Drug Design

Molecular docking

Karel Berka

**Dpt. Physical Chemistry, RCPTM, Faculty of Science,
Palacky University in Olomouc**

Motto

tj. 18 hod 8 min 18 s

Outline

- Structure-based drug design (SBDD)
  - Docking
  - Virtual screening
  - de novo design
  - Pharmacophore search
- Ligand-based drug design (LBDD)
  - Similarity matching
  - Pharmacophore search
  - QSAR

Possibilities of Drug Design

	Known ligand	Unknown ligand
Known target structure	**Structure-based drug design (SBDD)** Docking	***De novo* design**
Unknown target structure	**Ligand-based drug design (LBDD)** *1 or more ligands* • Similarity search *Several ligands* • Pharmacophore *Large number of ligands (20+)* • Quantitative Structure-Activity Relationships (QSAR)	**CADD not possible** some experimental data needed ADMET filtering

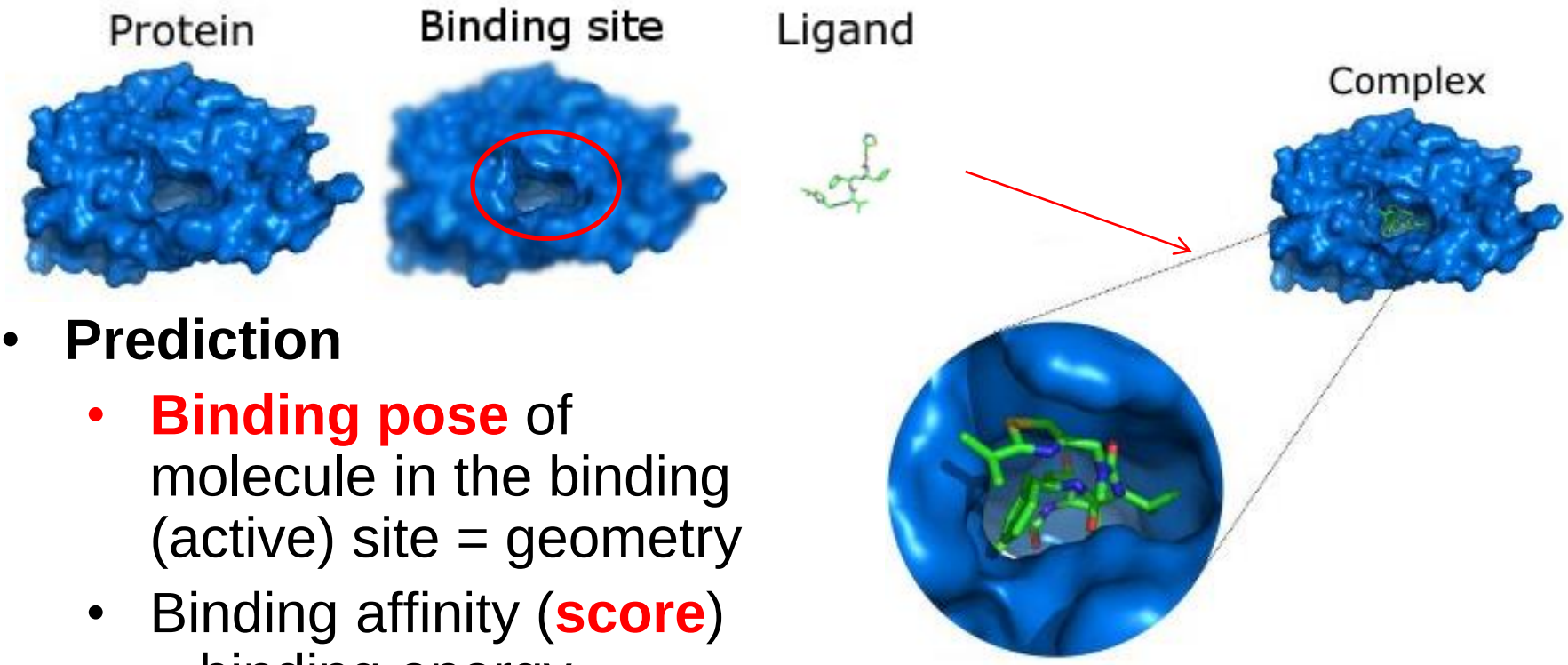
Molecular Docking Idea

- Finding the best ”fit“ of ligand to receptor

Bill Watterson: Calvin a Hobbes

Molecular Docking

Computational method mimicking binding of ligand to receptor

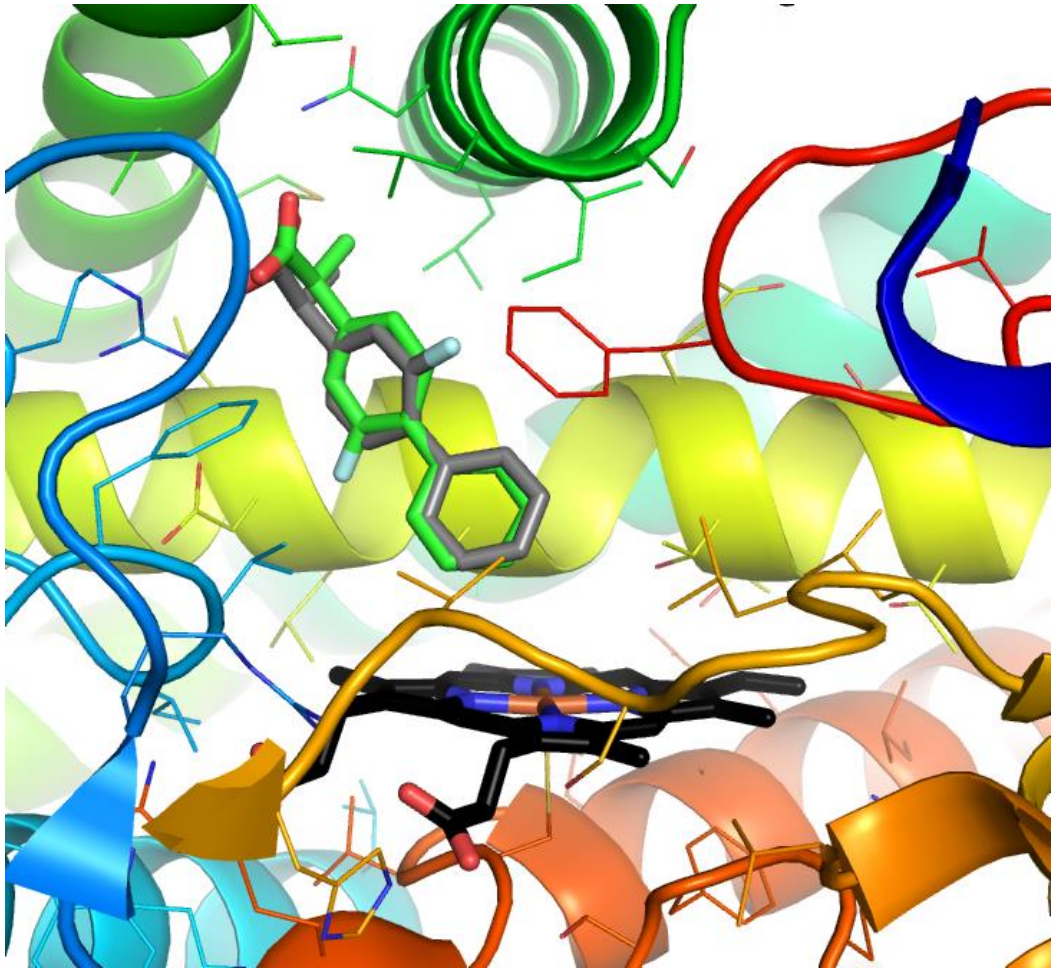


- **Prediction**
  - **Binding pose** of molecule in the binding (active) site = geometry
  - Binding affinity (**score**) = binding energy

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

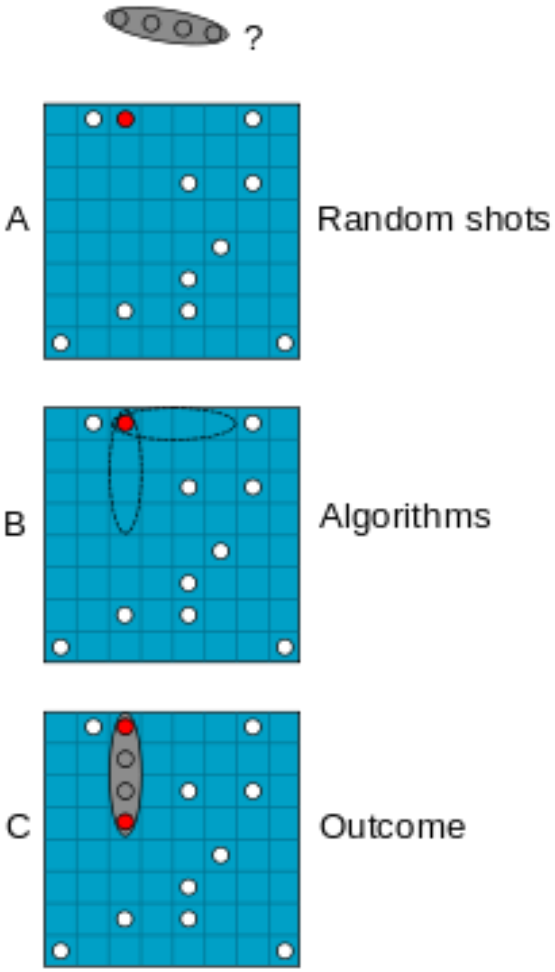
Binding Pose

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



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



Energetics

- Equilibrium binding constant

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

- correspond to free energy of binding:

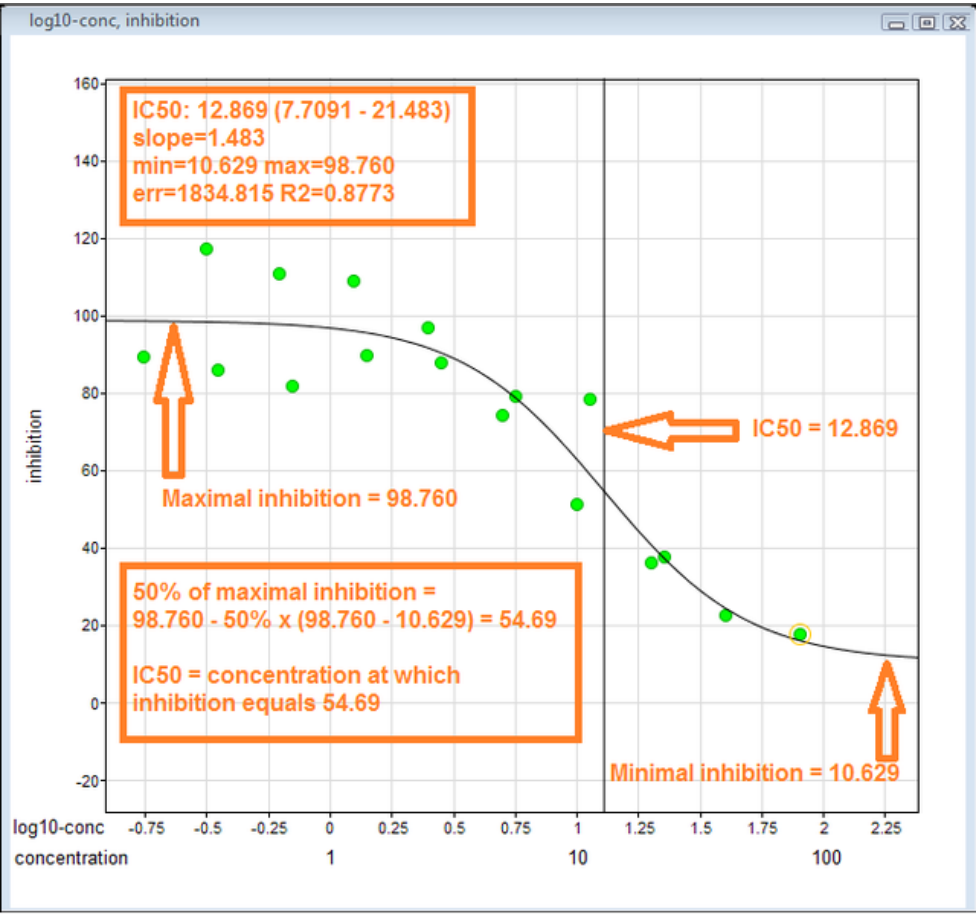
$$\Delta G_{bind} = -RT \ln K_d$$

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)

Molecular Interactions

Enthalpy:

- Electrostatics
  (partial charges)
- van der Waals
  (dispersion and repulsion)
- Hydrogen bonding
  (directionality)
- Desolvatation
  (cavitation energy)

Entropy

- Conformation selection
  (flexibility)
- Solvatation
  (hydrophobic effect)

Scoring Function

- Binding affinity approximation

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

- It should be:
  - Quick
  - Score the right pose the best
- Parameterized against known binding poses and affinities
- Types:
  - Force-field (DOCK, Autodock, GoldScore)
  - Empirical (Glide, ChemScore)
  - Knowledge-based (DrugScore)

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

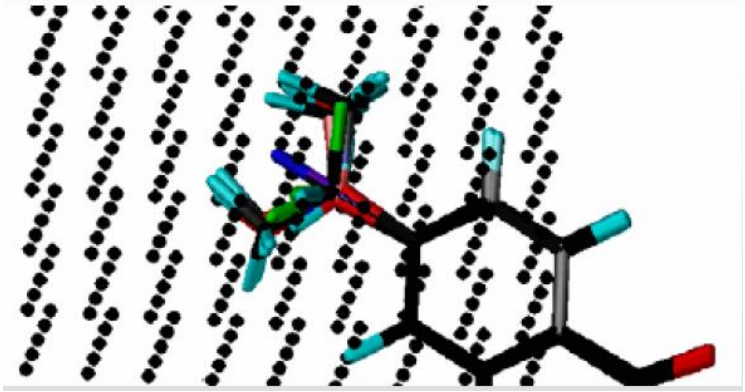
- Knowledge-based – based on protein-ligand complexes
  - Boltzmann hypothesis
    - typical binding motives -> stronger binding
  - PMF, DrugScore, ASP

Force-field Scoring Functions

- **Physical interaction terms**

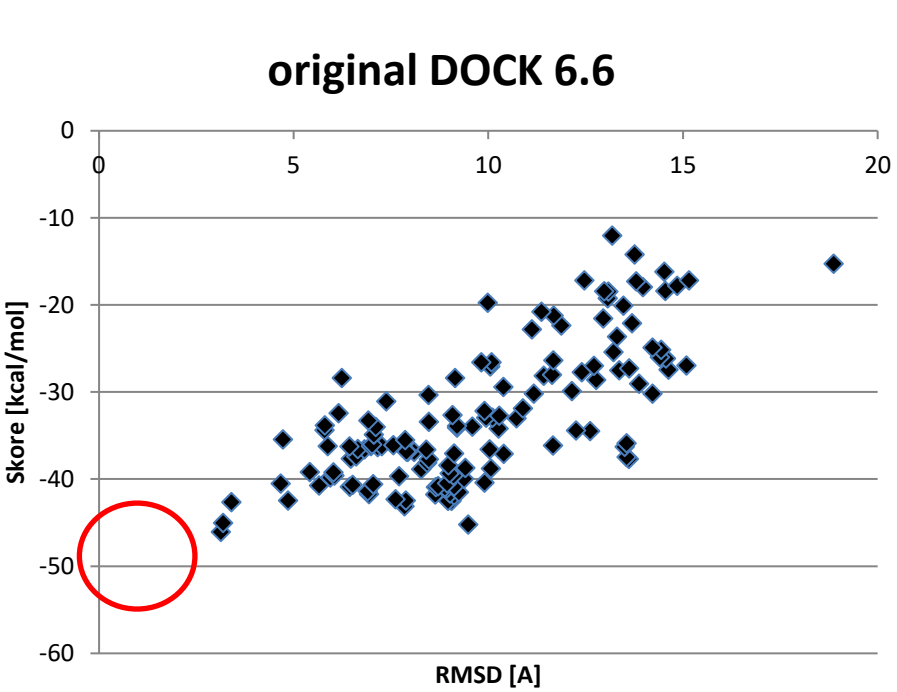
  $E = E_{bond} + E_{angle} + E_{dih} + E_{coulomb} + E_{vdw} + E_{solv}$

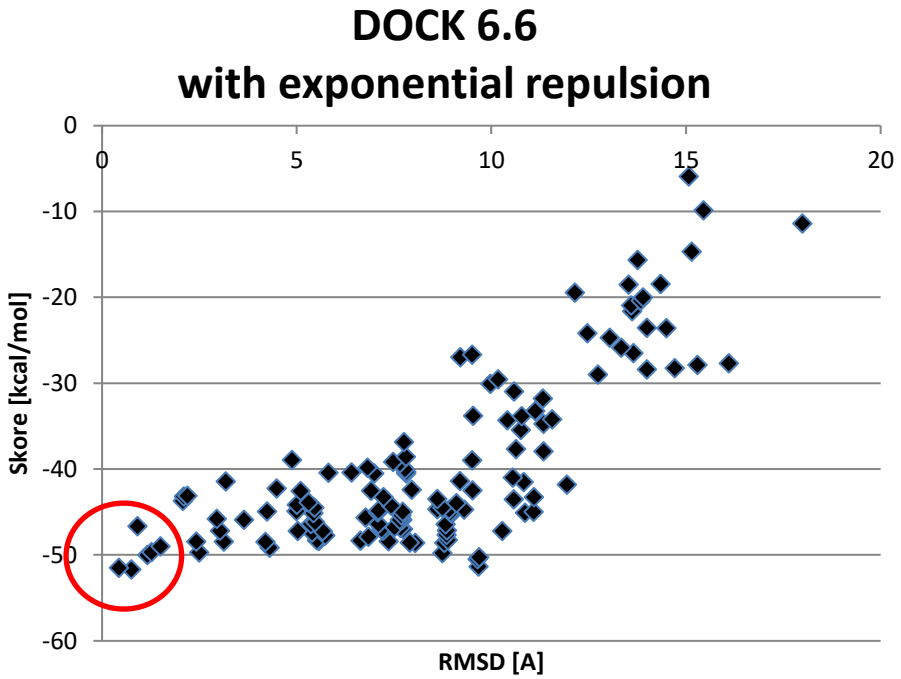
- Often only **intermolecular** terms ($E_{coul} + E_{vdw} + E_{solv}$)
- **Intramolecular** are usually changed to rigid (bonds, angles) or screened by some value (dihedrals by 5 deg)
- **Grid** – time-saving
  - Protein is divided into grid and interactions are pre-calculated at each point
  - Ligands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



Scoring Function Problems Example

- Problems:
  - Repulsion
  - Electrostatics





Bazgier V, Banáš P, **Berka K**, Otyepka M, *in preparation*

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

Chemscore

- Original Chemscore function for binding free energies

$$\Delta G_{binding} = \Delta G_o + \Delta G_{hbond} S_{hbond} + \Delta G_{metal} S_{metal} + \Delta G_{lipo} S_{lipo} + \Delta G_{rot} H_{rot}$$

- S_{hbond} – hydrogen bonding
- S_{lipo} – lipophilic interactions
- S_{metal} – acceptor-metal interactions
- H_{rot} – loss of conformational entropy on ligand binding

J. Comput. Aided Mol. Des. 11, 425-445, 1997

Chemscore

Chemscore for docking

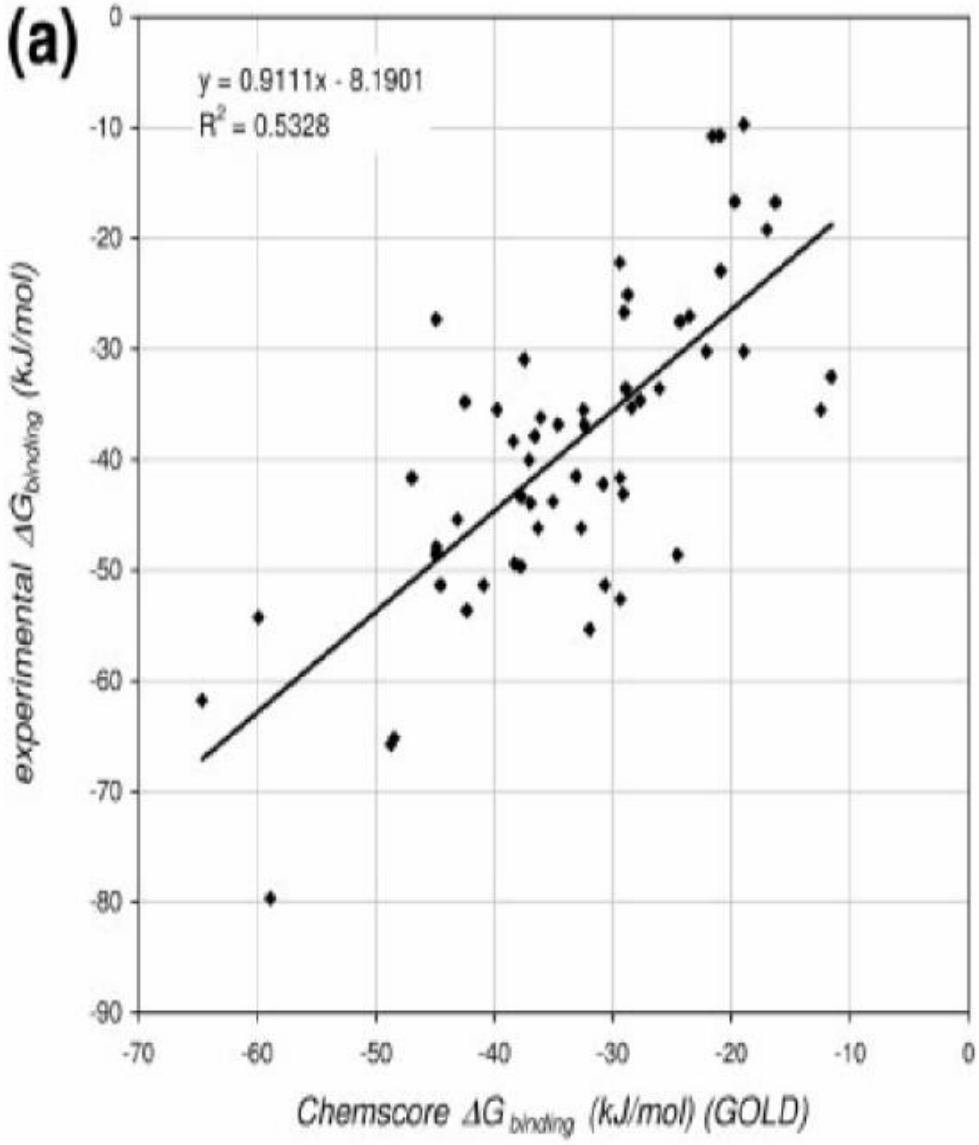
- Add more terms – clash, ligand internal, protein-ligand covalent

$$\Delta G'_{binding} = \Delta G_{binding} + E_{clash} + E_{int} + E_{cov}$$

- Complex functional forms – look them up!
- Parameters carefully rederived

Proteins 52, 609-623, 2003

Chemscore Accuracy



Correlation coefficient – r

r^2 <-1, 0, 1>

-1 – anticorrelation

0 – no correlation

1 – full correlation

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Empirical Scoring Functions Problems

- Heavy dependence on training set
- Can have missing interaction terms
  - metal-ion
- Parameterized on success
  - Use of molecules that bind in parameterization => artificial binding of molecules that otherwise would not bind

  => Use of **decoys** – molecules, which are of similar size as those really binding but not binding

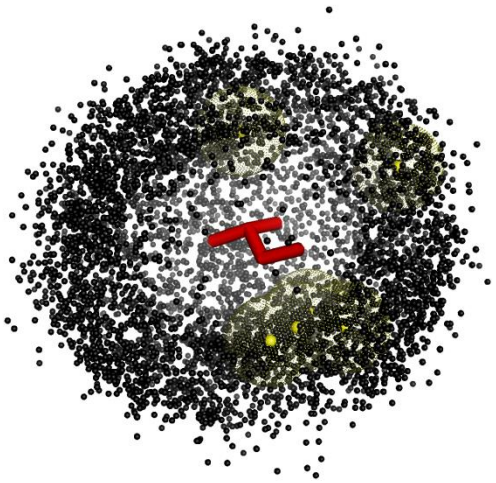
Knowledge-based Function

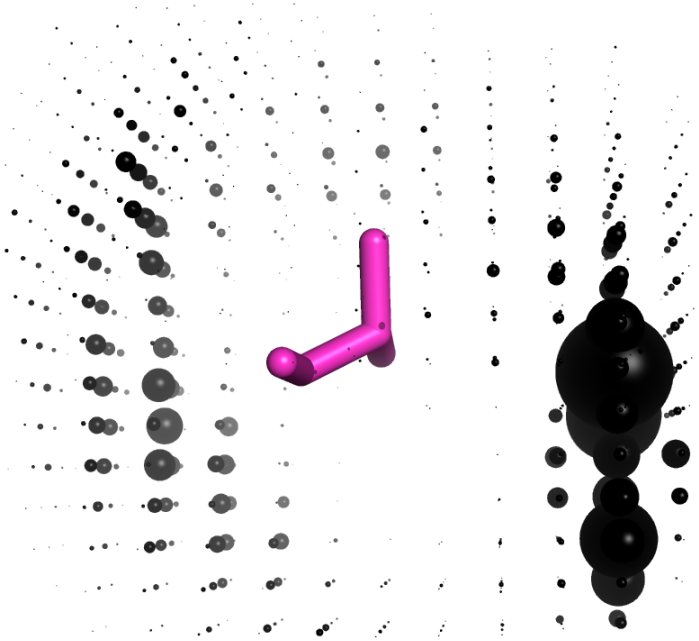
Correlation of structural data from ligand/protein complexes with free energy of binding

- 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
  - Flexible docking

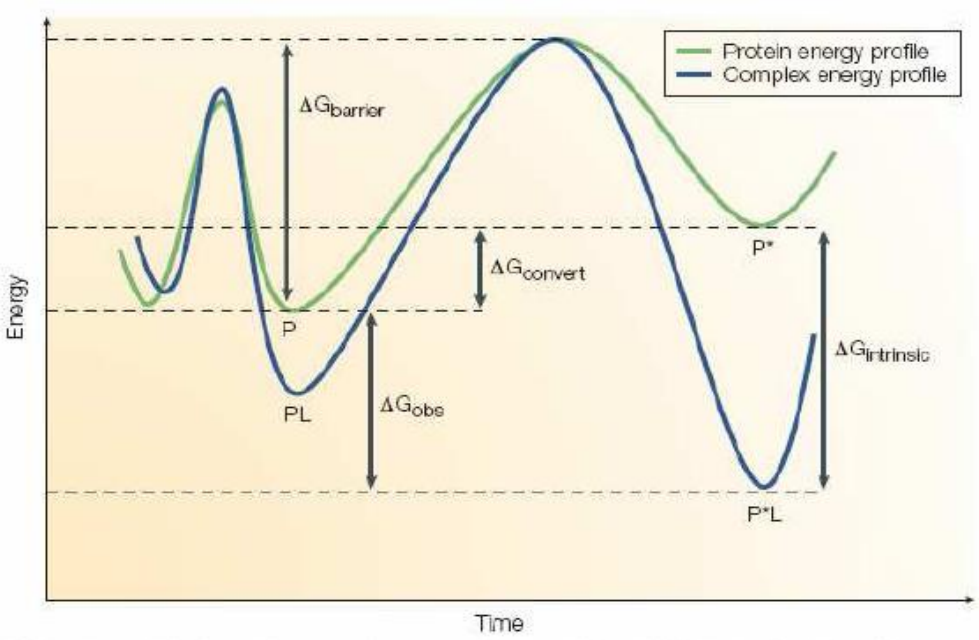
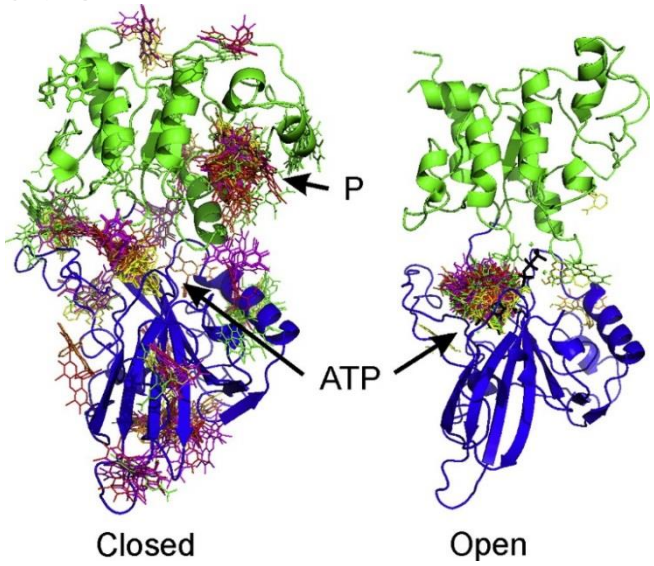


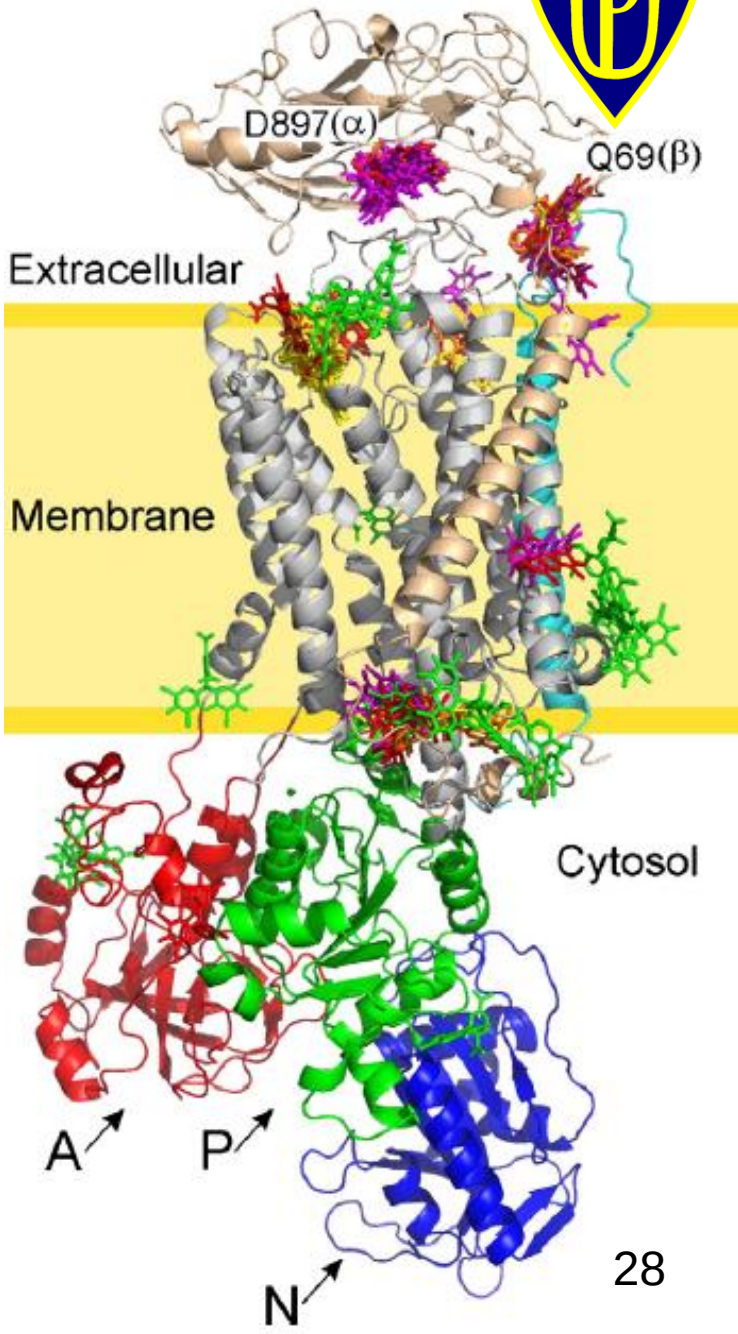
Figure 1 | **Protein mobility and ligand binding.** A protein is considered to exist in two conformations (P and P*) with an energy difference $\Delta G_{convert}$. The ligand (L) can bind the protein (P) to give a complex (PL), or bind to P* to give a complex (P*L). Although P* has a higher free energy, 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. *submitted, 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 c
  1. Side chains only
     (docking selected sidechains together with ligands)
  2. Docking into several structures of protein
     Larger movements can be taken into account

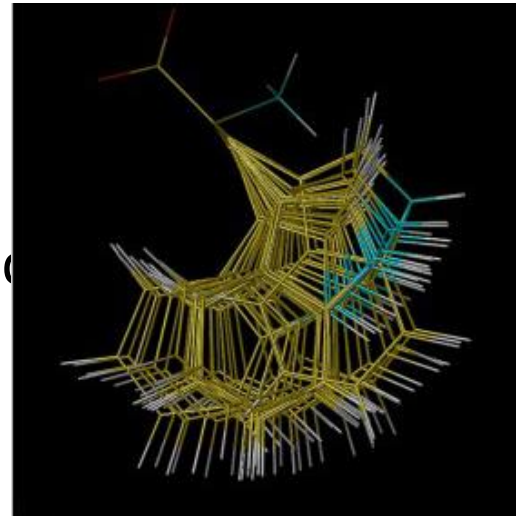
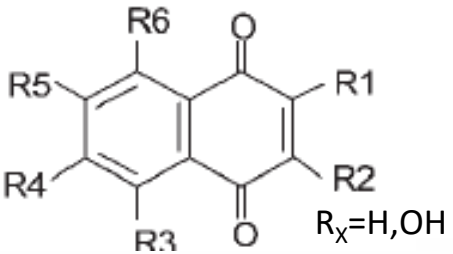


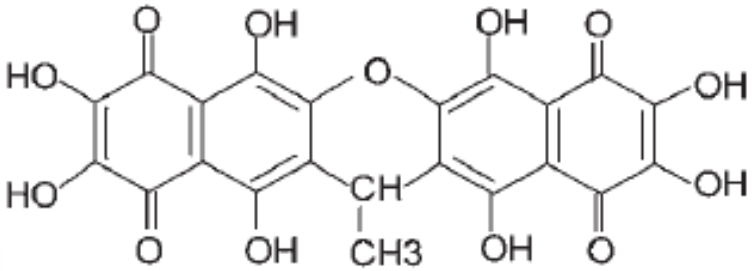
Image: Cláudio M. Soares, Protein Modelling Laboratory, http://www.itqb.unl.pt/labs/protein-modelling/activities/psccip-pf

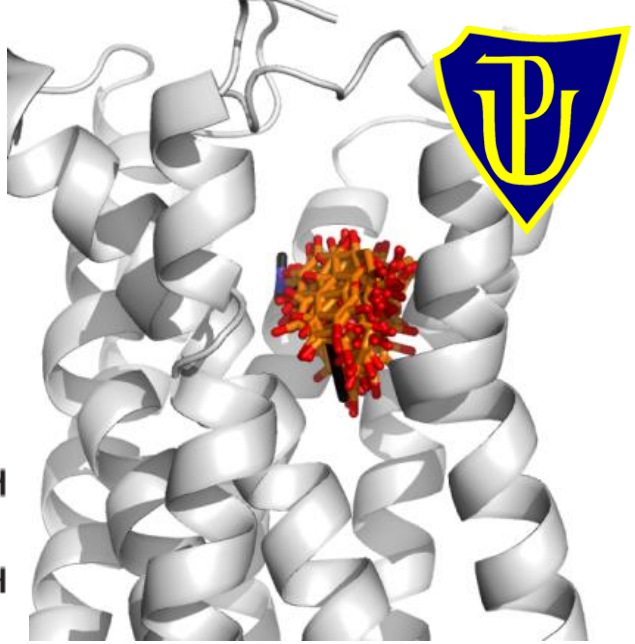
Example 2:
H1R receptor

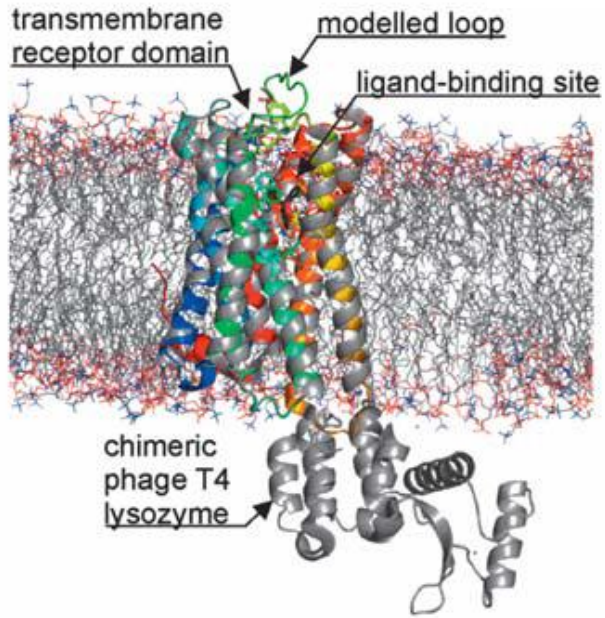
- Antiallergic compounds

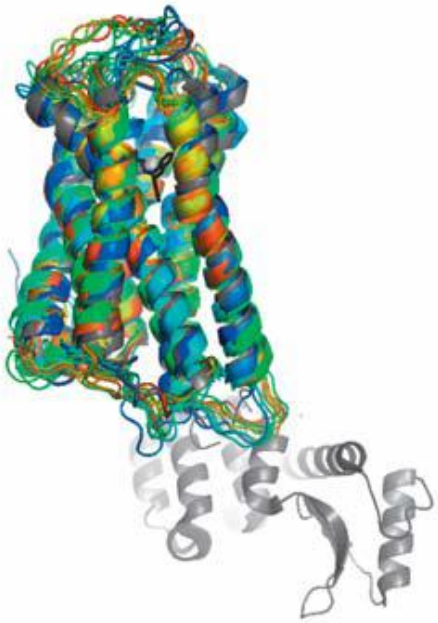
dG/n(atoms) – monomers are more active than dimers

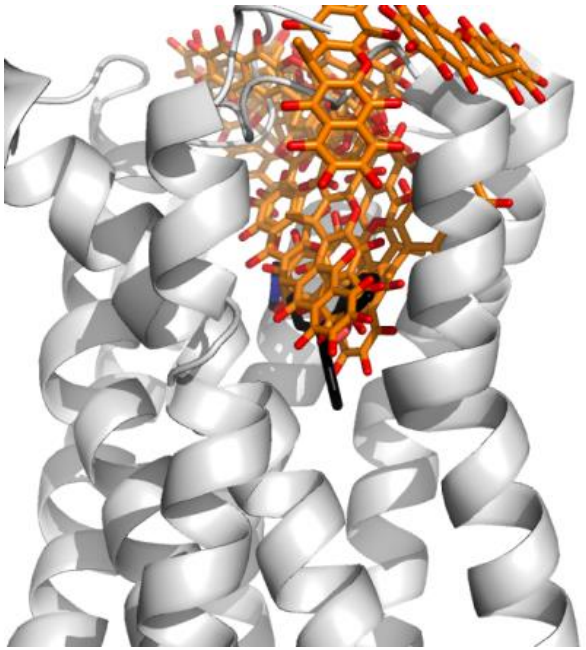








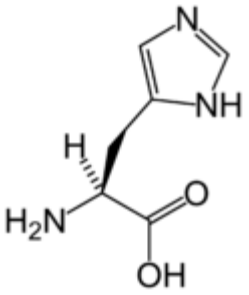


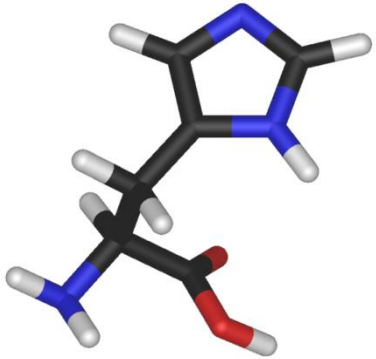


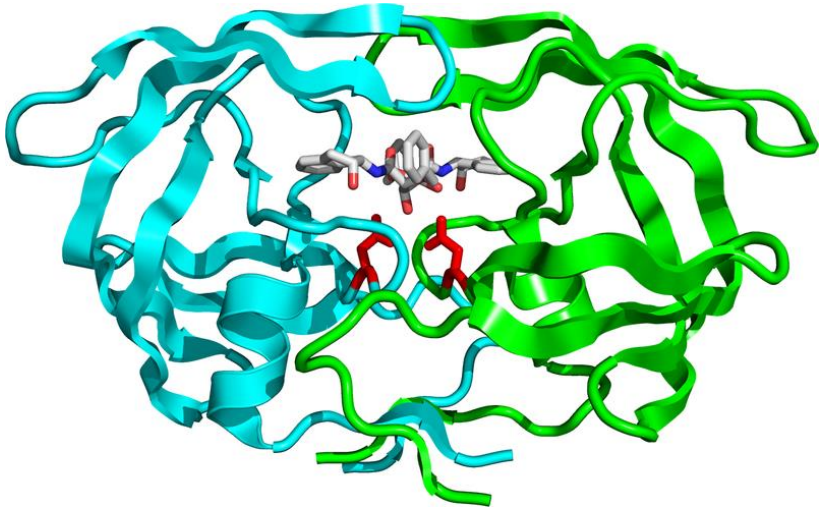
Pozharitskaya ON, …, **Berka K**, … et al. *Planta Med.* 79(18), 1698-1704, 2013.

Receptor Preparation

- protonation of aminoacids
  - His (pKa ~ 6.04)
  - Surroundings pKa shifts
    (Asp in HIV protease)
- tautomerization
- rotamers
- pre-selection change results significantly

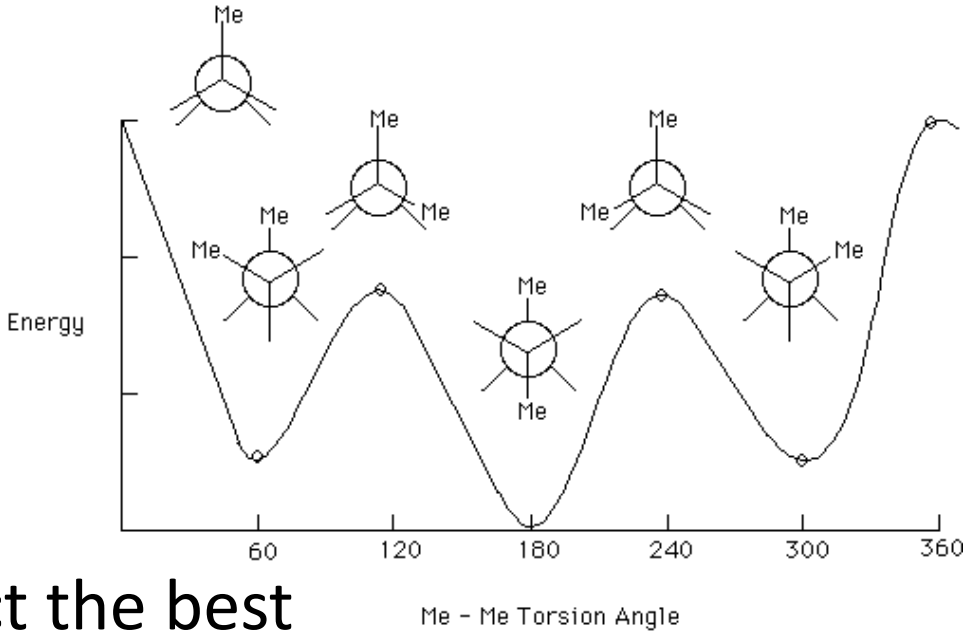




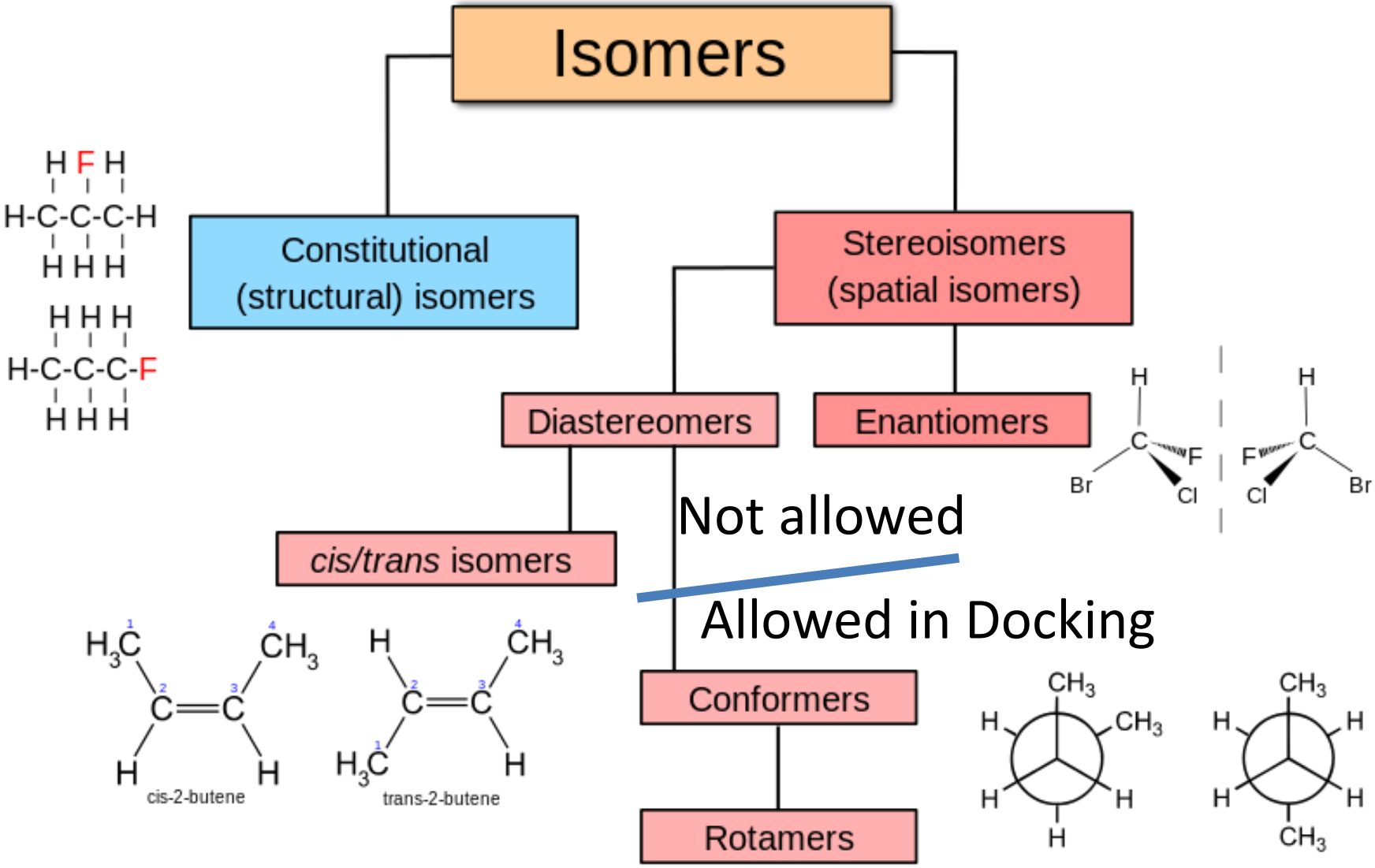


Ligand Preparation

- 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

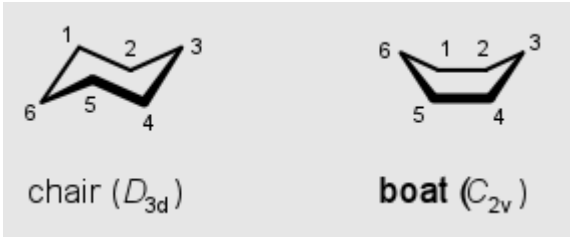
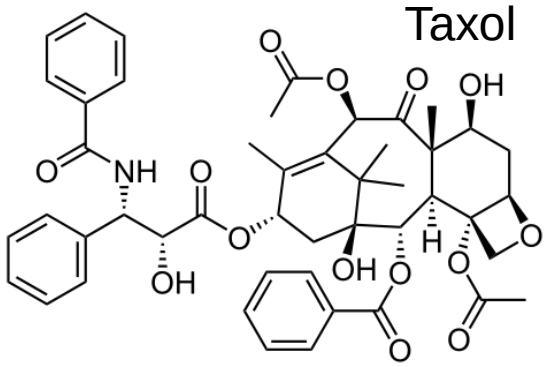
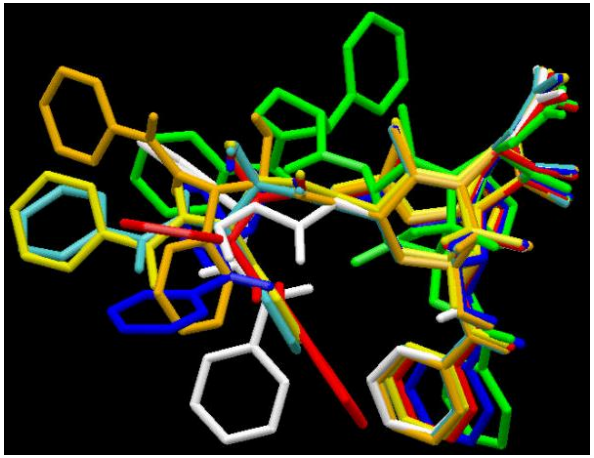


Image: IUPAC Gold Book

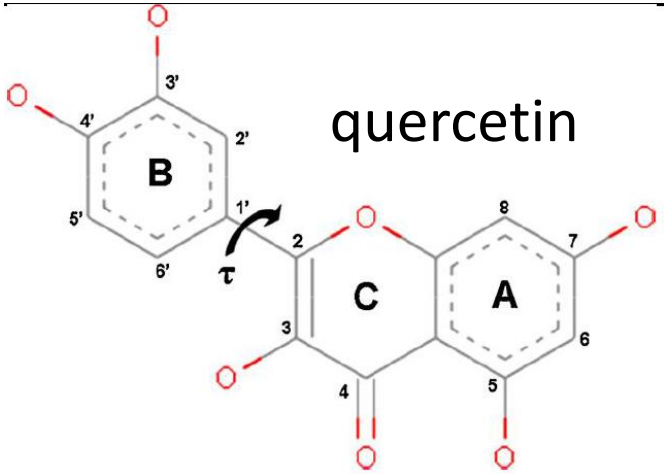


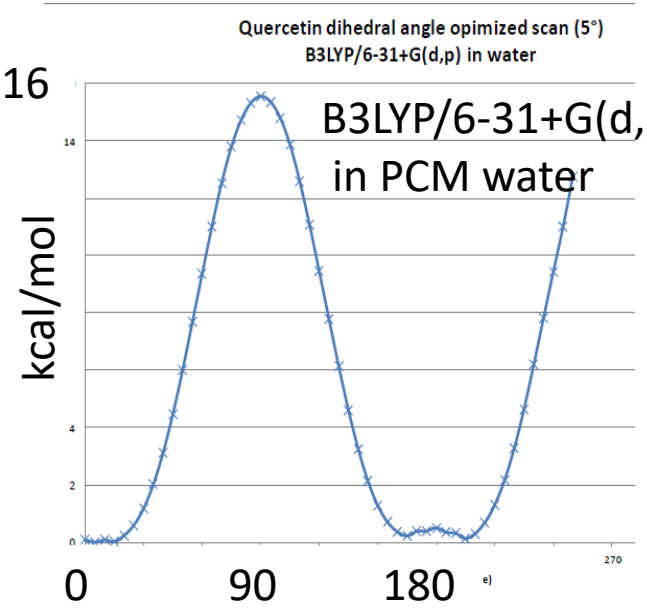


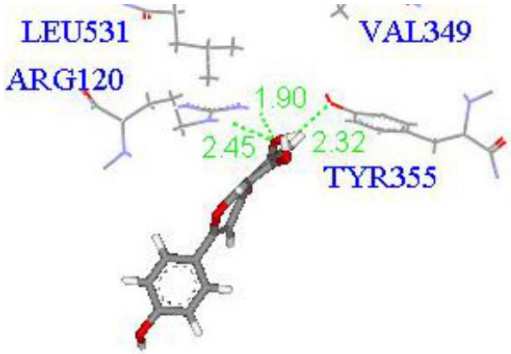
Lakdawala *et al. BMC Chemical Biology* 2001 **1**:2

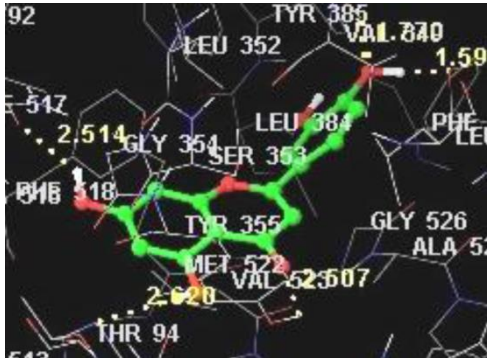
Ligand Structure Generation

Torsion angles



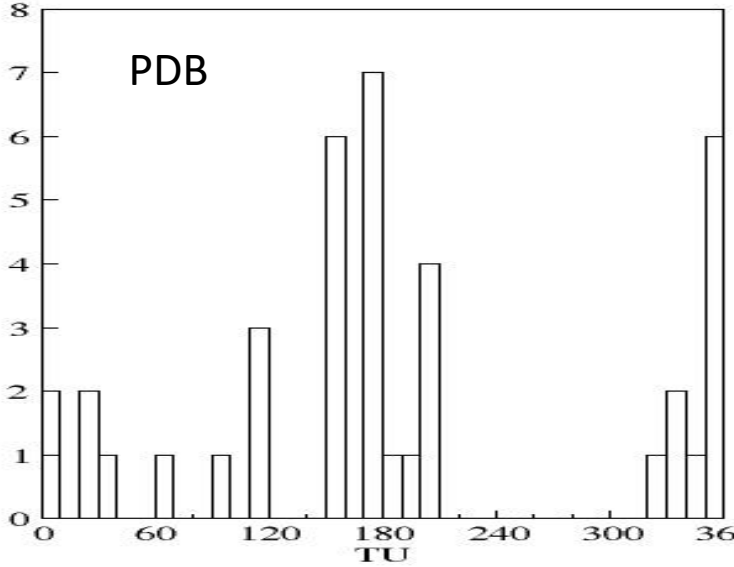






Accelrys DSmodelling 1.2 Wu, Chien-Ming et al *Int.J. Mol. Sci.* 8 (2007): 830–841.

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

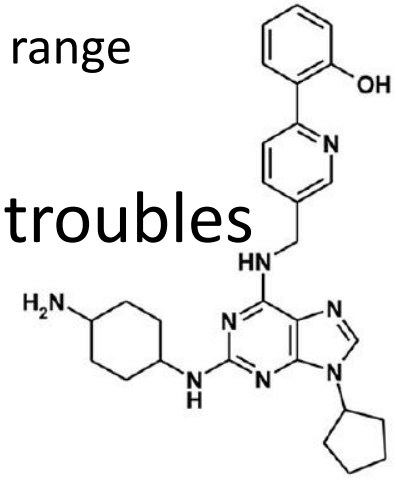


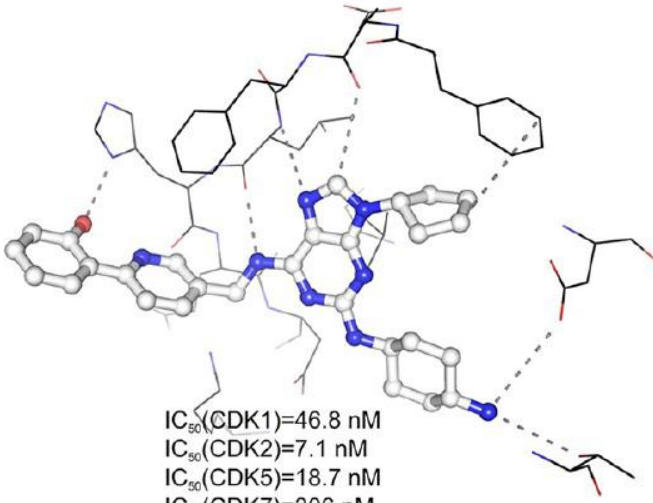
LigandFit, FlexX, DOCK 6.0, Autodock 4.0, MOE, Discover in Insight II, FlexiDock, Gold 3, …

Example: CDK2 kinase

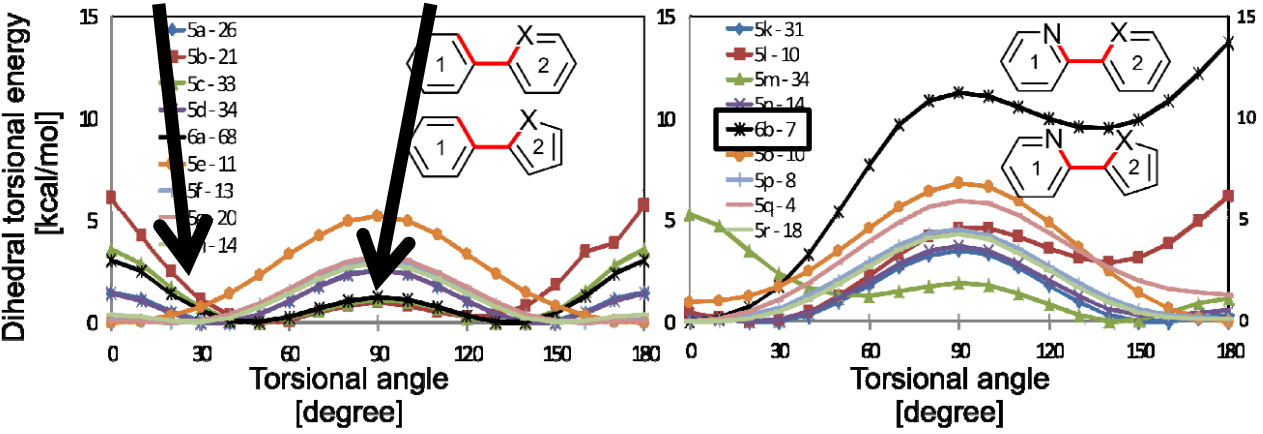
- Cell cycle regulation
  - Result: Inhibitors of CDK2 in nM range
  - Autodock Vina – speed
  - Ligand conformational troubles

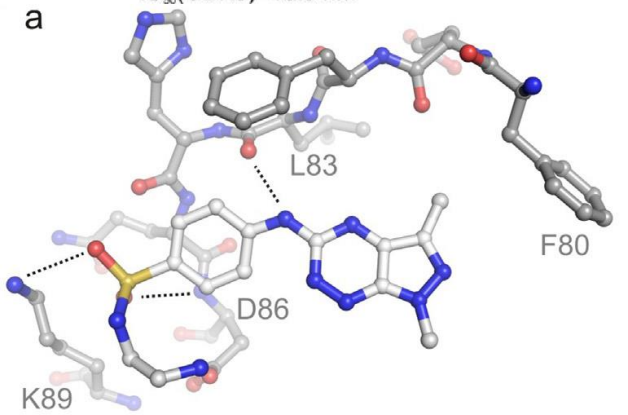
(planar NH close to aromatic ring,
torsional angle of biphenyl moiety)





DFT scan vz docking programs



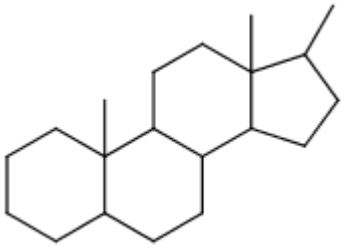


Gucký T,… **Berka K**, … et al *J. Med. Chem.*, 56 (15), 6234-6247, 2013
Mojzych M, … **Berka K**, … et al *Eur. J. Med. Chem.* 78, 217-224, 2014

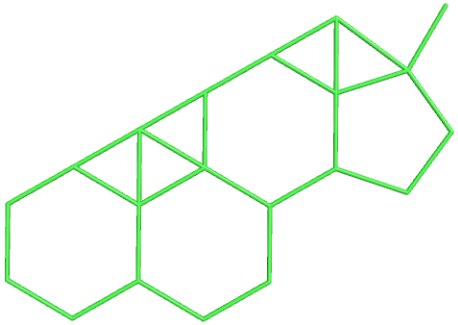
Ligand Structure Generation

Stereochemistry

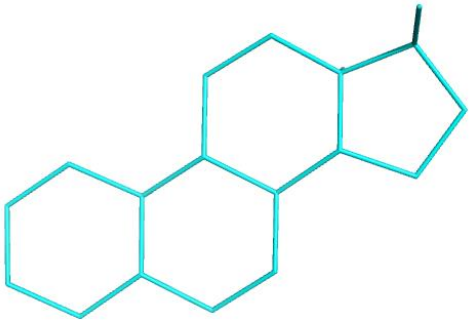
Experimentalist drawing

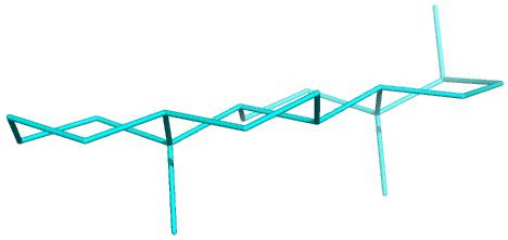


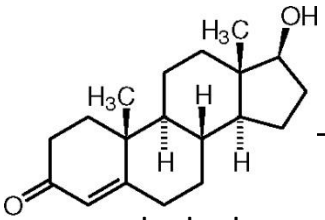
SDF



3D optimization

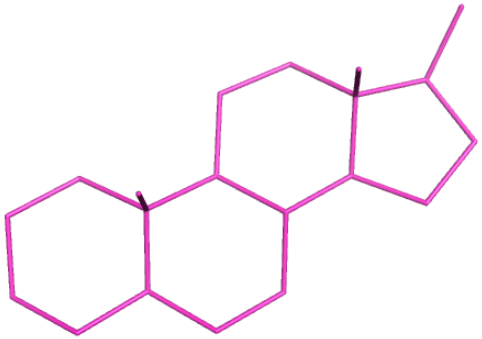


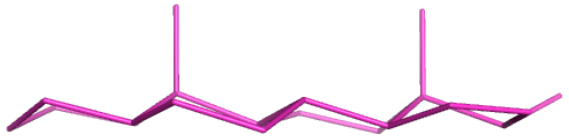




testosterone

right stereoisomer





Ligand Structure Generation
Ligand size

- Ligands in series
  - Typically similar size – easy comparison by score (e.g. Vina dG_{bind})
- Ligands in library
  - Diverse sizes

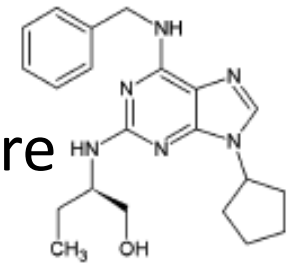
ZINCdb – propane (MW = 44 Da)

- TG(18:0/18:1(9Z) (MW = 1000 Da)

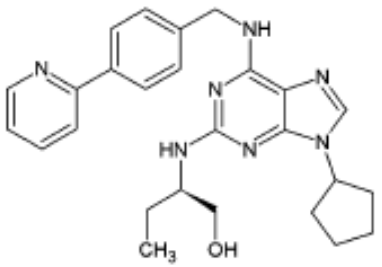
Larger ligand = more interactions = stronger binding (albeit arteficially - entropy)

- Size incorporating measures

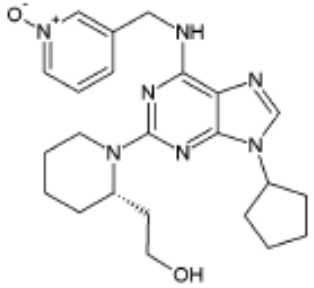
CDK2 inhibitors



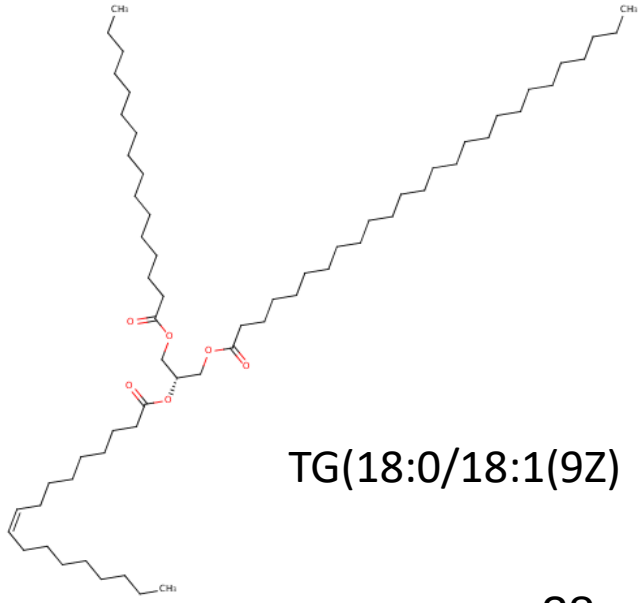
R-roscovitine



R-CR8



dinaciclib



TG(18:0/18:1(9Z)

http://zinc.docking.org/

Programs For Docking

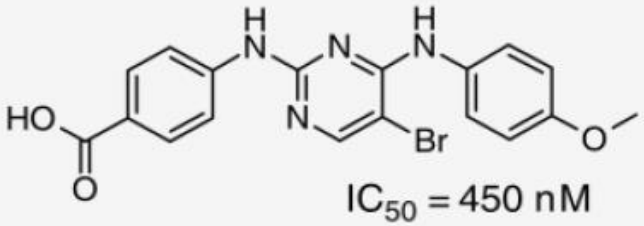
- **DOCK** (I. D. Kuntz, UCSF)
- **AUTODOCK** (Arthur Olson, The Scripps Research Institute)
- **Vina** (Arthur Olson, The Scripps Research Institute)
- RosettaDOCK (Baker, Washington Univ., Gray, Johns Hopkins Univ.)
- ArgusLabs
- **GOLD**
- **FlexX**
- Hex
- Glide (Schrodinger)

Virtual Screening

- Equivalent of biological screening (HTS – high throughput screening)
- Testing of thousands of compounds in silico
  - For further testing
  - For lead optimization
  - For leading organic synthesis

Virtual Screening Examples

Checkpoint kinase 1



$IC_{50} = 450$ nM

560 000 slouč.

Drug-like compound selection

199 000 slouč.

FlexX-Pharm docking

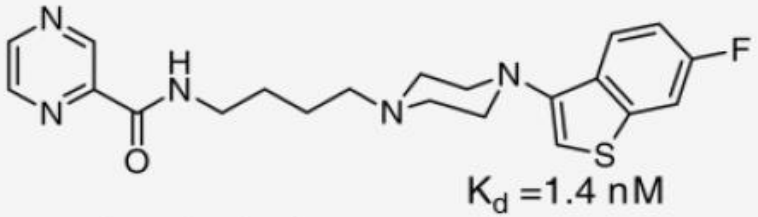
250 sloučenin

Hand picked

103 tested, 36 successful

J. Med. Chem. *47*, 1962 (2004).

α_{1A} adrenergic receptor



$K_d = 1.4$ nM

mnoho sloučenin

Drug-like compound selection

22 950 slouč.

GOLD docking

300 sloučenin

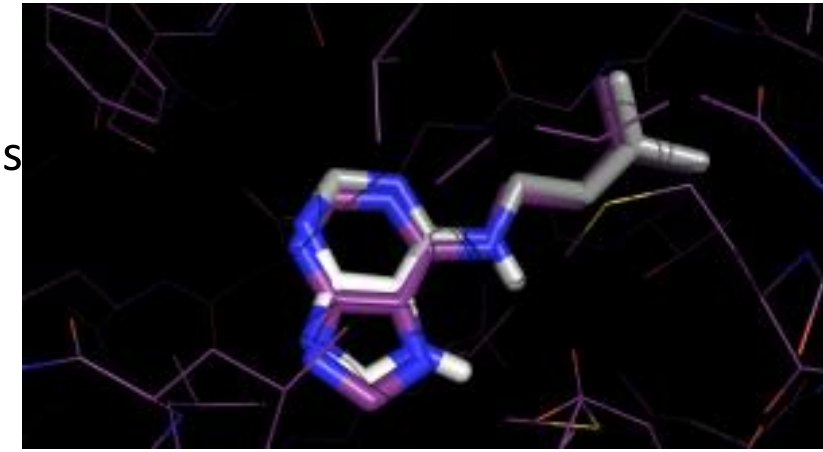
Statistical analysis

80 tested, 37 successful

J. Med. Chem. *48*, 1088 (2005).

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



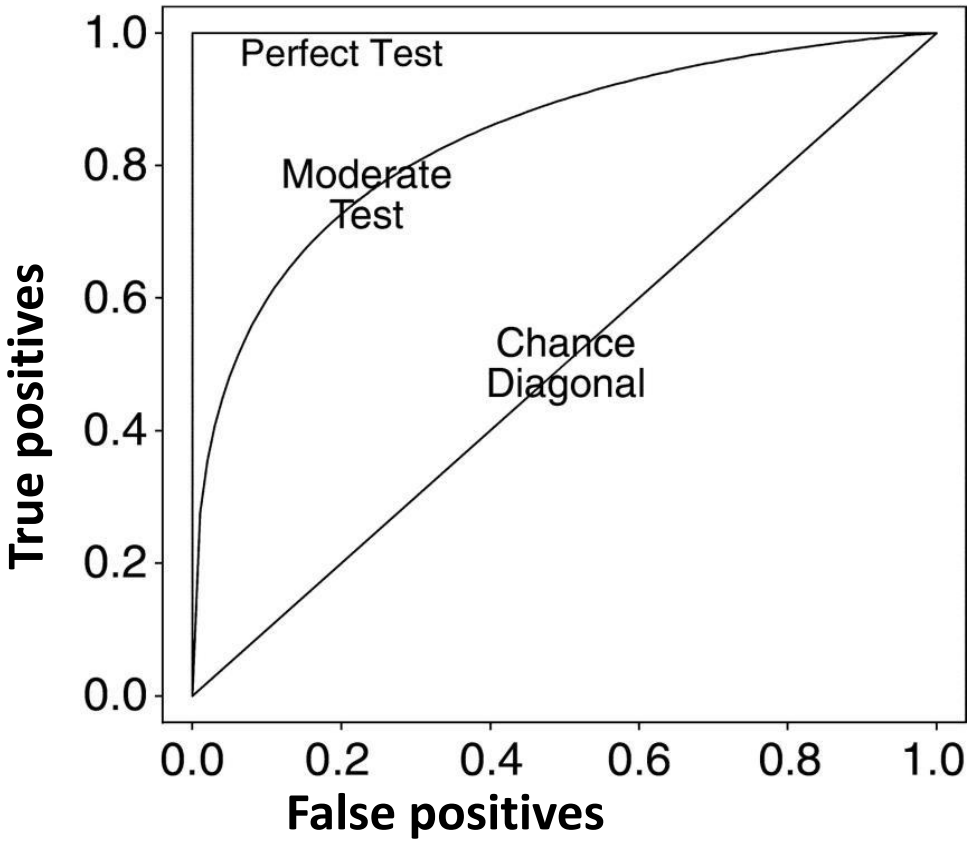
$$EF = \frac{a/n}{A/N}$$

- 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

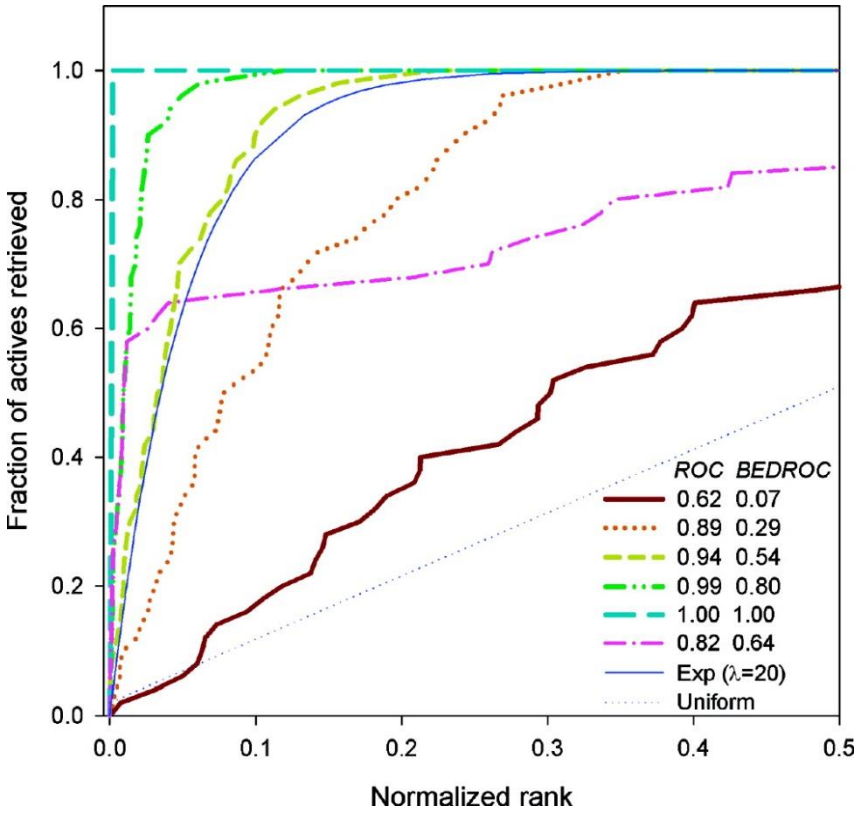


Figure 9 Different accumulation curves from sampling (n = 50, N = 25000) shown together with the corresponding *ROC* and *BEDROC* values where α = 20.0. An exact CDF with λ = 20 is also shown to highlight the fact that the *BEDROC* metric returns a value of 1/2 for a curve close to this CDF.

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

Docking Summary

- 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

Tutorial preparation

Programs installation:

- Python 2.7

http://www.python.org/download/releases/

- Pymol

http://www.lfd.uci.edu/~gohlke/pythonlibs/#pymol

- Autodock Tools

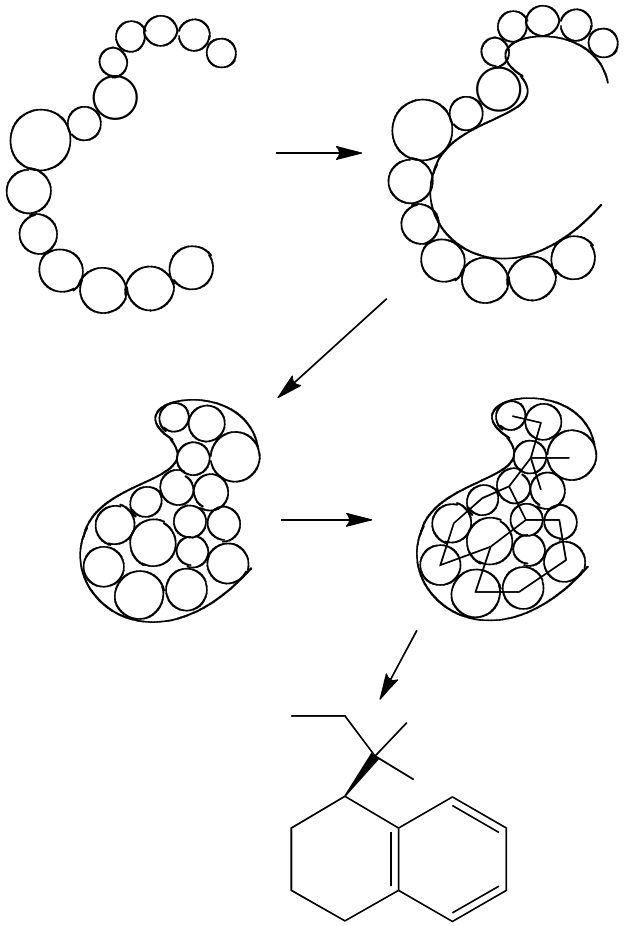
http://mgltools.scripps.edu/

- Autodock Vina

http://vina.scripps.edu/

de novo docking

1) Surface of receptor (SASA - Connelly)
2) “negative” of surface fill with spheres
3) Distances between spheres
4) Sphere distance to Bond distance conversion
5) Search in small molecule database
6) Selection of ligands with largest overlap with spheres
7) Scoring



Groupbuild

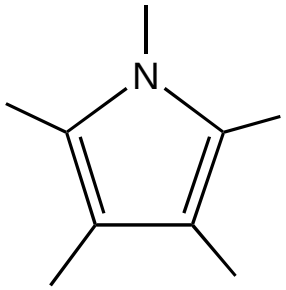
- Building of new compounds by filling of active site by random fragments

Algorithm:

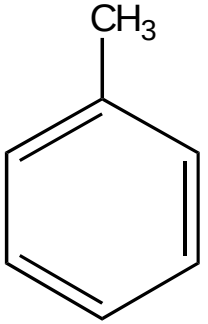
1) Grid for receptor binding site
2) Structure generation
    1) Docking of "core" fragment
    2) Build up (random fragments additions to core)
    3) Selection of best structures
    4) Iterate steps 2 and 3 until final criteria fulfilled (number of steps, minimal energy, etc.)
3) Selection of final structures for synthesis

FlexX, AutoGrow

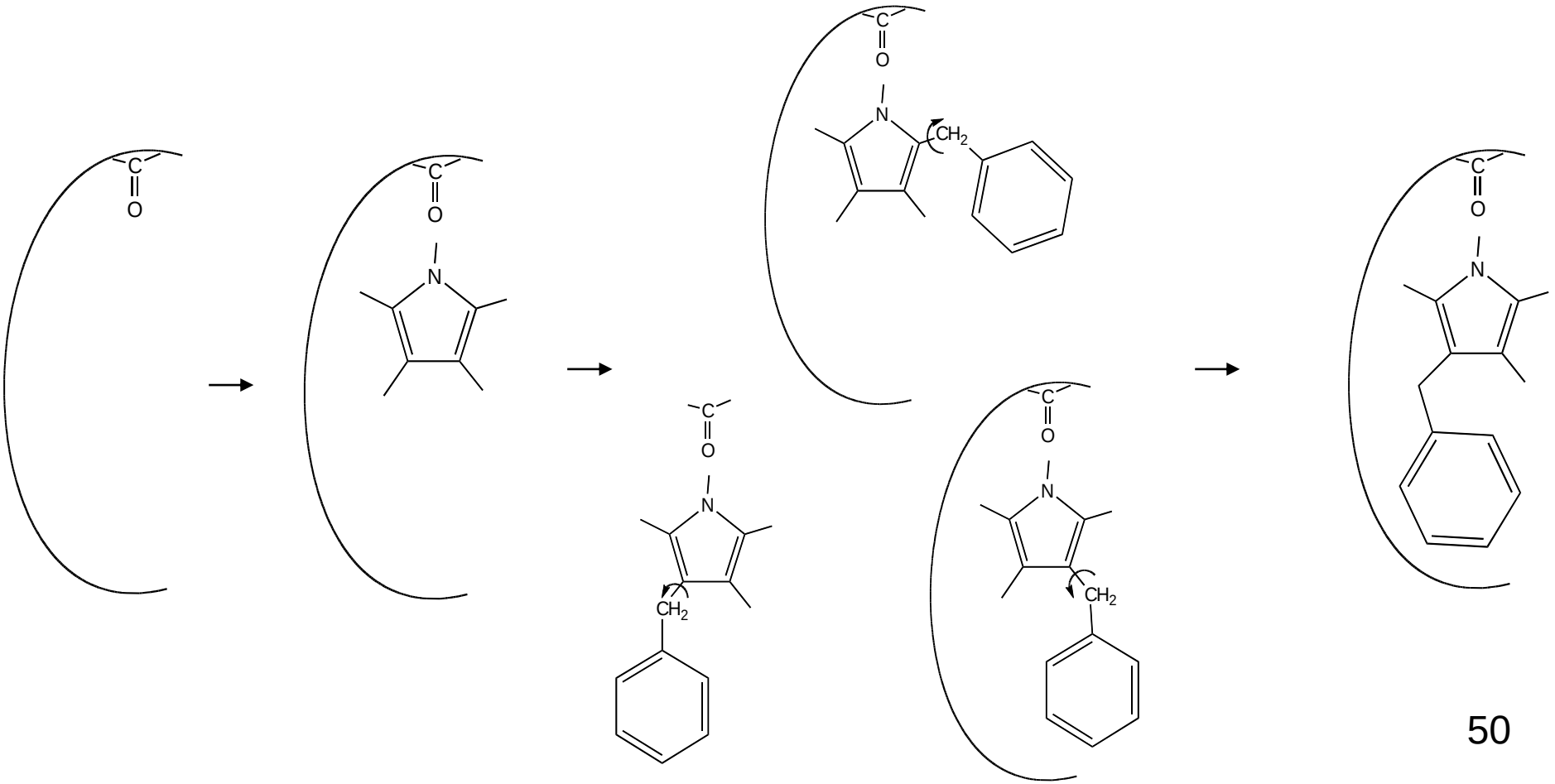
Groupbuild



initial core fragment

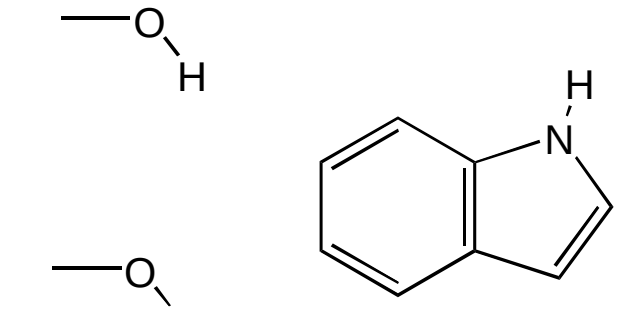


fragment to be added

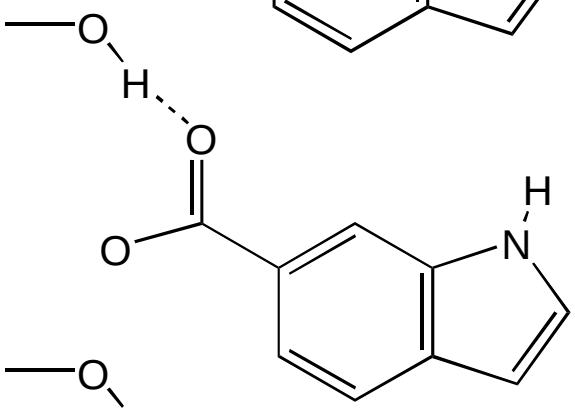


Groupbuild

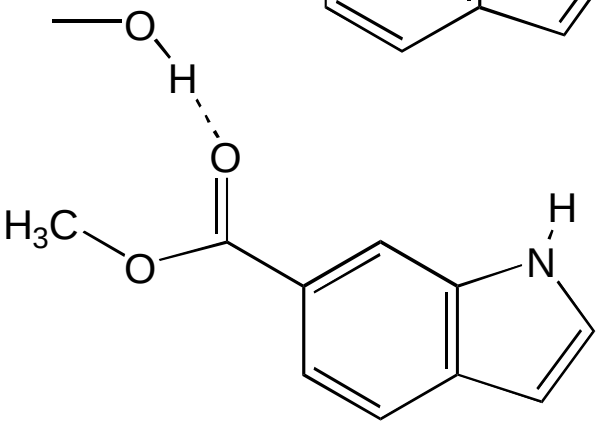
Example for Hypothetical Receptor



$G_{receptor}$	$G_{solution}$
-8	-10



$G_{receptor}$	$G_{solution}$
-30	-80



$G_{receptor}$	$G_{solution}$
-30	-20

$$\Delta\Delta G = \Delta G_{eq,2} - \Delta G_{eq,1} = \Delta G_{receptor} - \Delta G_{solution}$$