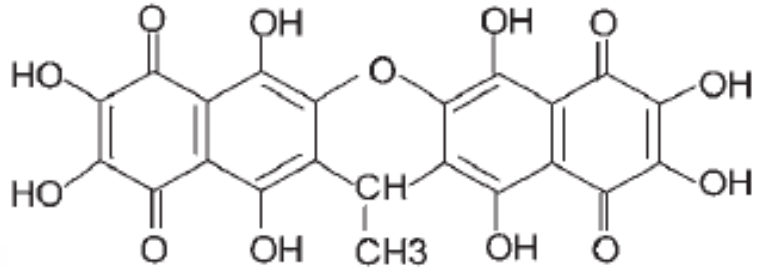
Image: Cláudio M. Soares, Protein Modelling Laboratory, http://www.itqb.unl.pt/labs/protein-modelling/activities/psccip-pf

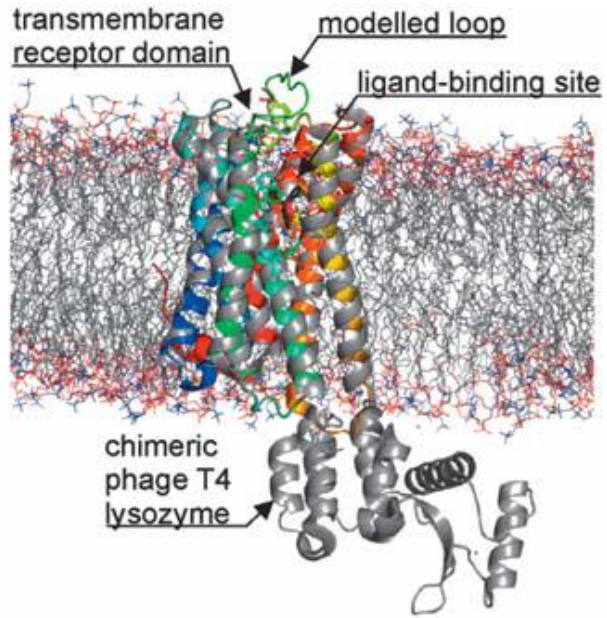
Example 2:
H1R receptor

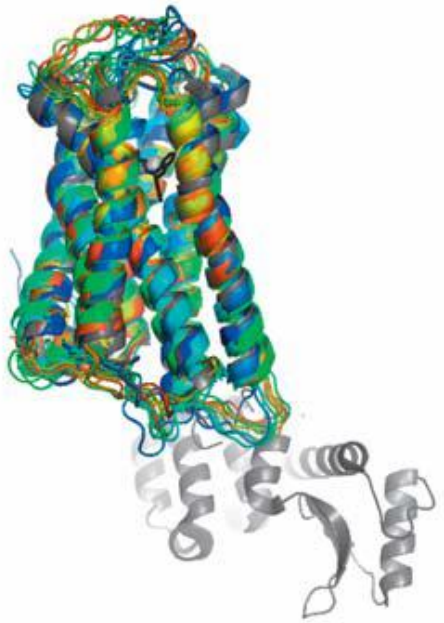
- Antiallergic compounds

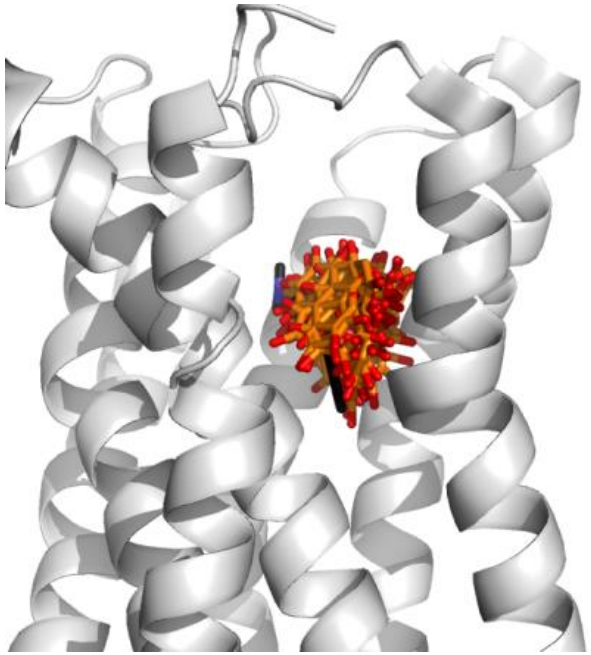
dG/n(atoms) – monomers are more active than dimers

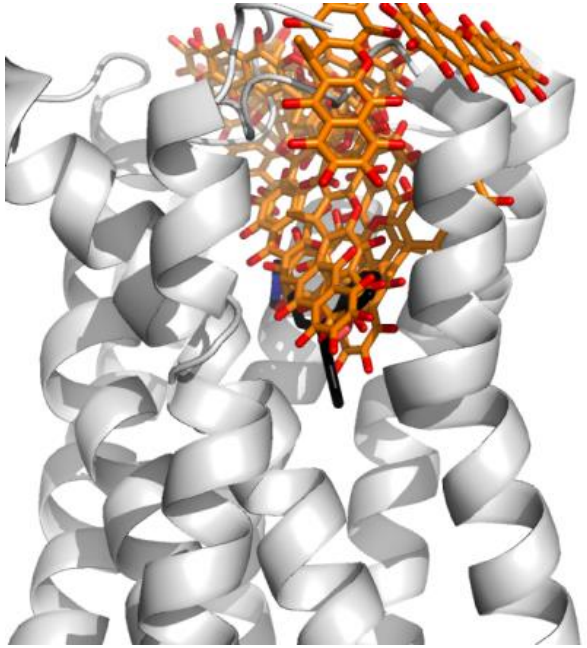










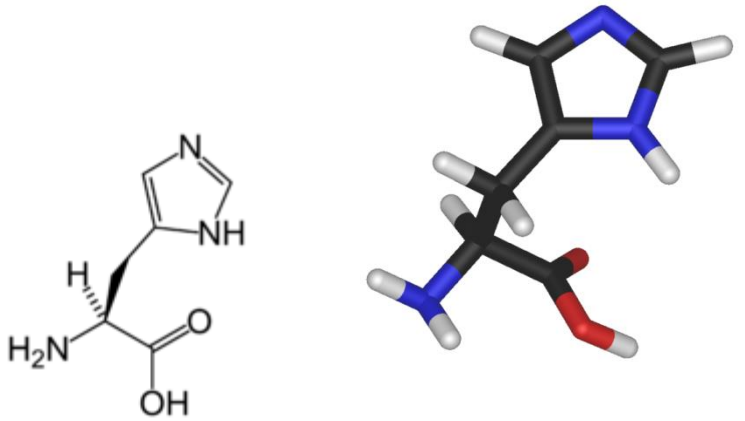


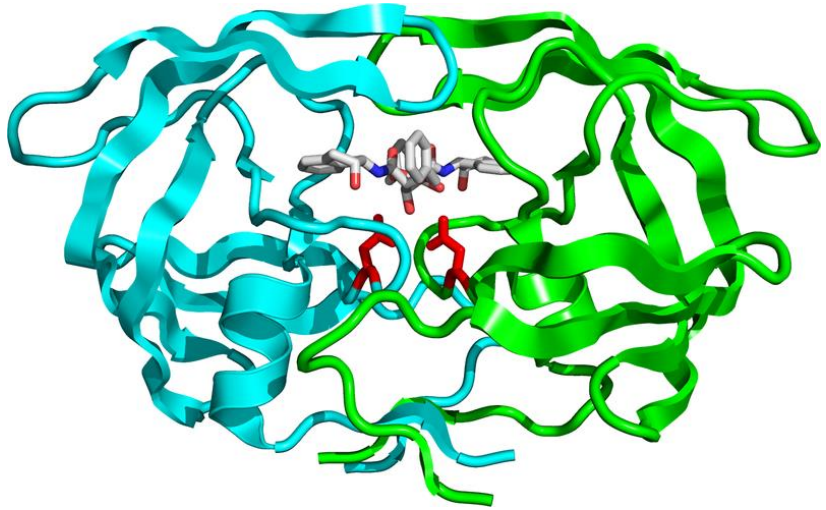
Pozharitskaya ON, …, **Berka K**, … et al. *Planta Med.* 79(18), 1698-1704, 2013.

Receptor Preparation

- protonation of aminoacids
  - His (pKa ~ 6.04)
  - Surroundings pKa shifts
    (Asp in HIV protease)
- tautomerization
- rotamers

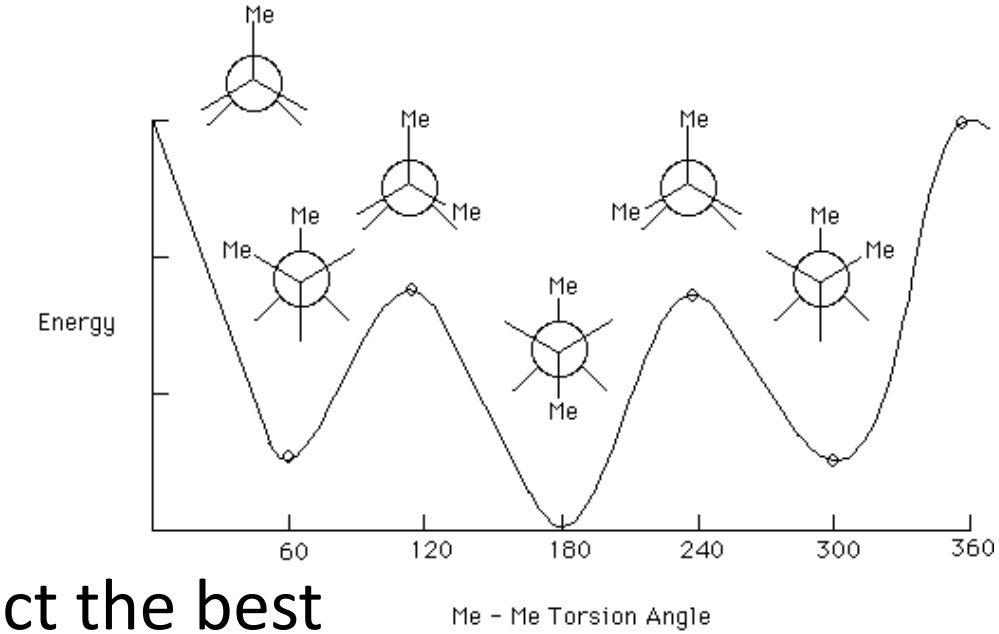
- pre-selection change results significantly



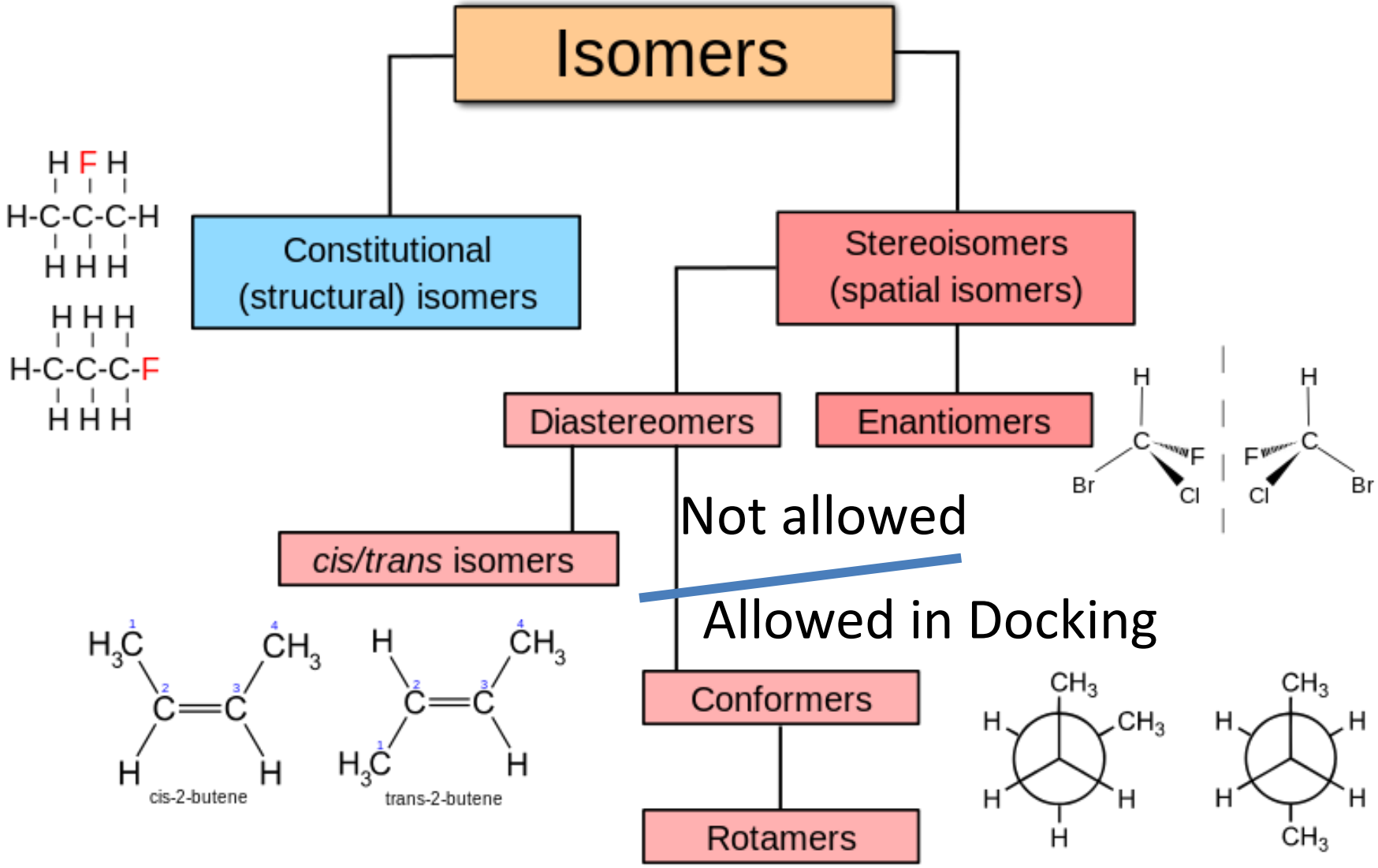


Ligand Preparation

- Ligand Flexibility
  - Ensemble of all possible ligand conformations
  - rotation C-C bonds,
    but not C=C or rings
  - Angles and bonds fixed
- Izomerization
  - Charge and tautomers
  - Prepare all and then select the best
    - Relative energy
  - Ask an expert! (organic chemists)



Isomers

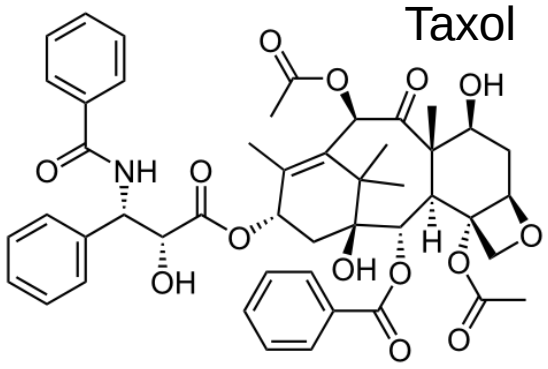


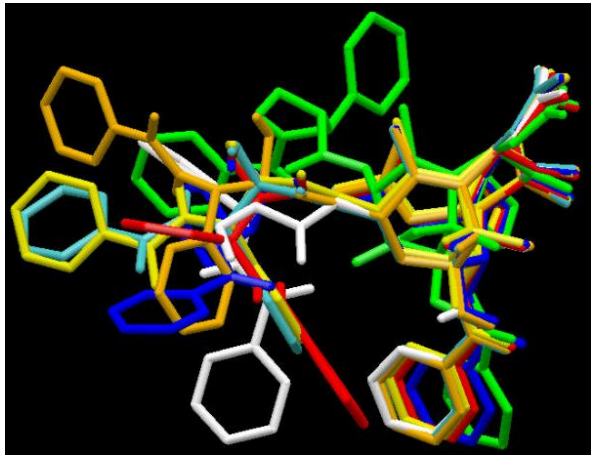
Wikipedia.org

Ligand conformation

- Conformation – rotation around torsion angles
  - N rotational bonds – rotate by θ degrees (5°)
  - Conformations: $(360º/\theta)^N$
- Question
  - If the torsion angles are incremented in steps of 30º, how many conformations does a molecule with 5 rotatable bonds have, compared to one with 4 rotatable bonds?
- Having too many rotatable bonds results in “combinatorial explosion”
- Also ring conformations

Image: IUPAC Gold Book

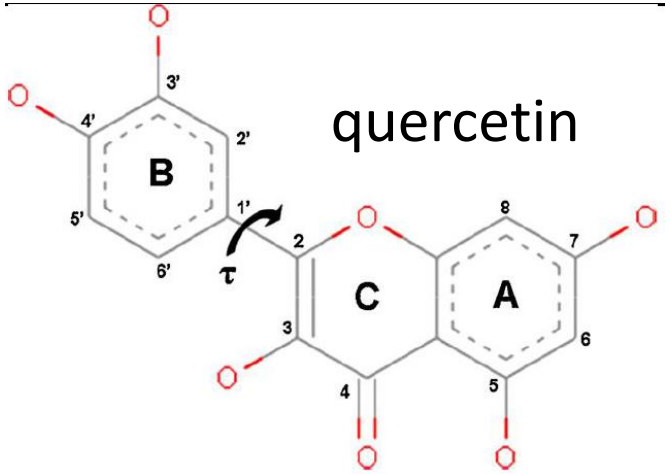


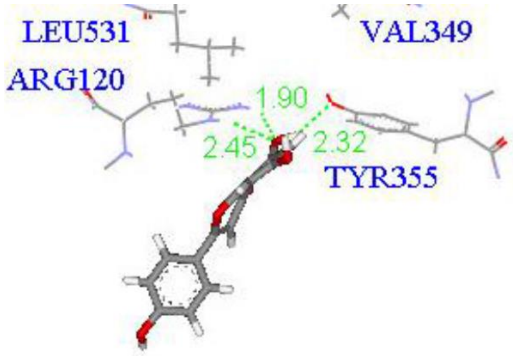


Lakdawala *et al. BMC Chemical Biology* 2001 **1**:2

Ligand Structure Generation

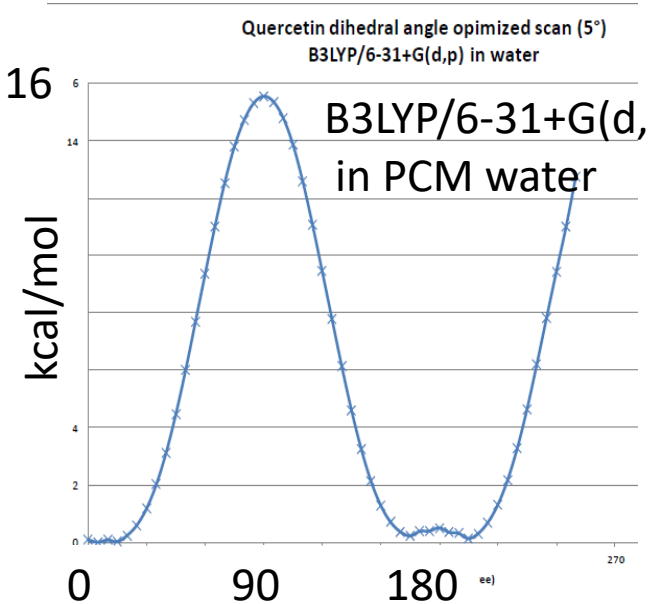
Torsion angles

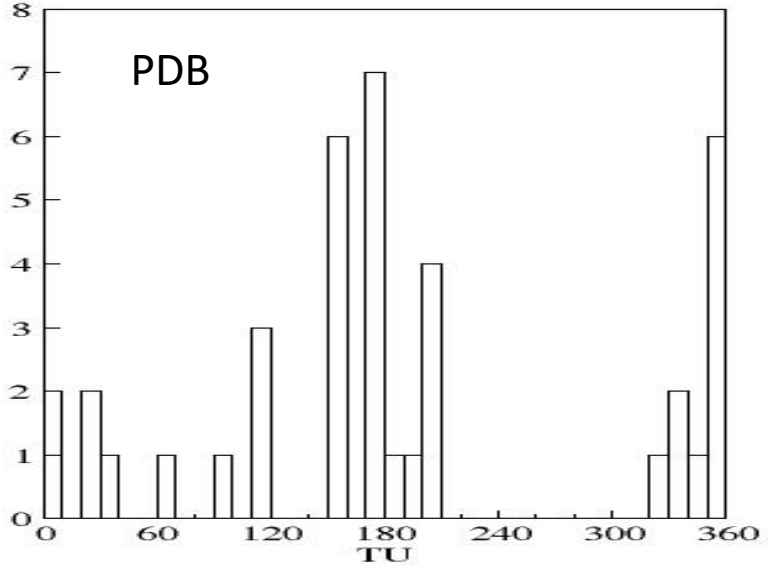




Accelrys DSmodelling 1.2 Wu, Chien-Ming et al *Int.J. Mol. Sci.* 8 (2007): 830–841.

GLIDE
D'mello, P et al *Int.J.Pharm. Sci.* 3 (2011): 33–40.



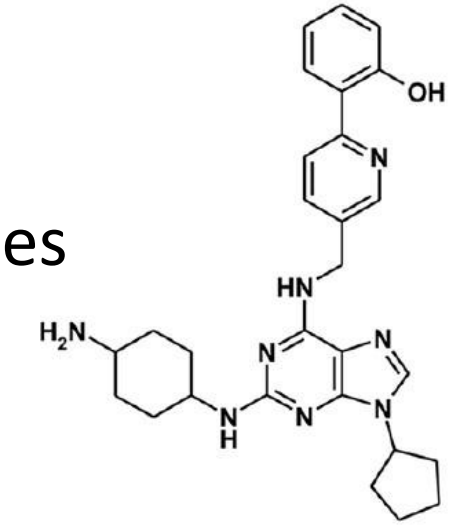


LigandFit, FlexX, DOCK 6.0, Autodock 4.0, MOE, Discover in Insight II, FlexiDock, Gold 3, …

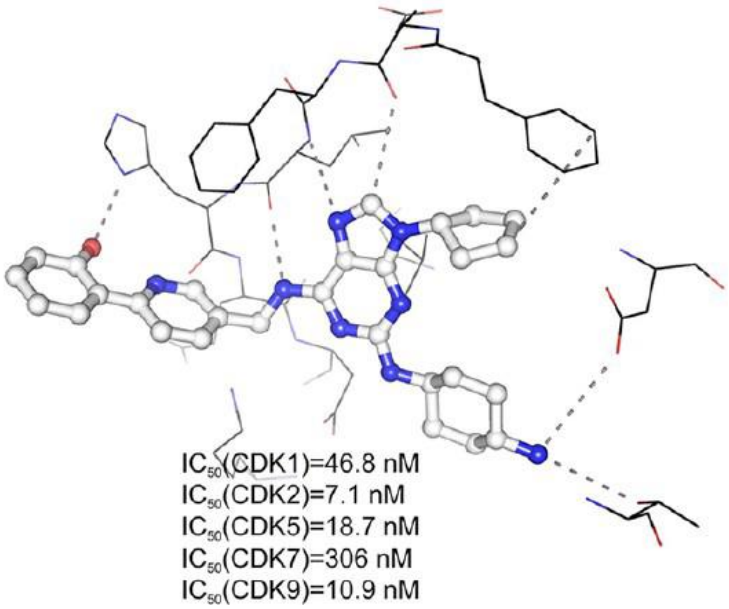
Example: CDK2 kinase

- Cell cycle regulation
  - Result: Inhibitors of CDK2 in nM range
  - Autodock Vina – speed
  - Ligand conformational troubles

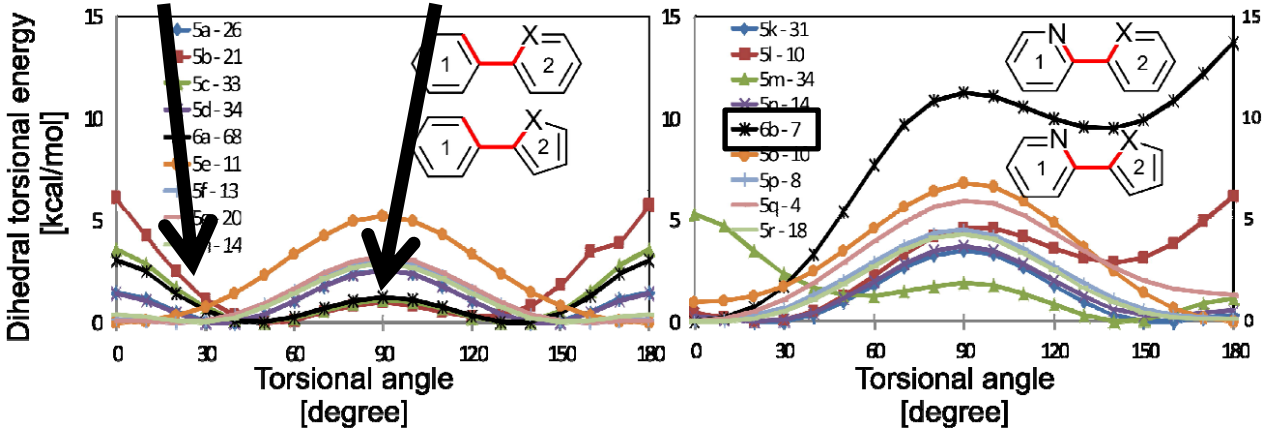
(planar NH close to aromatic ring,
torsional angle of biphenyl moiety)

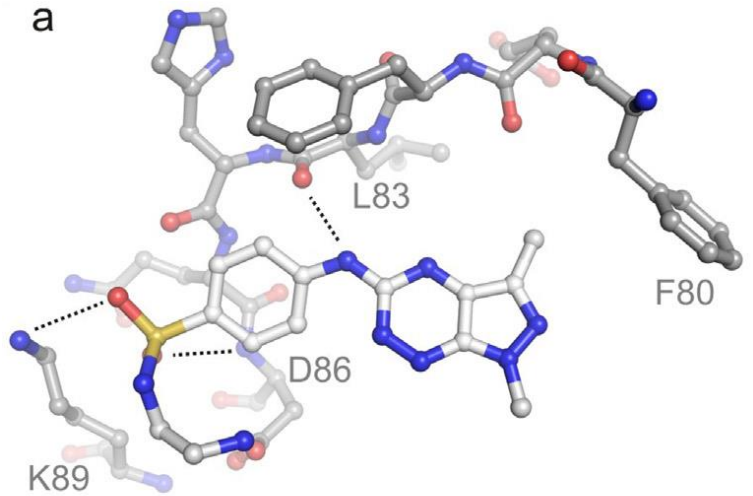


Compound **6b**



DFT scan vz docking programs



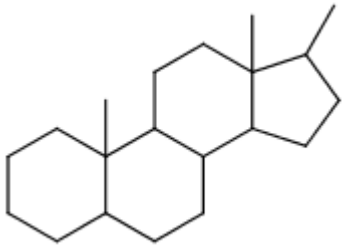


Gucký T,… **Berka K**, … et al *J. Med. Chem.*, 56 (15), 6234-6247, 2013
Mojzych M, … **Berka K**, … et al *Eur. J. Med. Chem.* 78, 217-224, 2014

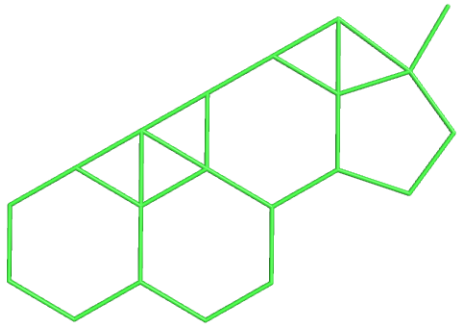
Ligand Structure Generation

Stereochemistry

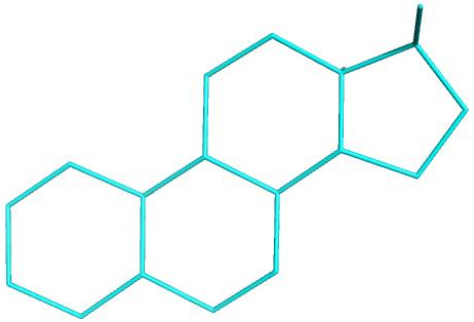
Experimentalist drawing

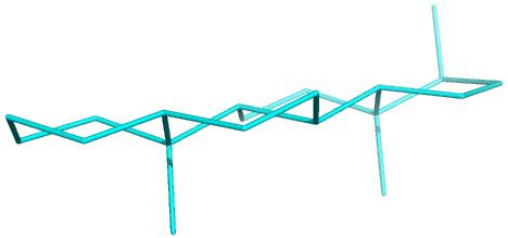


SDF

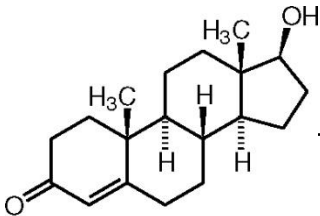


3D optimization

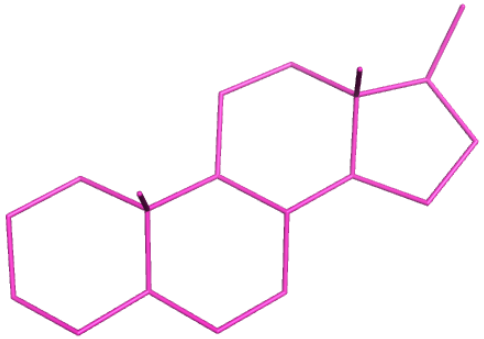


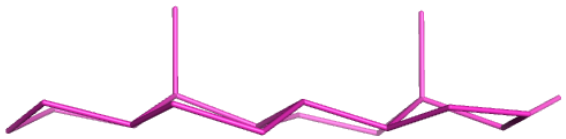


right stereoisomer



testosterone





Programs For Docking

- **DOCK** (I. D. Kuntz, UCSF)
- **AUTODOCK** (Arthur Olson, The Scripps Research Institute)
- **Vina** (Arthur Olson, The Scripps Research Institute)
- RosettaDOCK (Baker, Washington Univ., Gray, Johns Hopkins Univ.)
- ArgusLabs
- **GOLD**
- **FlexX**
- Hex
- Glide (Schrodinger)

Open Resources

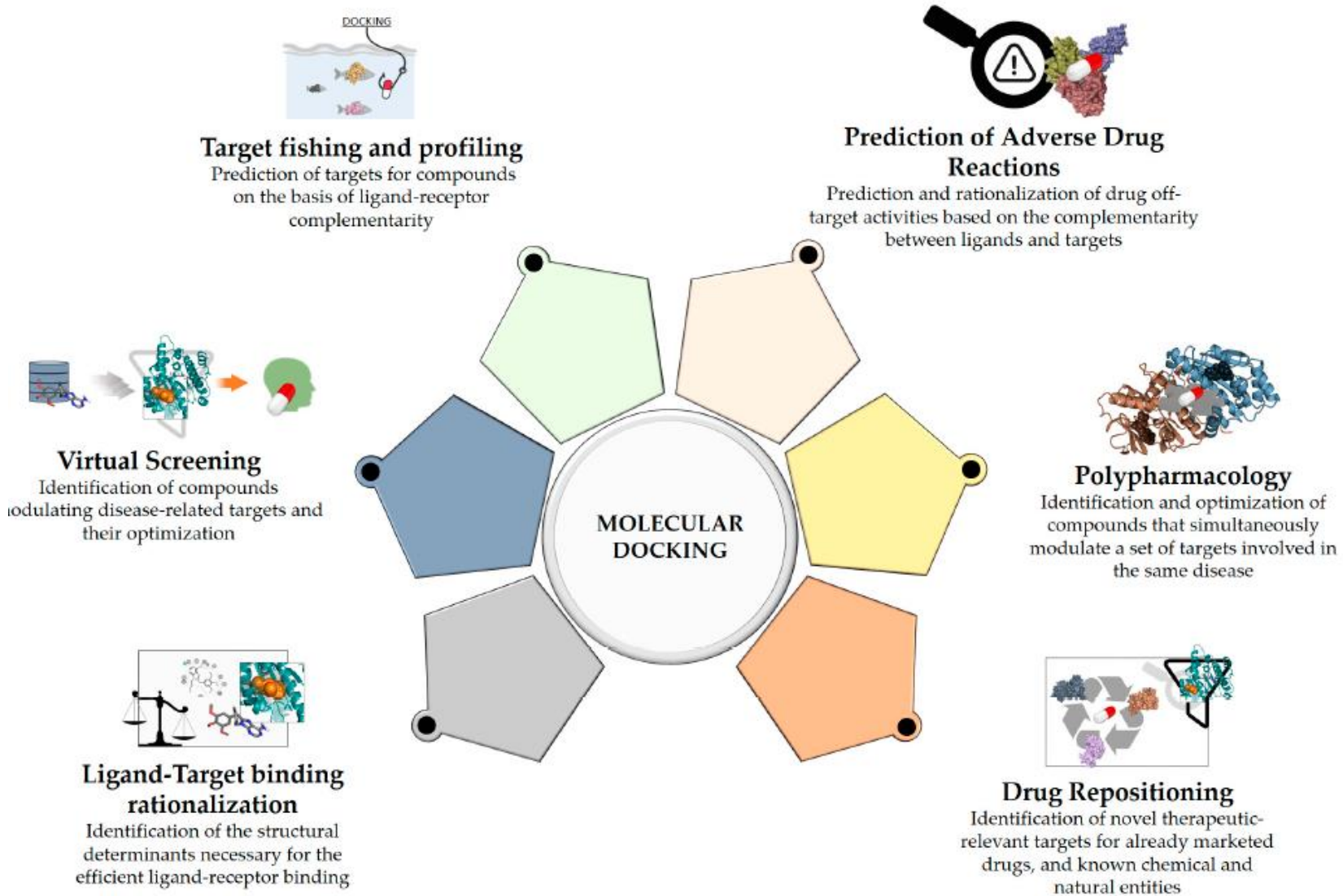
Name	URL	License	Activity
ADplugin	github.com/ADplugin	LGPL	A2
APBS	www.poissonboltzmann.org	BSD	A1
AutoDock	autodock.scripps.edu	GPL	C1
AutoDock Vina	vina.scripps.edu	Apache	C1
DockoMatic	sourceforge.net/projects/dockomatic	LGPL	B1
DOVIS	bhsai.org/software	GPL	C2
idock	github.com/HongjianLi/idock	Apache	A2
MOLA	www.esa.ipb.pt/~ruiabreu/mola	GPL	C2

A- active development
1- active usage

Name	URL	License	Activity
NNScore	nbcr.ucsd.edu/data/sw/hosted/nnscore	GPL	C1
Paradocks	github.com/cbaldauf/paradocks	GPL	A2
PyRx	pyrx.sourceforge.net	BSD	A1
rDock	rdock.sourceforge.net	LGPL	C1
RF-Score	github.com/HongjianLi/RF-Score	Apache	A2
smina	sourceforge.net/projects/smina	GPL	A1
VHELIBS	urvnutrigenomica-ctns.github.io/VHELIBS	GPL	A2
VinaLC	mvirdb1.llnl.gov/static_catsid/vina	Apache	C2
VinaMPI	cmb.ornl.gov/~sek	Apache	C2
Zodiac	sourceforge.net/projects/zodiac-zeden	GPL	C1

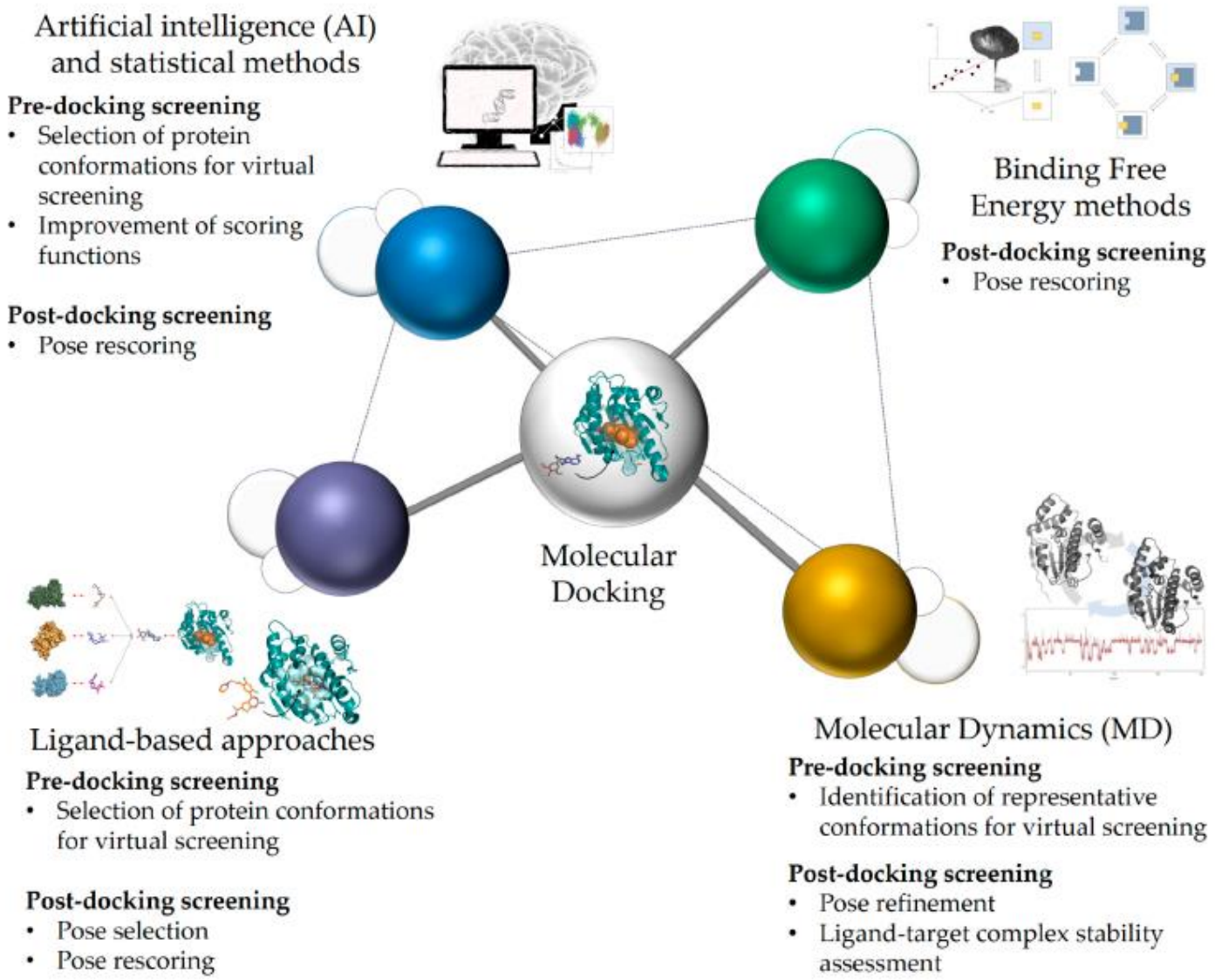
Pirhadi S, Sunseri J, Koes DR: Open source molecular modeling. J. Mol. Graph. Model. 2016, 69, 127-143
https://opensourcemolecularmodeling.github.io/

Usage of Docking



Pinzi L, Rastelli G: Molecular Docking: Shifting Paradigms in Drug Discovery. Int. J. Mol. Sci. **2019**, 20, 4331 doi:10.3390/ijms20184331

Usage of Docking – Interactions with other methods



Pinzi L, Rastelli G: Molecular Docking: Shifting Paradigms in Drug Discovery. IJMS **2019**, 20, 4331 doi:10.3390/ijms20184331

Quality control

- Redocking (back to Xtal)
  - RMSD < 2A
  - flexible ligand docking ~70%
  - test for scoring functions/docking programs
- Correlation plot ($r^2 > 0.5$)
  - ΔG_{eff}
- test sets – validation
  - GOLD test set, Astex set
  - decoys – ZINC, DUD (similar phys-chem., different structures)
- Virtual Screening
  - Enrichment factor
  - (BED)ROC curves

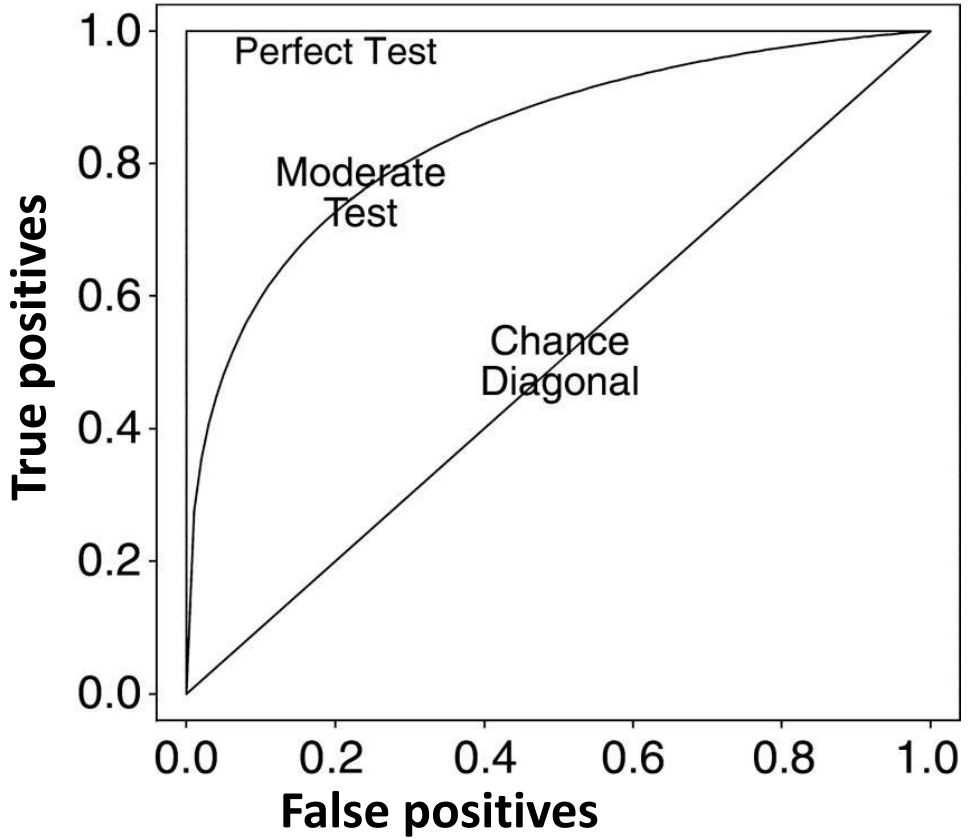
$$EF = \frac{a/n}{A/N}$$

- top (e.g. top10)
a – active
n - total
- overall

$$\Delta G_{eff} = \Delta G_{eff}/N_{nonHatoms}$$

ROC curve

- Receiver operator characteristic curve
- – signal to noise ratio



Sorting Quality

- ROC
  - receiver operating characteristic curve
- AUAC
  - area under the accumulation curve
- average rank of actives
- EF
  - enrichment factor
- RIE
  - robust initial enhancement
- BEDROC
  - Boltzmann-enhanced discrimination of receiver operating characteristic

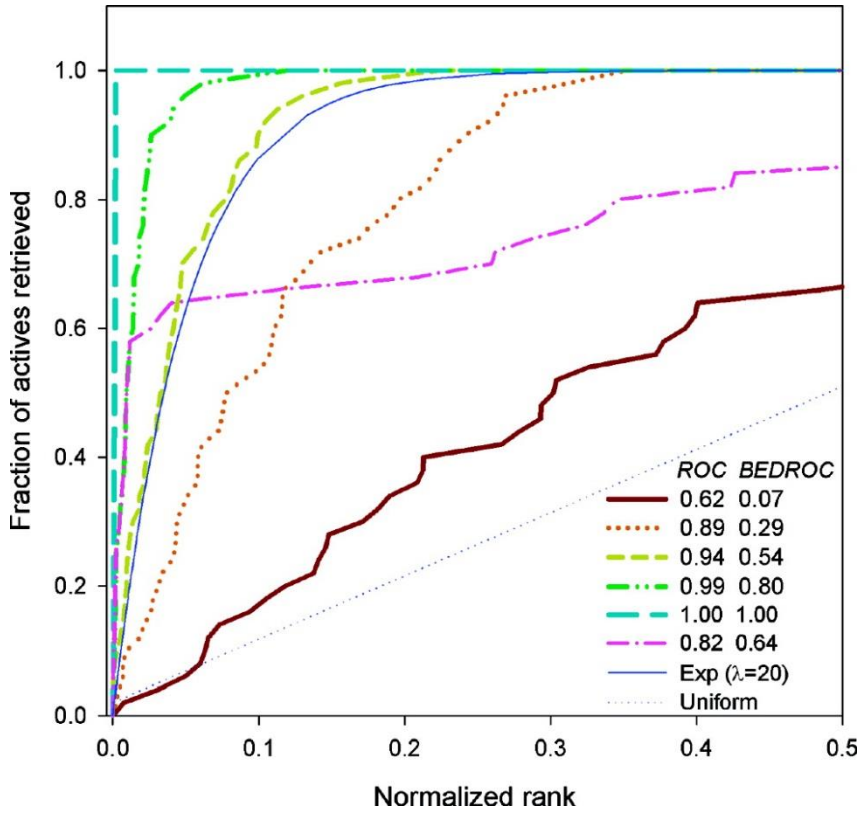


Figure 9 Different accumulation curves from sampling (n = 50, N = 25000) shown together with the corresponding *ROC* and *BEDROC* values where α = 20.0. An exact CDF with λ = 20 is also shown to highlight the fact that the *BEDROC* metric returns a value of 1/2 for a curve close to this CDF.

Published in: Jean-François Truchon; Christopher I. Bayly; *J. Chem. Inf. Model.* **2007,** 47, 488-508.
DOI: 10.1021/ci600426e

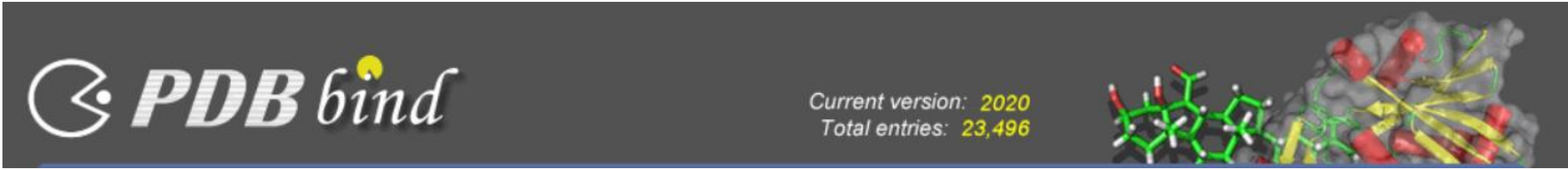
Benchmark datasets

Dataset	No. of drugs	No. of proteins	Drug-target interactions
Davis et al (2011)	68	442	30,056
Metz et al (2011)	1,421	156	35,259
Kinase Inhibitor BioActivity (KIBA)	2,116	229	118,254
ToxCast -	7,675	335	530,605

Thafar M, et al: Comparison Study of Computational Prediction Tools for Drug-Target Binding Affinities. Front. Chem. (2019) 7:782. doi: 10.3389/fchem.2019.00782

CASF-2016

- Comparative Assessment of Scoring Functions (SF)
  - scoring power – linear correlation with exp. binding data
  - ranking power – rank known ligands in precise binding poses
  - docking power – identify native ligand binding pose
  - screening power – identify true binders from decoys
- 285 protein-ligand complexes with good structures and reliable binding constants from http://www.pdbbind-cn.org/

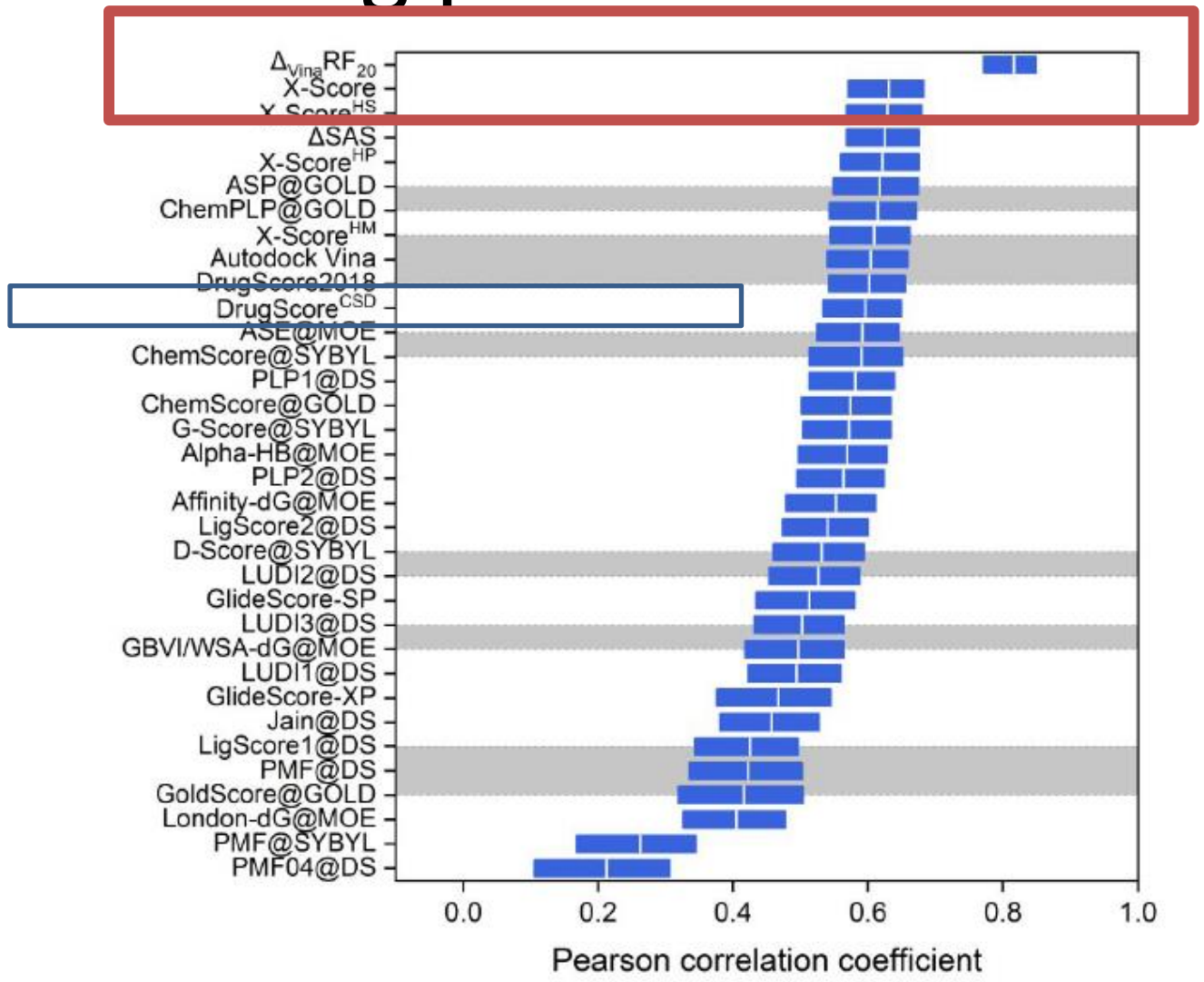


Su M. et al J. Chem. Inf. Model, 2019;59(2):895-913, doi:10.1021/acs.jcim.8b00545

CASF-2016 – scoring power

correlation with exp. binding data

- $\Delta_{Vina}RF_{20}$
  - descriptor ML
- X-Score
  – empirical SF
- ΔSAS
  - single descriptor
- ASP@GOLD
  – knowledge-based SF
- ChemPLP@GOLD
  – empirical SF
- Autodock Vina
  – empirical SF

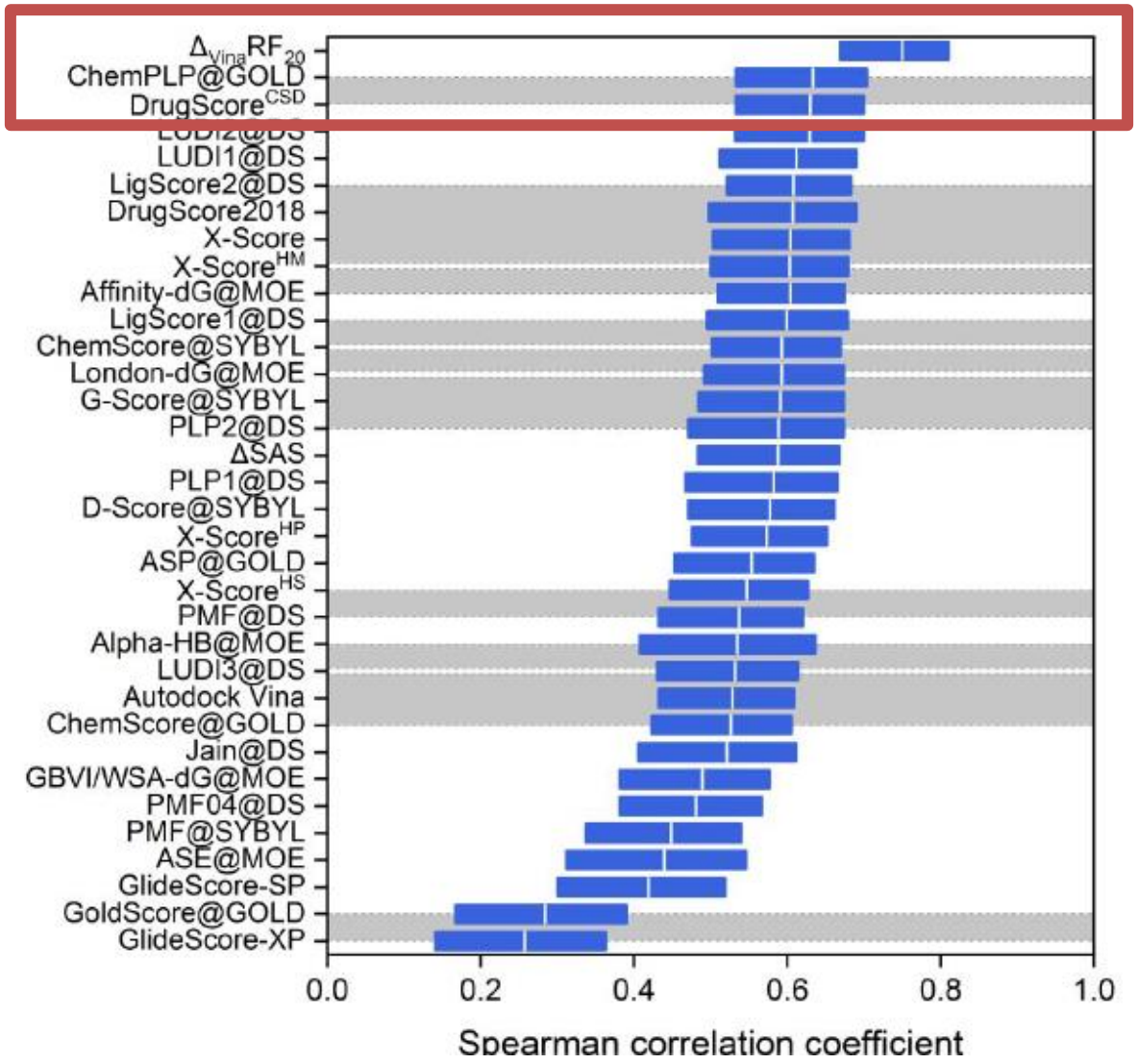


Su M. et al J. Chem. Inf. Model, 2019;59(2):895-913, doi:10.1021/acs.jcim.8b00545

CASF-2016 – ranking power

rank known ligands in precise binding poses

- $\Delta_{Vina}RF_{20}$
  - descriptor ML
- ChemPLP@GOLD
  – empirical SF
- $DrugScore^{CSD}$
  – knowledge-based SF
- LUDI@DiscoveryStudio
  - empirical SF
- LigScore@DiscoveryStud
  – empirical SF
- X-Score
  – empirical SF

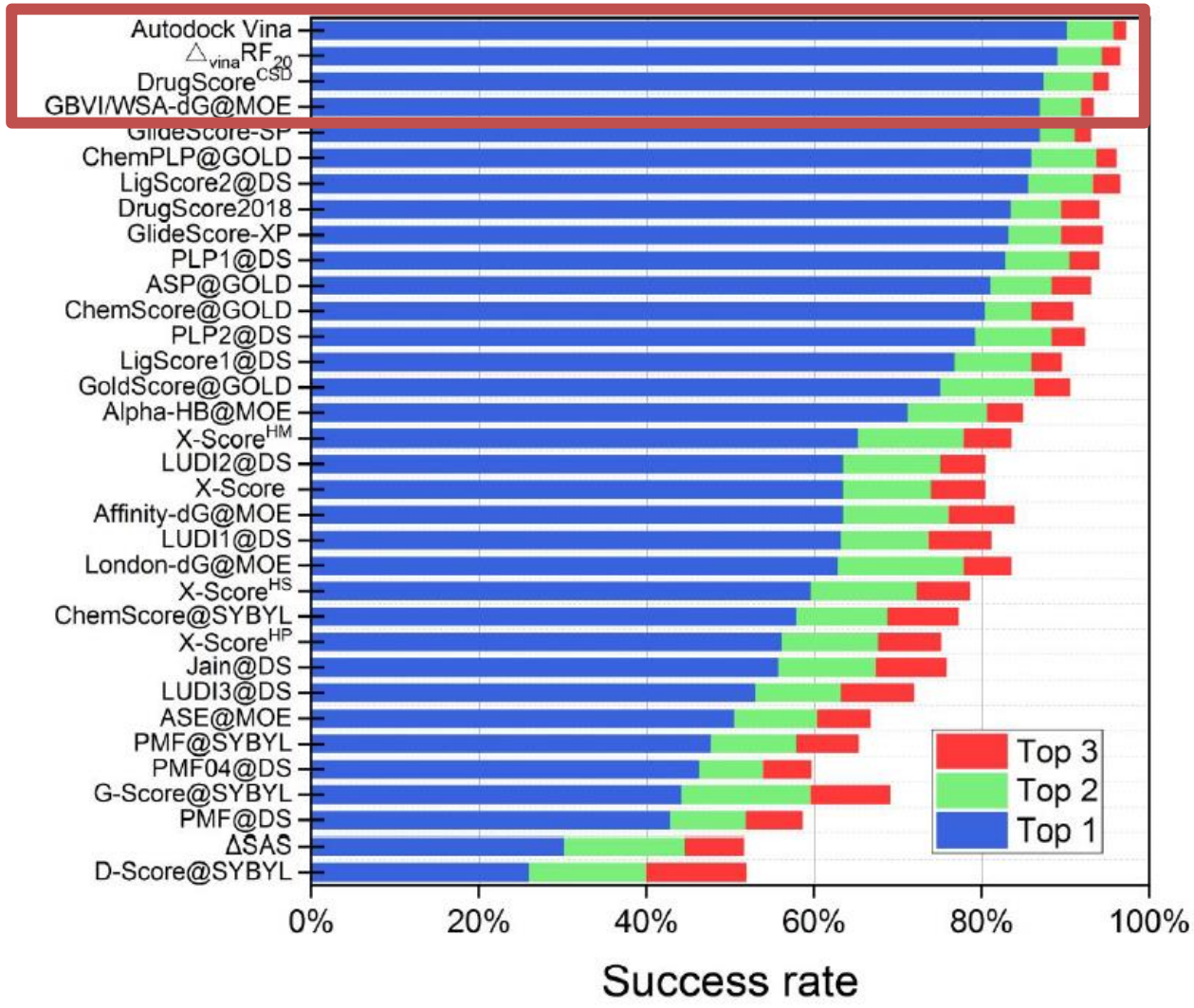


Su M. et al J. Chem. Inf. Model, 2019;59(2):895-913, doi:10.1021/acs.jcim.8b00545

CASF-2016 - docking power

identify native ligand binding pose

- Autodock Vina
  – empirical SF
- $\Delta_{Vina}RF_{20}$
  – descriptor ML
- $DrugScore^{CSD}$
  – knowledge-based SF
- GBVI/WSA-dG@MOE
  – physics-based SF
- GlideScore@Schrodinger
  – empirical SF
- ChemPLP@GOLD
  – empirical SF

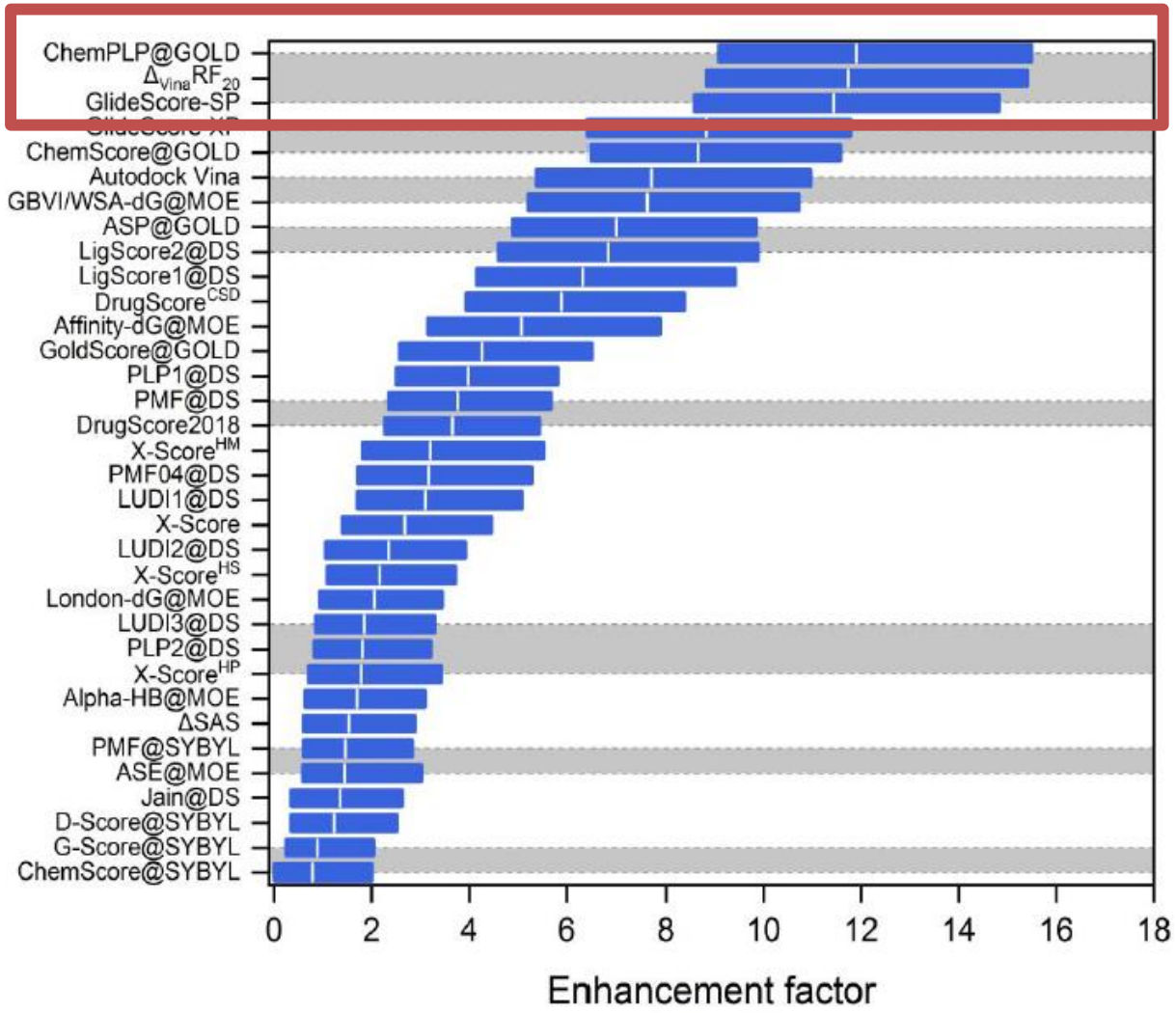


Su M. et al J. Chem. Inf. Model, 2019;59(2):895-913, doi:10.1021/acs.jcim.8b00545

CASF-2016 - screening power

identify true binders from decoys

- ChemPLP@GOLD
  – empirical SF
- $\Delta_{Vina}RF_{20}$
  – descriptor ML
- GlideScore@Schrodinger
  – empirical SF
- ChemScore@GOLD
  - empirical SF
- Autodock Vina
  – empirical SF
- GBVI/WSA-dG@MOE
  – physics-based SF



Su M. et al J. Chem. Inf. Model, 2019;59(2):895-913, doi:10.1021/acs.jcim.8b00545

Docking Take Home Message

- Usable in SAR (structure-activity relationship)
  - explore the interactions between ligands and receptor
  - can lead drug development
- Troubles
  - Ligand preparation – 3D generation, torsion angles
  - Receptor preparation – protein flexibility
  - Scoring function – identification of right binding pose, size of ligand issue