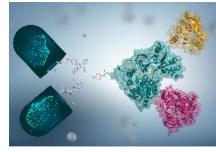


Docking and Scoring

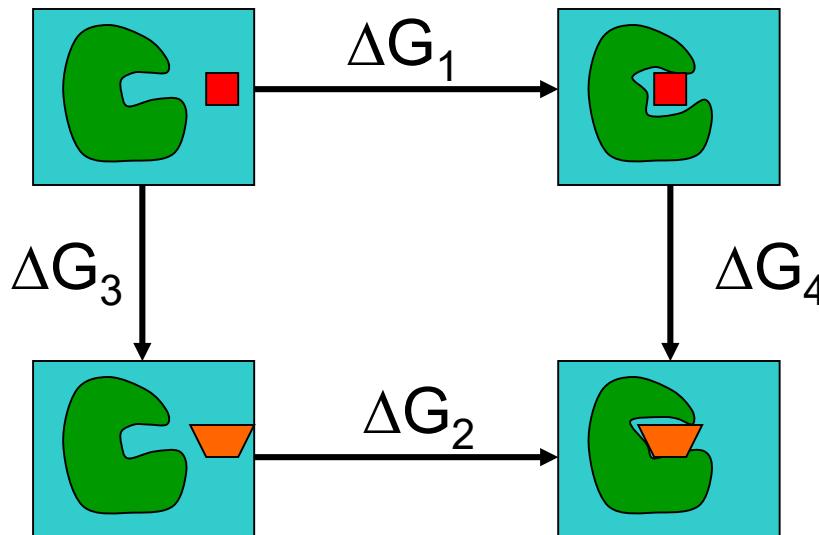


Introduction



Docking & Scoring

Calculation of binding affinities
using simulation techniques
(MD/MC):

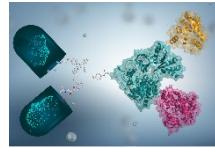


Calculate full thermodynamic
cycle (FEP) or endpoints only
(MM/PBSA, LIE, SIE)

BUT: Sampling very time consuming

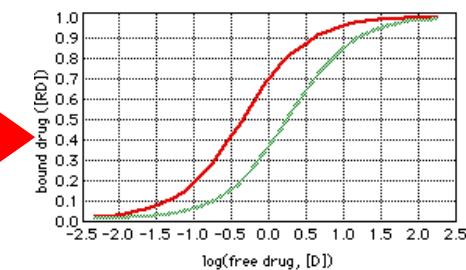
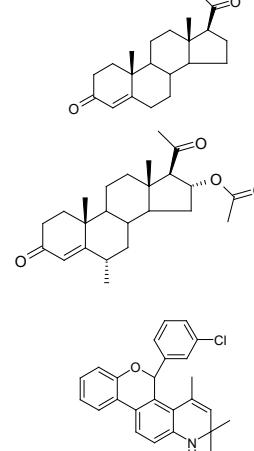
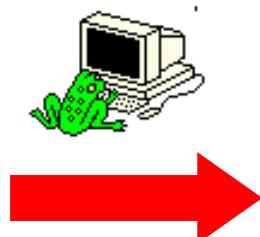
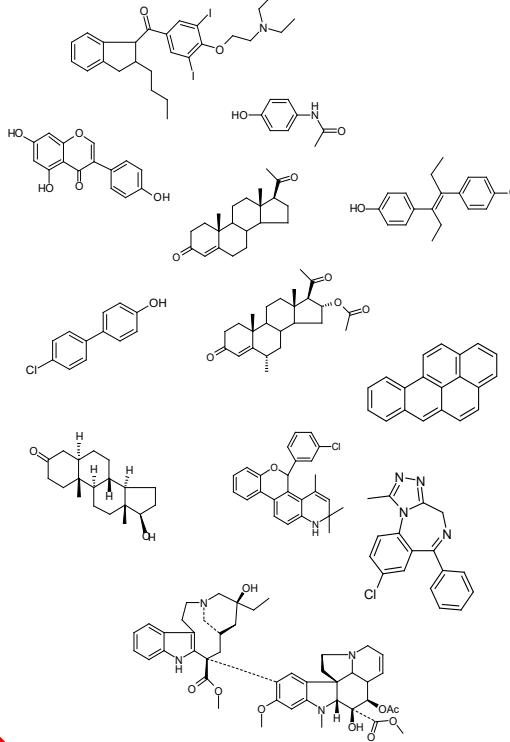
Limitations:

- For FEP: only structurally similar compounds
- **What is the overall binding mode?** Standard MD simulations only locally sample ligand-protein conformations.
- Very time consuming (1+ days per compound)



Docking & Scoring

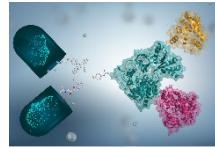
Aim: Predict binding mode and affinities in reasonable amount of time



Library of options
100'000 – 1'000'000
possible compounds

Synthesis
100 – 1000
compounds

Biological testing
100 – 1000
compounds



Docking & Scoring

For a given protein:

Which ligand is able to bind?

How does the ligand bind?

→ Orientation, position, conformation

How strong does the ligand bind?

→ Binding affinity

Docking & Scoring

FEP, MM/PBSA, LIE, SIE

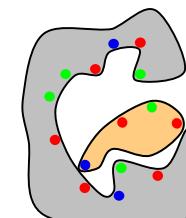
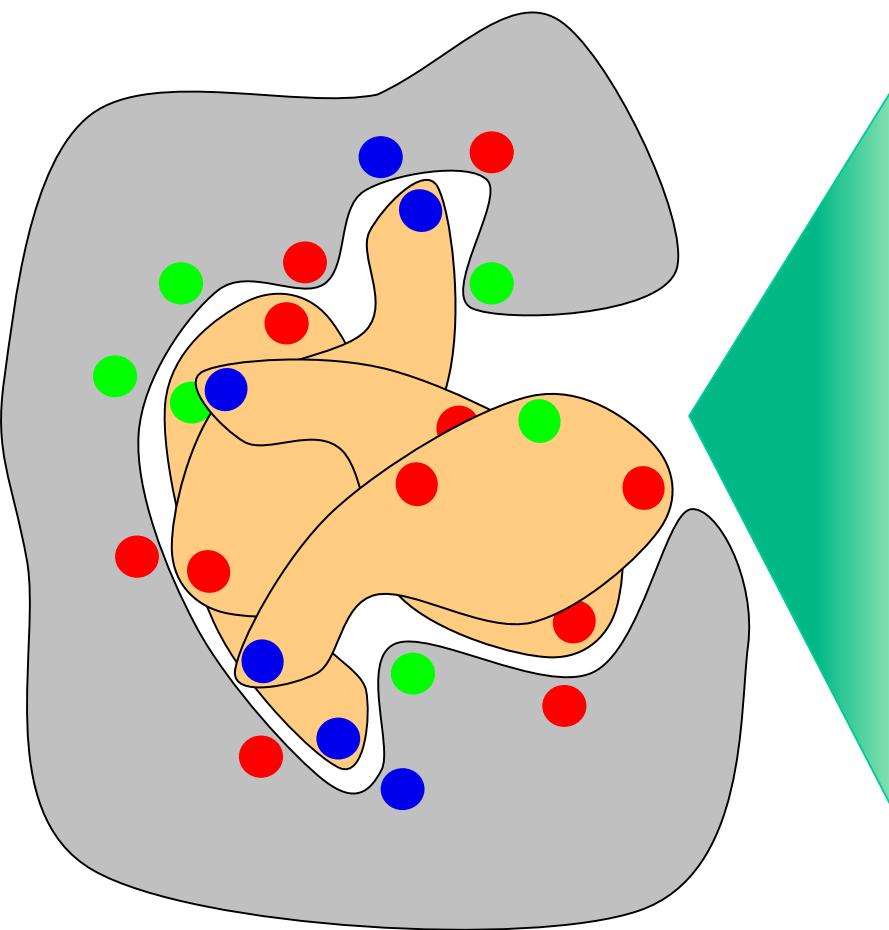


General procedure

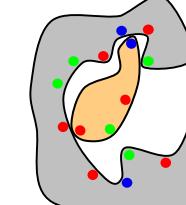
Generate a large set of possible ligand-protein configurations using an efficient **search algorithm** (posing)



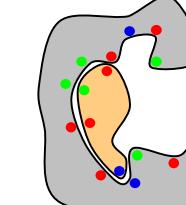
Compute protein-ligand interaction using an efficient **scoring function**



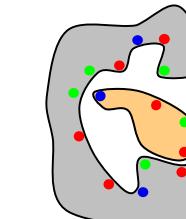
Score = -6.2



Score = -8.4

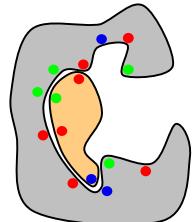


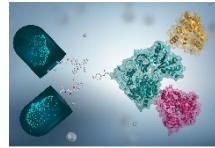
Score = -9.3



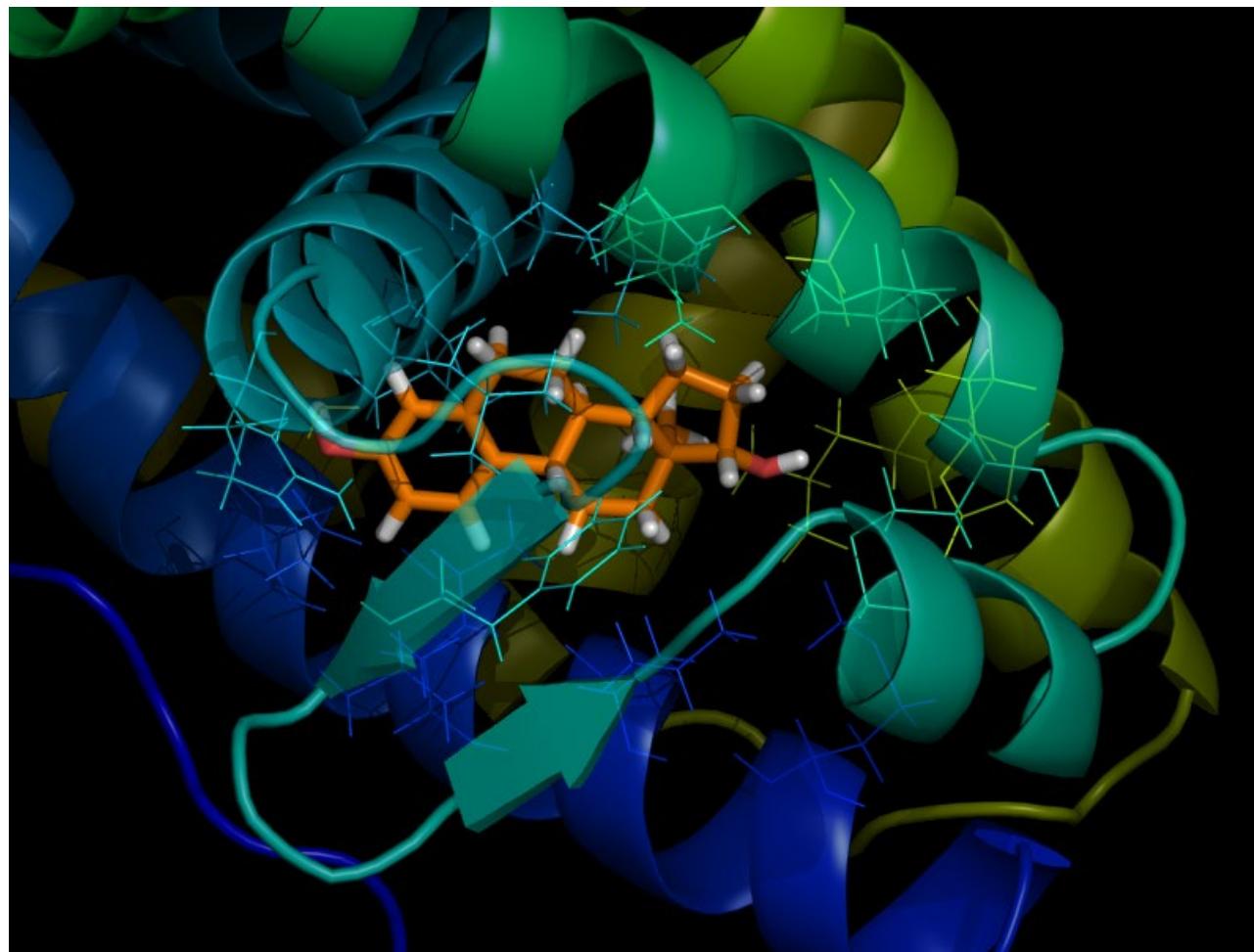
Score = -7.2

Predicted pose

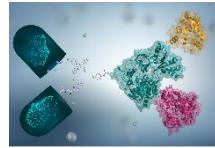




Docking & Scoring



Search algorithms



Search algorithms

Characteristics:

- Extensive sampling of different poses
- Limiting duplicate visits of similar/identical poses



Search algorithm

Degrees of flexibility:

rigid ligand &
rigid protein

rigid ligand, but multiple pre-generated conformations (library) & rigid protein

||

flexible ligand &
rigid protein

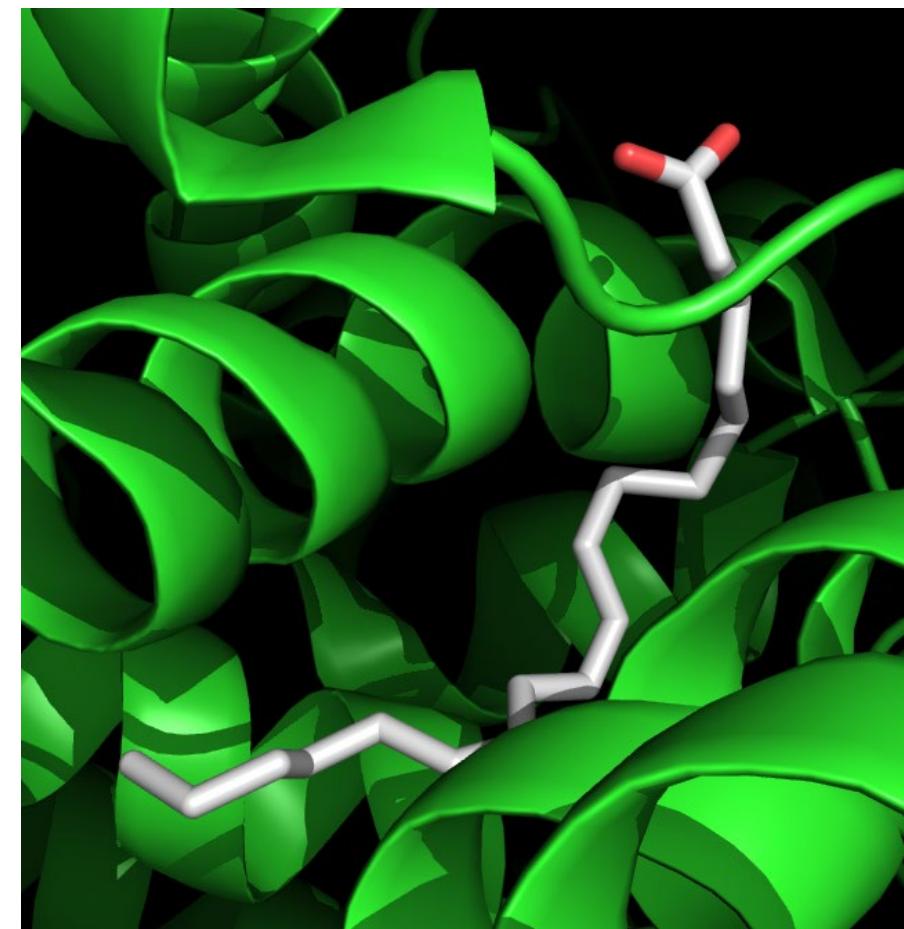
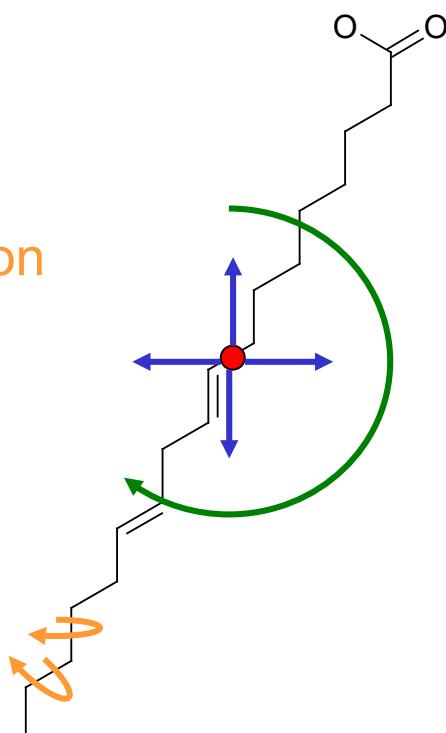
flexible ligand &
flexible protein



Search algorithm

Flexible ligand: Degrees of freedom

- Translation
- Rotation
- Torsion rotation



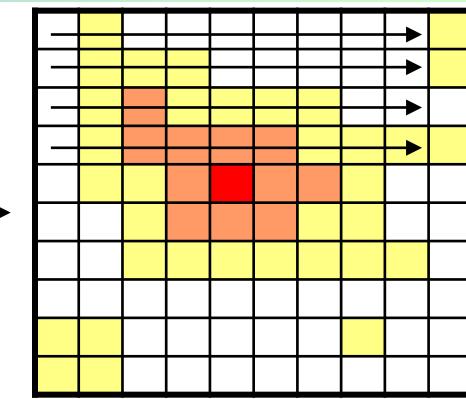
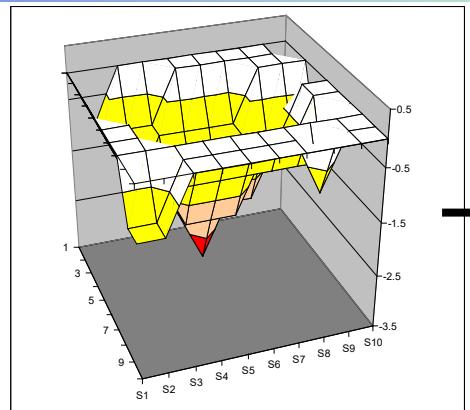


Search algorithm

General categories:

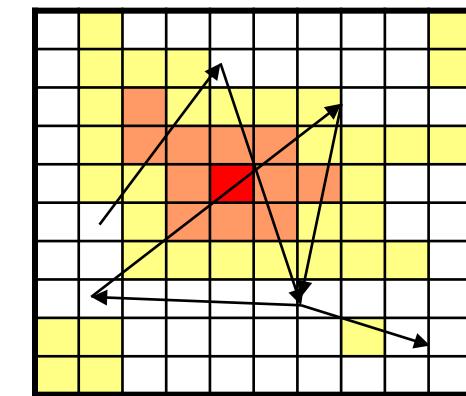
- **Systematic search:**

Each degree of freedom is systematically explored in intervals in a combinatorial fashion



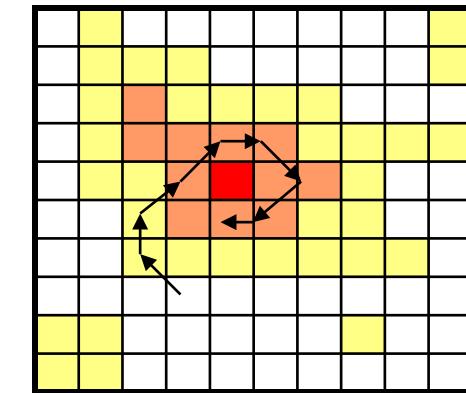
- **Stochastic search:**

Conformations are changed by random steps under certain rules



- **Deterministic search (via simulation methods):**

New conformation is generated from previous conformation following certain rules (e.g. forces) without randomness





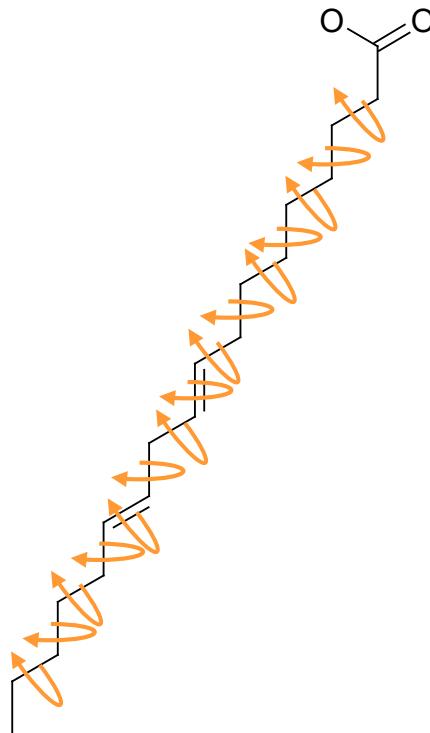
Systematic search algorithm

Example: Conformational search method

Systematic scan for all degrees of freedom

Fine scan → combinatorial explosion

Coarse scan → correct binding mode is possibly not identified



15 rotatable bonds (torsions):

Scan each 10° → $36^{15} = 2.2 \cdot 10^{23}$ conformations

Scan each 30° → $12^{15} = 1.5 \cdot 10^{16}$ conformations

Scan each 120° → $3^{15} = 1.4 \cdot 10^7$ conformations

Usually used to pre-generate conformational library (e.g. for FLOG, SLIDE):
Fast, because only internal energy has to be calculated

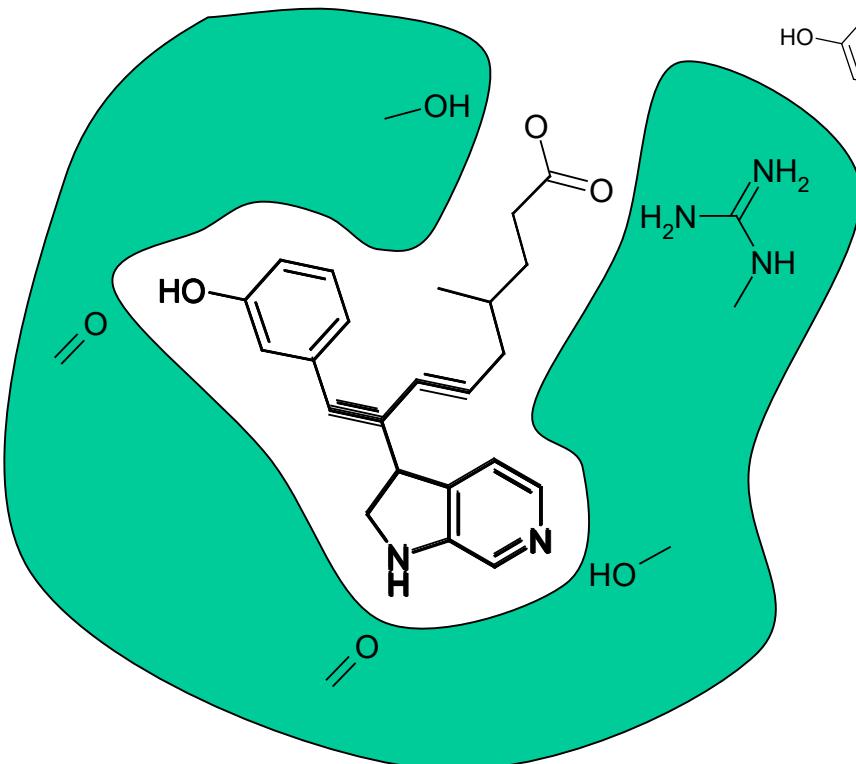


Systematic search algorithm

Example: **Fragmentation methods** (e.g DOCK 4.0, FlexX)

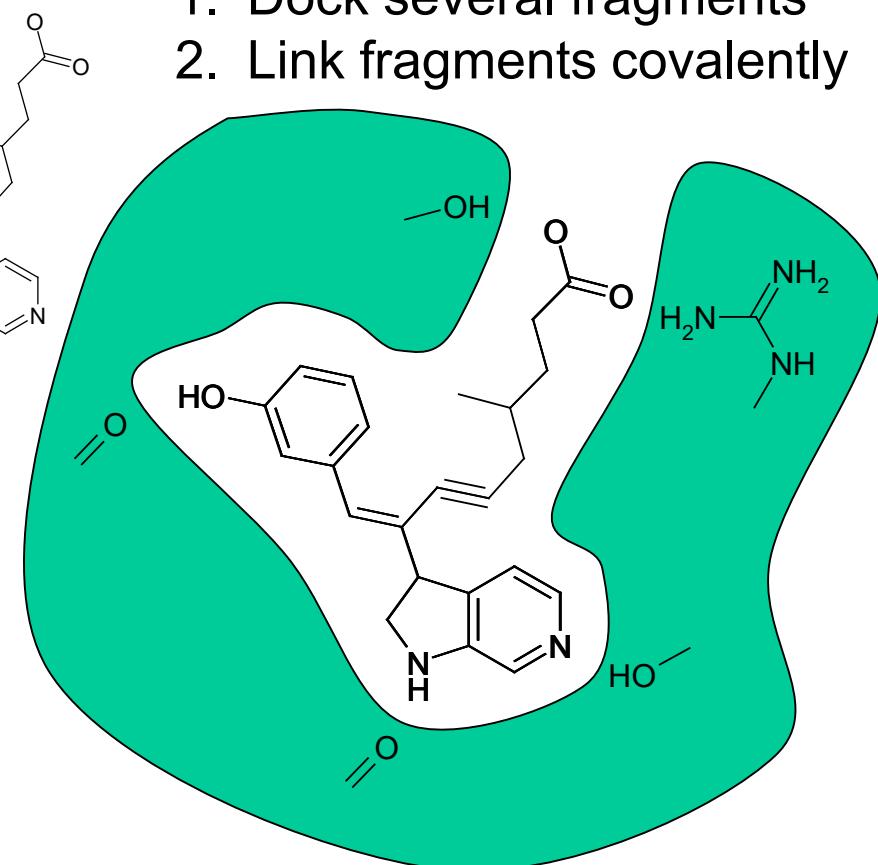
Incremental approach:

1. Dock core fragment
2. Add flexible regions incrementally

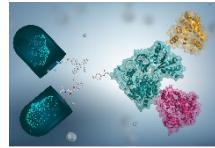


Place-and-join approach:

1. Dock several fragments
2. Link fragments covalently



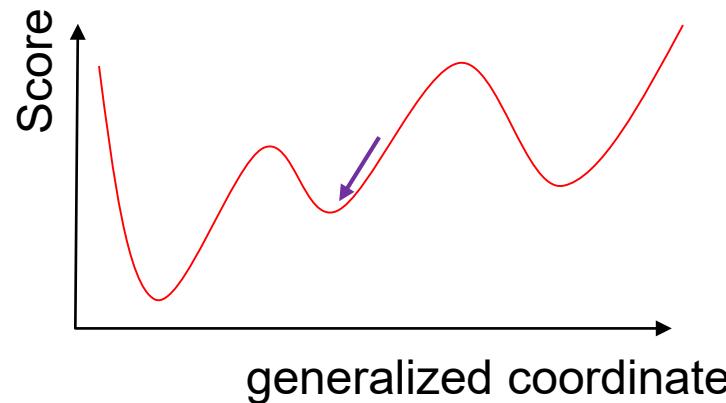
Assumption: Core fragment position in the intact ligand is among n lowest docked poses \rightarrow repeat process with n docked poses for each growing fragment



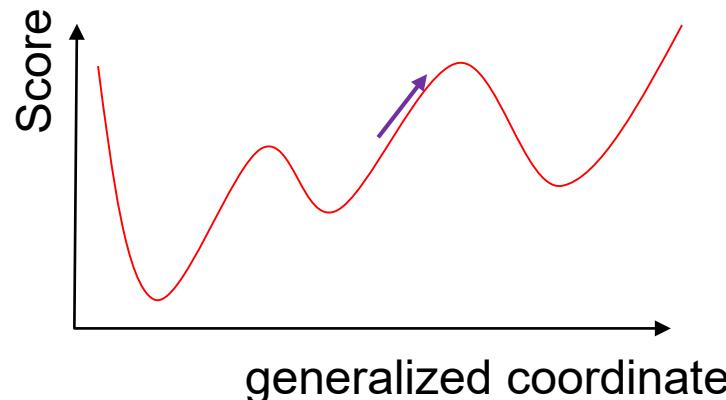
Stochastic search algorithm

Example: **Monte Carlo algorithms** (e.g. ICM)

1. Random changes in translational, rotational and torsional degrees of freedom
2. Usually: Minimization
3. Accept new conformation based on Metropolis criteria



⇒ always accept



⇒ accept with probability

$$p = \exp\left(-\frac{\Delta Score}{k_B T}\right)$$

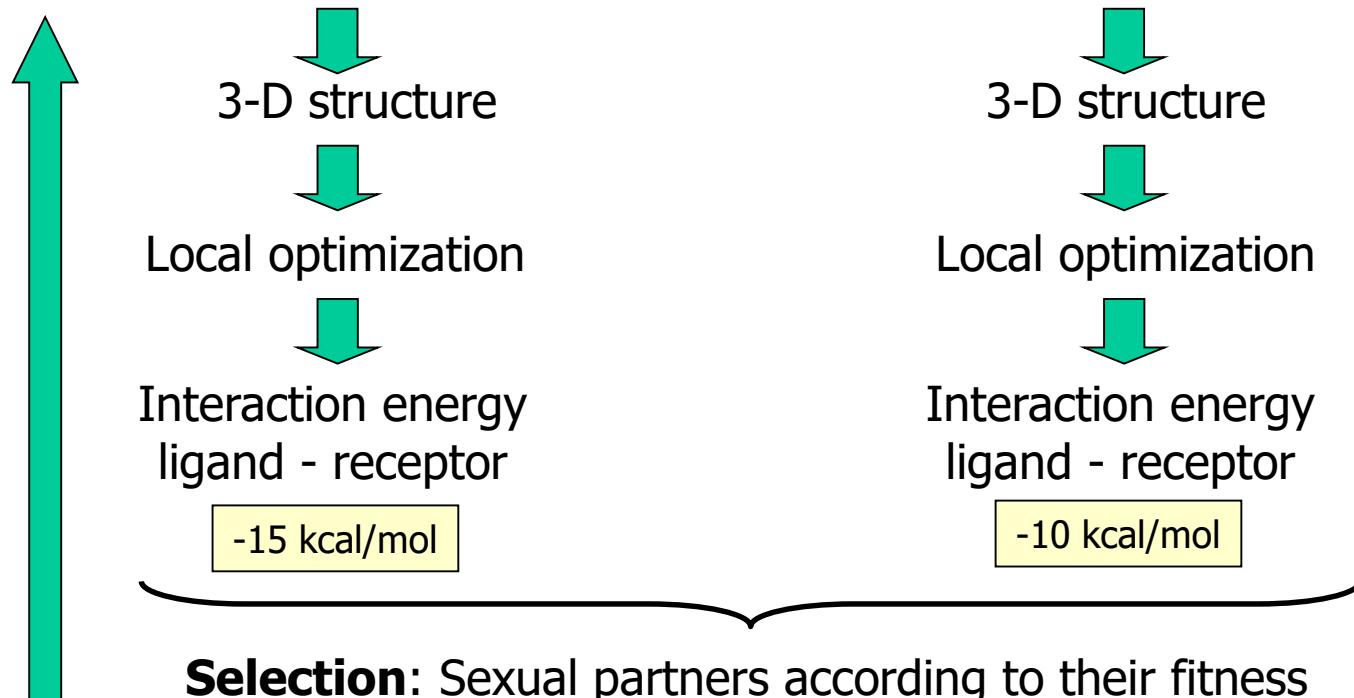


Stochastic search algorithm

Example: (Lamarckian) Genetic algorithm methods (e.g. AutoDock, GOLD)

Translation	Rotation	Torsions	Translation	Rotation	Torsions	...												
0.7	0.2	-0.4	-0.2	0.5	-1.1	0.3	1.3	0.0	0.1	0.8	0.3	-0.1	0.1	-0.4	-0.3	0.3	0.2	...

n models
(individuals)



0.7	0.2	-0.4	-0.2	0.2	-1.1	0.3	1.3	0.0	0.1	0.8	0.3	-0.1	0.1	-0.4	-0.3	0.3	0.2
-----	-----	------	------	-----	------	-----	-----	-----	-----	-----	-----	------	-----	------	------	-----	-----

Mutation

Crossover



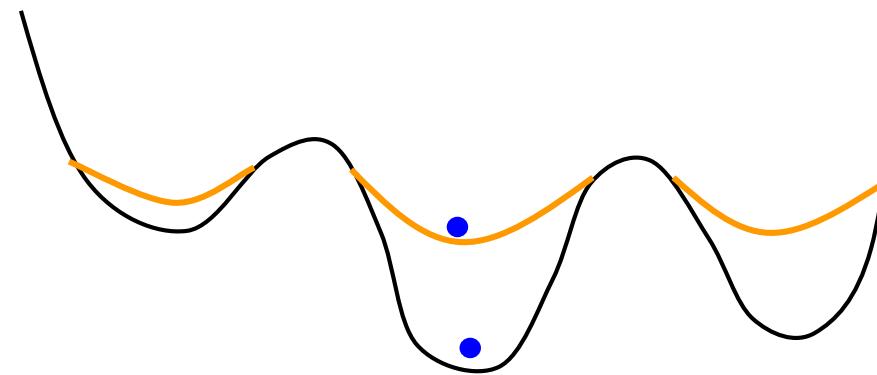
Deterministic search methods

Example: Molecular dynamics simulations

Problem: MD simulations can not overcome large energy barriers → only local sampling

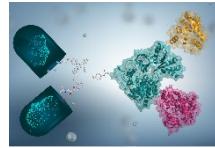
Possible Solution: MD at high temperature or reduce barriers

Accelerated MD



Hamelberg, D.; Mongan, J.; McCammon, J.A. *J. Chem. Phys.* **2004**, *120*, 11919-11929.

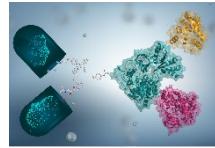
Scoring functions



Scoring functions

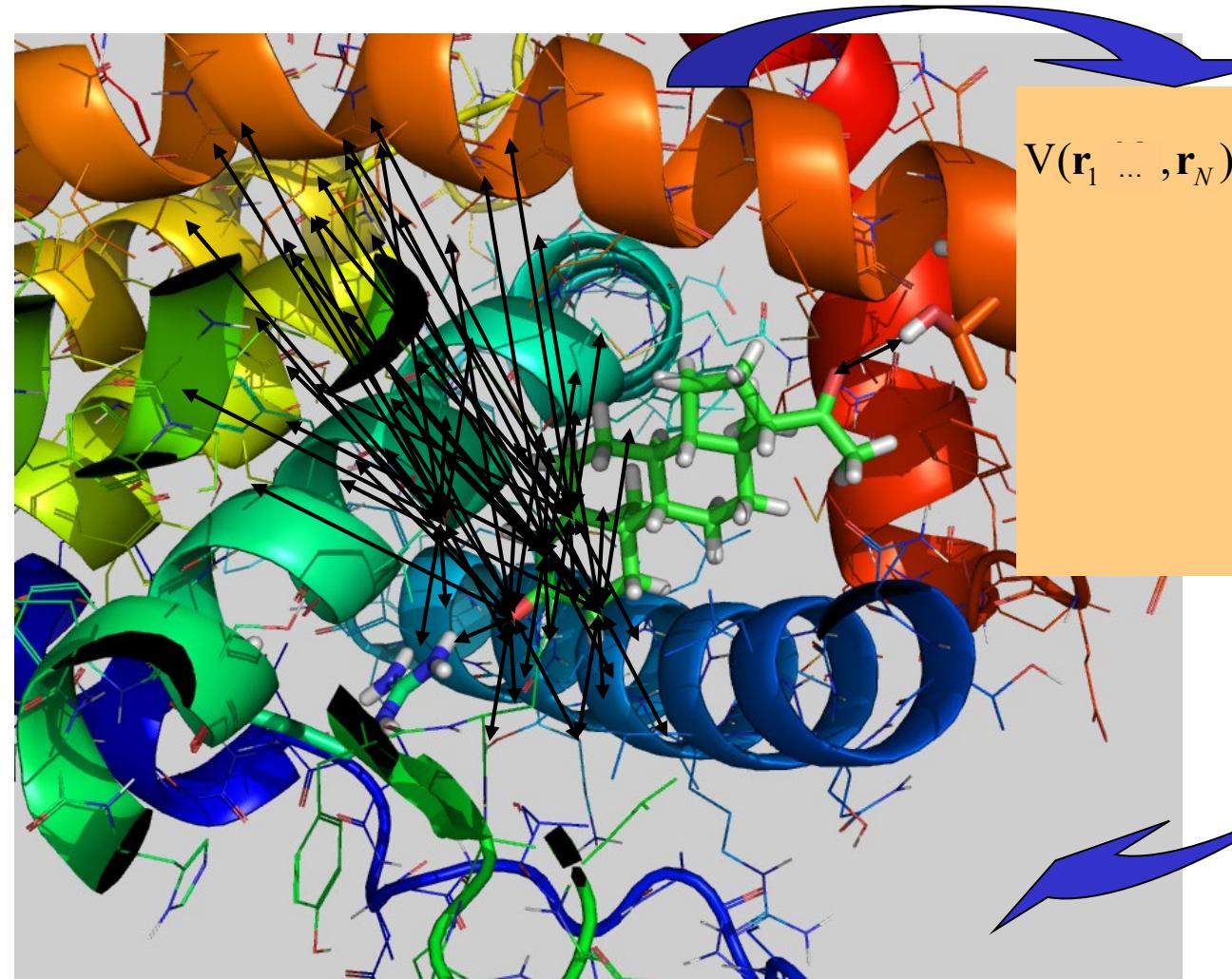
Characteristics:

- **Accurate** inclusion of dominant protein-ligand interaction terms to assign the most favorable scores to the experimentally determined complex
 - Steric complimentarity
 - Physico-chemical complimentarity
 - **Fast** to allow scoring of many different poses
- ⇒ Compromises necessary

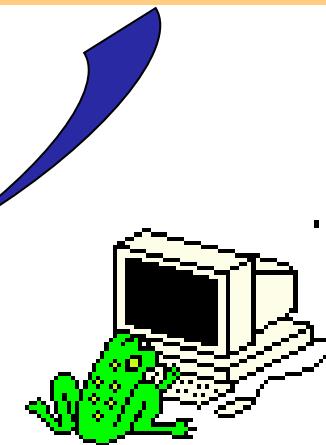


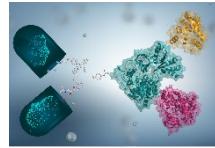
Scoring functions

Number of non-bonded interactions e.g. for N=10000 atoms: $N(N-1)/2 \sim 5 \cdot 10^7$



$$\begin{aligned} V(\mathbf{r}_1, \dots, \mathbf{r}_N) = & \sum_{i < j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_{i < j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \\ & + \sum_{bonds} \frac{1}{2} k_{ij}^b (r_{ij} - r_{ij}^0)^2 \\ & + \sum_{angles} \frac{1}{2} k_{ijk}^a (\theta_{ijk} - \theta_{ijk}^0)^2 \\ & + \sum_{dihedrals} \frac{1}{2} k_{ijkl}^d (1 + \cos(n(\phi_{ijkl} - \phi_{ijkl}^0))) \end{aligned}$$





Scoring function

A scoring function ranks the different ligand-protein configurations.

Aims:

1. Calculate binding affinity.

2. Enable distinction between true binding mode and all the other alternative modes explored during configurational search.
3. Efficient calculation of ligand-protein interactions to allow for extensive configurational search.

General categories:

- **Force-field based scoring functions** (e.g. AutoDock 3.0, GoldScore)

$$\Delta G = \Delta G_{vdW} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \Delta G_{HBond} \sum_{i,j} E(\theta) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + \Delta G_{el.st.} \sum_{i,j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}}$$
$$+ \Delta G_{Tors} N_{Tors} + \Delta G_{Solv} \sum_{i,j} (S_i V_j + S_j V_i) \exp \left(-\frac{r_{ij}^2}{2\sigma^2} \right)$$

Stored on grid → Efficiency

Issue: Steep potential function ↔ protein flexibility



Scoring function

- **Empirical scoring functions** (e.g. LUDI, ChemScore)

also called: **Regression-based scoring functions:**

Optimize pre-factors (ΔG_0 , ΔG_{HBond} , etc.) e.g. with multi-linear regression on dataset of experimental binding data (i.e. binding modes, affinity)

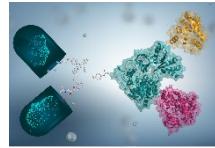
$$\Delta G = \Delta G_0 + \Delta G_{\text{HBond}} \sum_{\text{HBonds}} f(\Delta r, \Delta \theta) + \Delta G_{\text{ionic}} \sum_{\text{ionic contacts}} f(\Delta r, \Delta \theta) \\ + \Delta G_{\text{hydrophob}} |A_{\text{hydrophob}}| + \Delta G_{\text{Tors}} N_{\text{Tors}}$$

Typical difference between empirical scoring functions:

pre-factors, explicit aromatic term, function of hydrophobic term

Issues:

- Scoring function (i.e. pre-factors) only as good as the experimental data set on which it was optimized (experimental uncertainty in ΔG !!!)
- Is data set representative?
- Additivity of ΔG ?



Scoring function

- **Knowledge-based scoring functions** (e.g. PMF, DrugScore)

Idea:

- Represent ligand and protein by individual atom-types
- Frequency of occurrence of individual contacts (atom-type j ligand – atom-type i protein) is a measure of their energetic contribution to binding

$$\Delta G = \sum_{\text{pairs}} E_{i,j}(r)$$

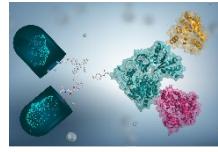
$$\text{with } E_{i,j}(r) = -k_B T \ln [\rho_{i,j}(r)/\rho(r)]$$

$\rho_{i,j}(r)$: normalized probability of atom pair i,j to be in contact at distance r

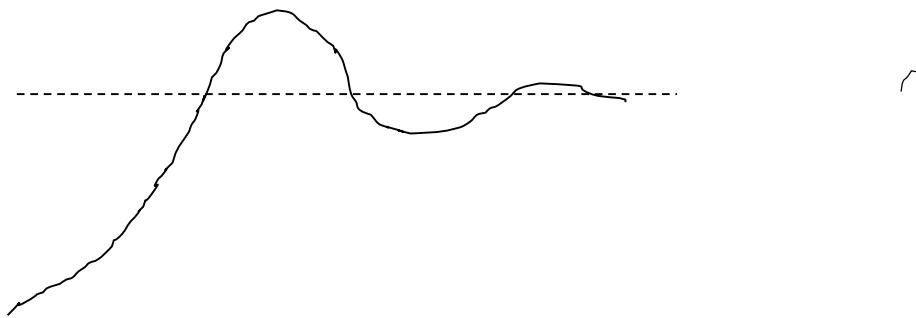
$\rho(r)$: normalized reference probability

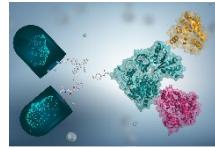
Issues:

- Is data set representative?
- Directionality of hydrogen bonds?



Scoring function





Docking & Scoring

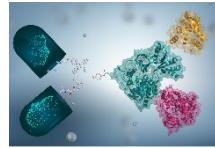
References

N. Brooijmans, I. D. Kuntz (2003) Molecular Recognition And Docking Algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **32**, 335–73.

S. F. Sousa, P. A. Fernandes, M. J. Ramos (2006) Protein–Ligand Docking: Current Status and Future Challenges. *Proteins* **65**, 15–26.

I. Muegge, M. Rarey (2001) Small Molecule Docking and Scoring. *Rev. Comput. Chem.* **17**, 1-39.

Applications



Applications/Validation

Types of applications:

- Prediction of binding modes

Generate large set of binding poses (ligand orientation and conformation) and use scoring function to select most likely pose (which should be similar to that seen crystallographically) → base pose for lead optimization.

Speed: - Accuracy: +

- Lead identification (Virtual screening)

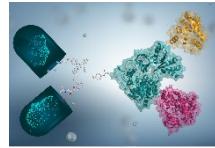
Dock large database of compounds to target protein and rank compounds
→ The early part of list of top ranked compounds should be enriched with active compounds.

Speed: + Accuracy: - (often different settings as in binding mode prediction)

- Lead optimization (Prediction of binding affinity)

Score accurate binding mode to predict relative potencies among compounds.

Speed: -- Accuracy: ++

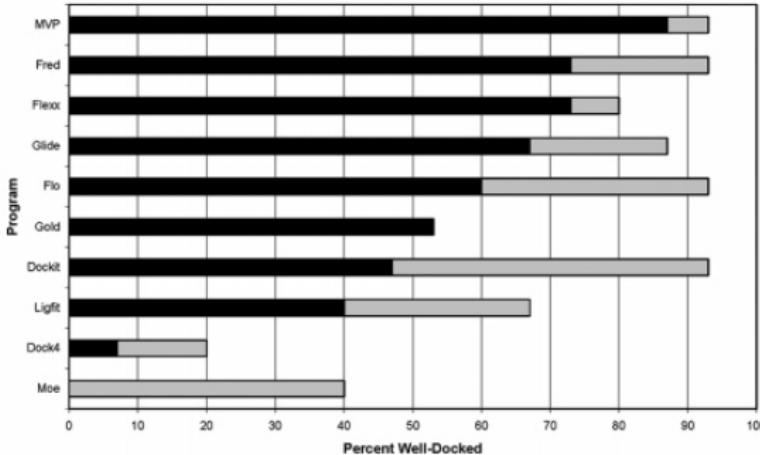


Prediction of binding modes

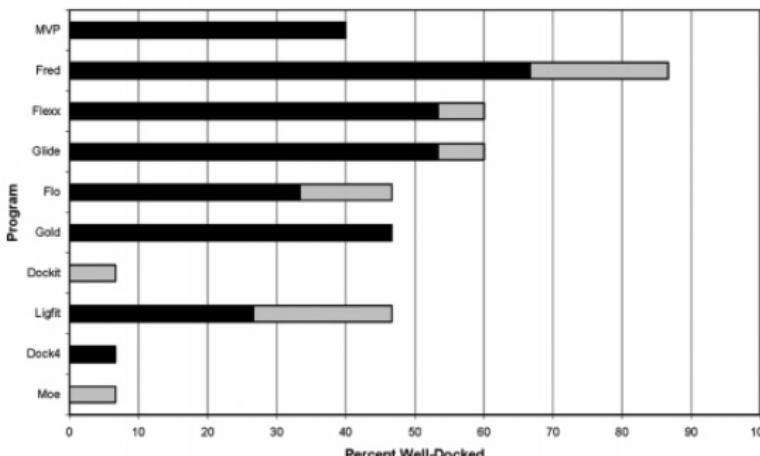
Percentage of successfully docked ligands (compared to X-ray):

G.L.Warren *J. Med. Chem.* 2006, 49, 5912-5931

Any Pose, Chk1 Kinase



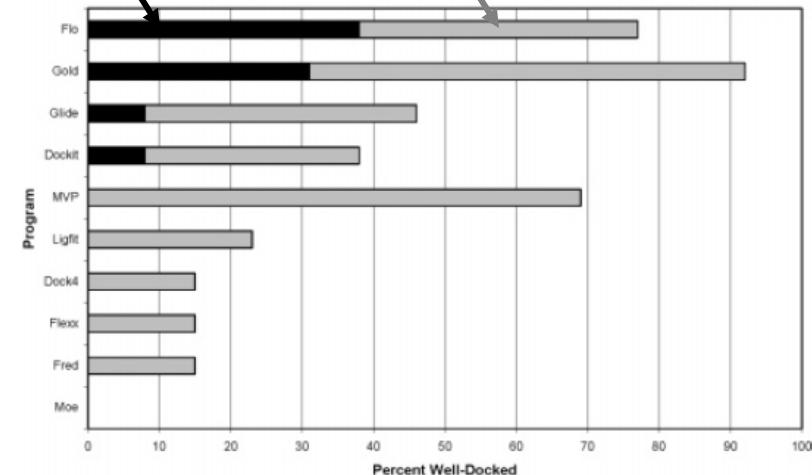
First Pose, Chk1 Kinase



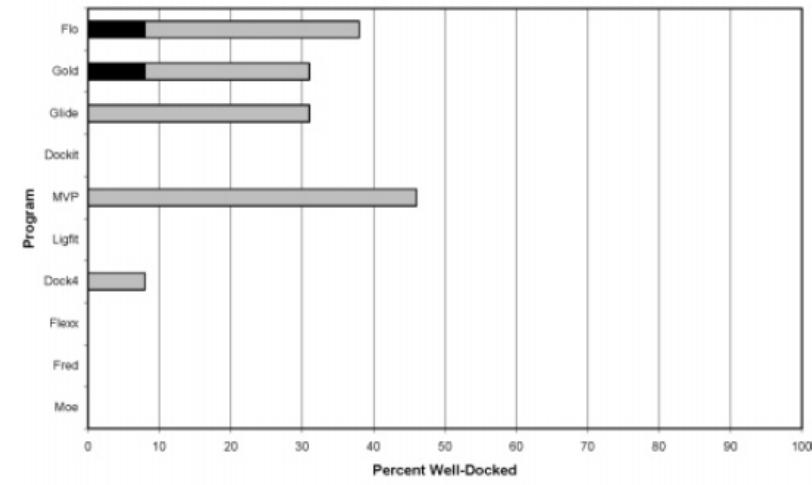
RMSD < 2.0 Å

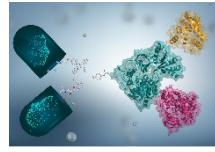
RMSD < 4.0 Å

Any Pose, HCV Polymerase



First Pose, HCV Polymerase





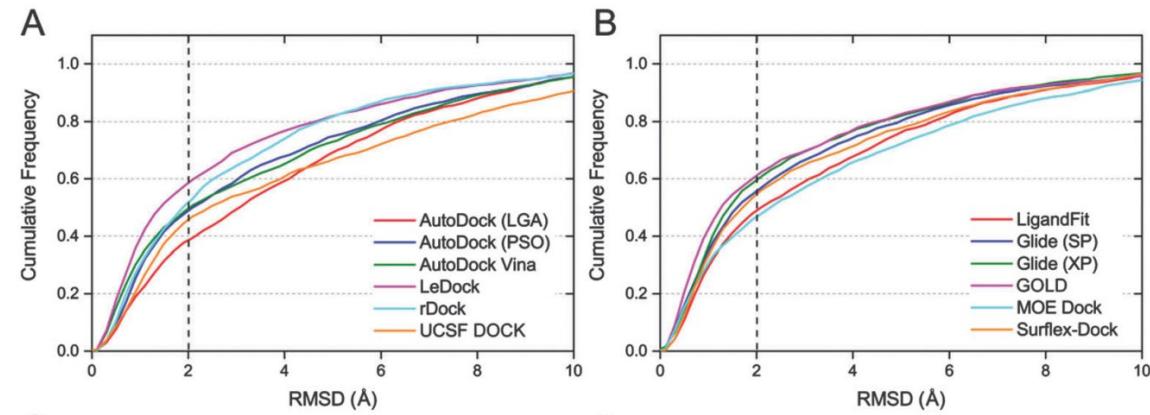
Prediction of binding modes

Success rates (7 protein systems): RMSD < 2.0 Å

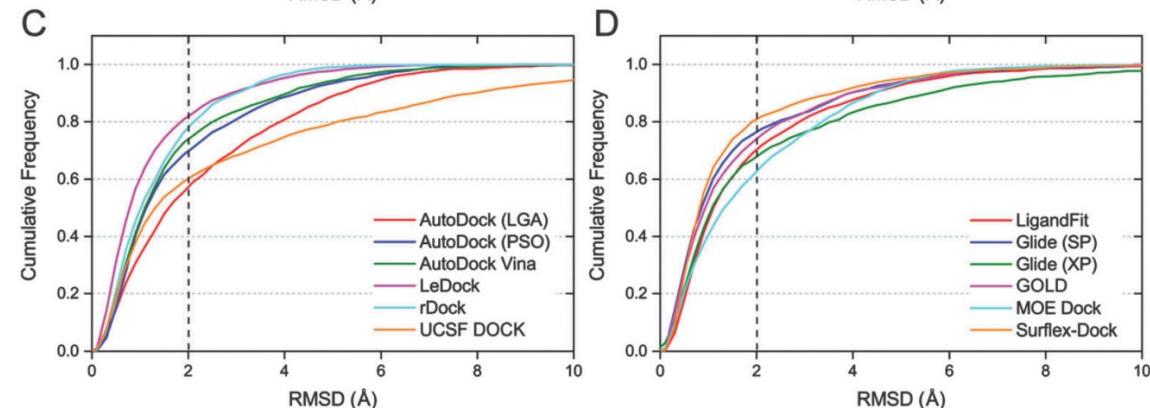
	Any pose	Top pose
Best docking/scoring method:	38 - 94%	8 - 74%
e.g. GLIDE	8 - 74%	0 - 55%

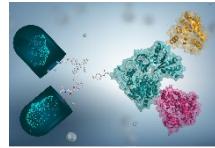
New Study: Phys. Chem. Chem. Phys., 2016, 18, 12964-12975
(>2000 systems)

Top pose:



Best/Any pose:



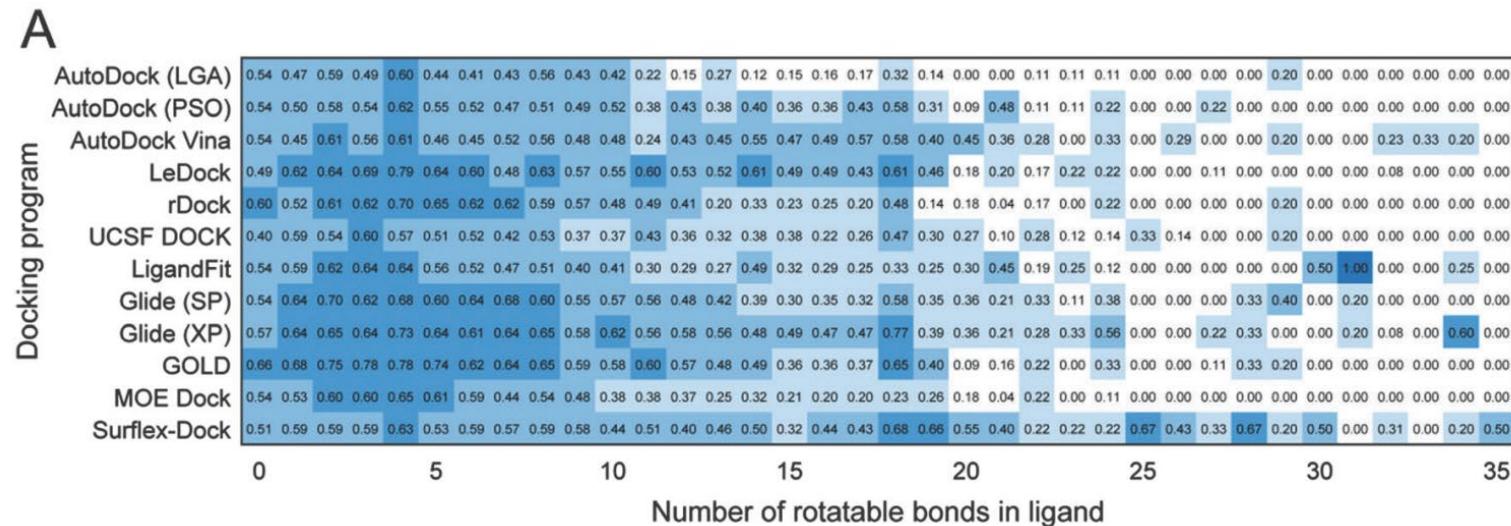


Prediction of binding modes



In general:

- Docking methods can generate poses close to X-ray data for about 60-80% of all protein-ligand systems.
 - However, RMSD for best scored pose shows significant drop in success rates (40-60%). Scoring functions are not always successful in separating X-ray conformation from decoy conformations.
 - Best docking/scoring method depends strongly on protein system
 - Success rate drops with
 - increasing ligand size/flexibility (in particular >10 rotatable bonds)





Prediction of binding modes

→ Peptides (high flexibility) are difficult to dock:

Table 2 Success rates of docking for regular organic molecule ligands and peptides or peptide mimic ligands. A docking pose is considered successful if the RMSD between the docking pose and the experimentally determined conformation of a ligand is less than 2.0 Å

Docking program	Regular organic molecule		Peptide or peptide mimic	
	Top scored pose	Best pose	Top scored pose	Best pose
AutoDock (LGA)	0.378	0.559	0.216	0.324
AutoDock (PSO)	0.477	0.686	0.331	0.439
AutoDock Vina	0.485	0.726	0.384	0.597
LeDock	0.574	0.808	0.352	0.465
rDock	0.503	0.763	0.283	0.465
UCSF DOCK	0.445	0.591	0.340	0.415
LigandFit	0.479	0.689	0.267	0.504
Glide (SP)	0.544	0.754	0.403	0.547
Glide (XP)	0.584	0.666	0.403	0.484
GOLD	0.599	0.726	0.371	0.472
MOE Dock	0.457	0.612	0.195	0.245
Surflex-Dock	0.533	0.800	0.440	0.673

- increasing size of binding pocket (ratio binding pocket : ligand size)
- decreasing number of specific non-hydrophobic contacts

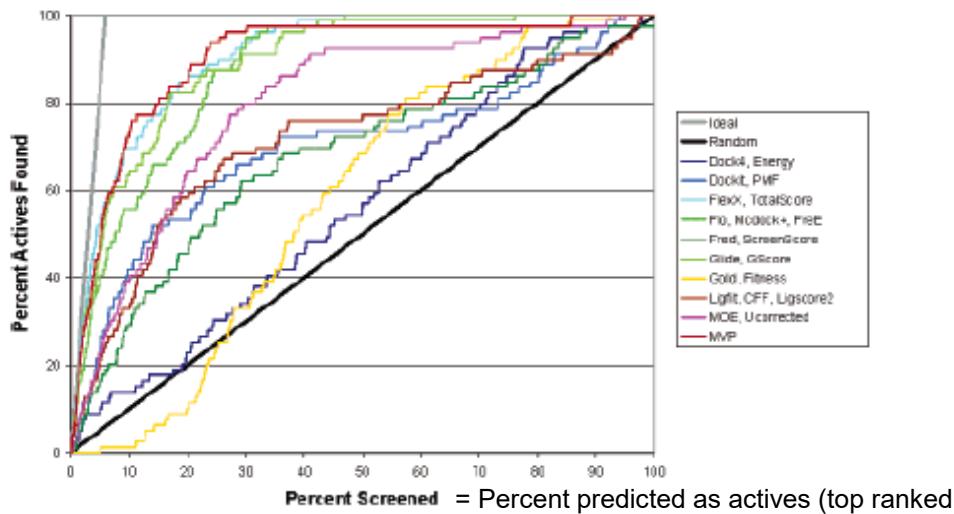


Virtual screening

Lead identification (virtual screening): G.L.Warren *J. Med. Chem.* 2006, 49, 5912-5931

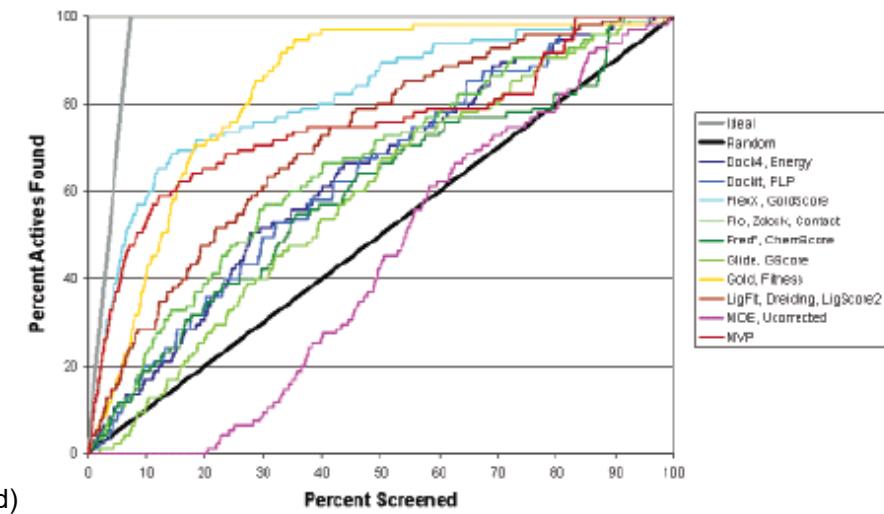
A.

Chk1 Kinase



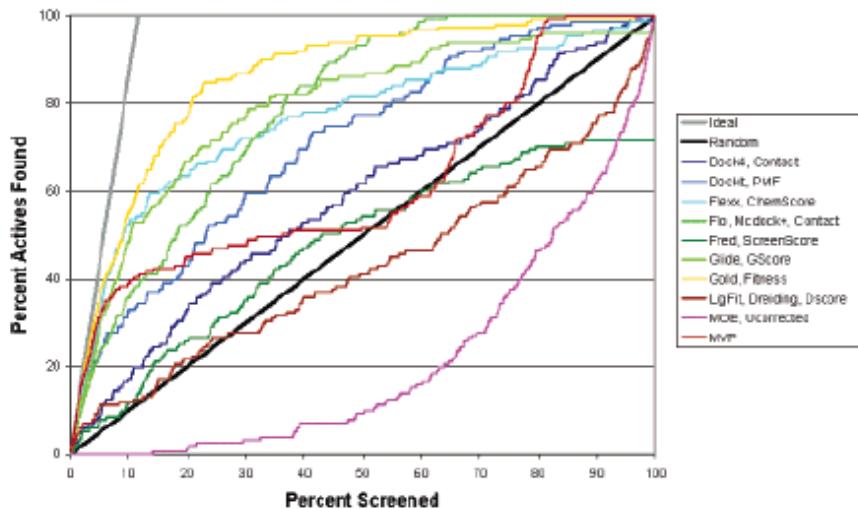
G.

Gyrase B



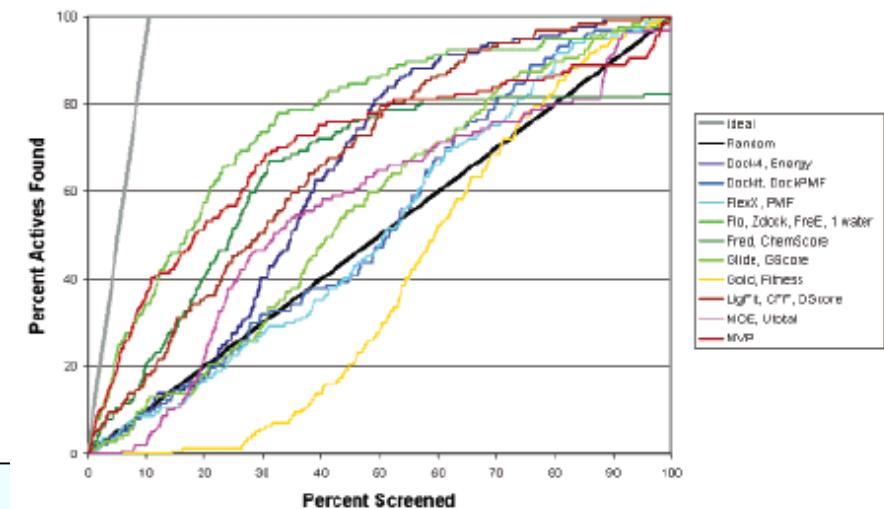
B.

PPAR δ



H.

HCV Polymerase





Virtual screening

G.L.Warren *J. Med. Chem.* **2006**, *49*, 5912-5931

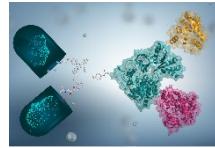
Enrichment factor:

$$EF = \frac{\% \text{ real actives found (identified in experiment)}}{\% \text{ predicted as actives (highest ranked)}}$$

Table 4. Enrichment Factor for Actives ($\leq 1 \mu\text{M}$) Found at 10% of the Docking-Score-Ordered List

program	Chk1	FXa	gyrase B	HCVP	MRS	<i>E. coli</i> PDF	<i>Strep</i> PDF	PPAR δ
ideal	10.0	9.8	10.0	9.5	10.0	7.6	8.3	8.6
Dock4	1.4	4.1	1.7	1.8	4.2	0.9	0.8	1.7
DockIt	4.2	2.0	2.0	1.0	1.0	0.2	0.0	3.2
FlexX	7.0	2.2	5.8	0.9	3.9	0.8	0.8	5.2
Flo+	5.6	2.7	2.3	3.4	1.7	1.5	0.8	3.6
Fred	2.9	4.1	1.9	2.0	0.6	3.2	1.2	1.1
Glide	6.3	3.4	1.0	1.0	5.3	0.6	0.4	4.8
Gold	0.1	4.1	4.0	0.0	0.8	1.0	0.1	5.5
LigandFit	3.3	1.9	2.8	1.8	2.9	2.9	1.7	1.2
MOEDock	3.9	0.6	0.0	0.0	1.0	2.1	0.6	0.0
MVP	7.2	5.8	5.3	3.6	6.4	6.7	6.9	3.9

For MVP: Manually defined pharmacophore points → MedChem knowledge can improve results significantly



Virtual screening

In general:

- Docking methods yield enrichment above random for most protein systems
- Best docking/scoring method depends strongly on protein system
- Success rate drops with
 - increasing similarity of actives and decoys



Prediction of binding affinities

Prediction of binding affinities (rank order): G.L.Warren *J. Med. Chem.* **2006**, *49*, 5912

Weak correlation between predicted and experimental binding affinities (r value ranges from 0 to -0.57 → r^2 from 0 to 0.33)

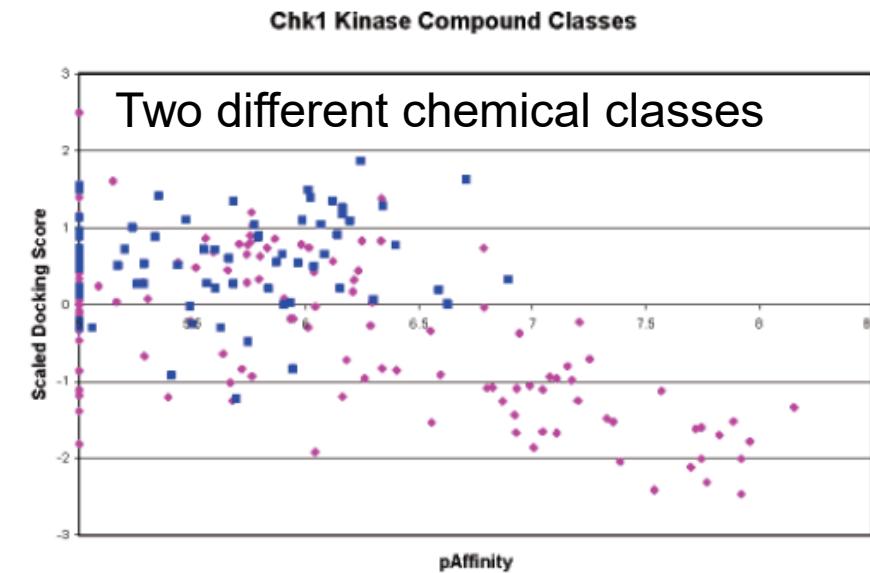
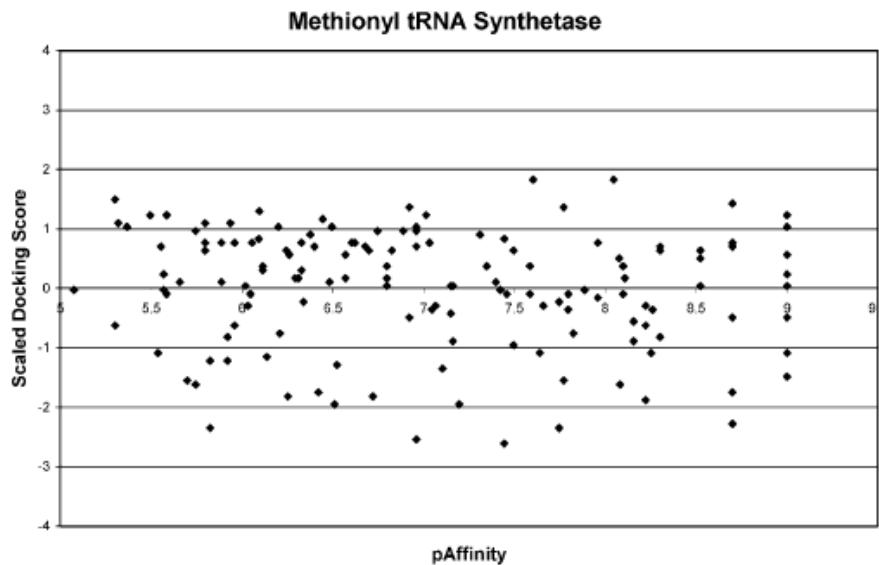
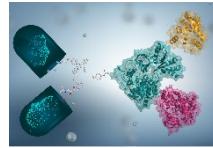


Table 7. Best Correlation Coefficient r between the $-\log$ Affinity (pAffinity) and Docking Score for All Programs across All Targets

program	Chk1	FXa	gyrase B	HCVP	MRS	<i>E. coli</i> PDF	<i>Strep</i> PDF	PPAR δ
Dock4	-0.33	-0.31	-0.39	0.00	-0.13	-0.38	-0.34	0.07
DockIt	-0.49	-0.19	-0.37	0.04	-0.28	-0.13	-0.30	-0.34
FlexX	-0.57	-0.31	-0.39	-0.12	-0.01	-0.42	-0.25	-0.36
Flo+	-0.44	-0.38	-0.36	-0.09	0.05	-0.27	-0.39	-0.42
Fred	-0.14	0.01	-0.13	-0.07	0.13	0.07	-0.24	0.06
Glide	-0.47	-0.08	-0.21	-0.04	0.08	-0.13	-0.12	-0.35
Gold	-0.42	-0.05	-0.14	-0.09	0.04	-0.12	-0.11	-0.43
LigandFit	-0.45	-0.13	-0.39	-0.06	-0.15	-0.21	-0.49	-0.10
MOEDock	-0.29	0.00	0.07	-0.01	-0.13	0.08	0.20	0.17
MVP	-0.26	0.10	-0.33	-0.01	-0.18	-0.17	-0.16	-0.18



Prediction of binding affinities

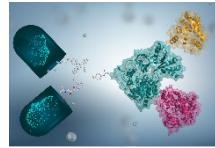
Phys. Chem. Chem. Phys., 2016,
18, 12964-12975:

- regression over all systems
- ! inconsistencies in experimental binding affinities

Table 3 Overall prediction accuracies of all docking programs in the scoring power test

Docking program	Correlation coefficient	Top scored pose	Best pose
AutoDock (LGA)	r_p^a	0.433 ± 0.009^c	0.404 ± 0.009
	r_s^b	0.477 ± 0.008	0.450 ± 0.009
AutoDock (PSO)	r_p	0.492 ± 0.008	0.466 ± 0.008
	r_s	0.534 ± 0.007	0.513 ± 0.008
AutoDock Vina	r_p	0.564 ± 0.008	0.569 ± 0.008
	r_s	0.580 ± 0.008	0.584 ± 0.008
LeDock	r_p	0.442 ± 0.009	0.463 ± 0.009
	r_s	0.462 ± 0.010	0.486 ± 0.009
rDock	r_p	-0.015 ± 0.011	-0.021 ± 0.011
	r_s	-0.017 ± 0.011	-0.005 ± 0.011
UCSF DOCK	r_p	0.291 ± 0.010	0.276 ± 0.011
	r_s	0.331 ± 0.011	0.323 ± 0.011
LigandFit	r_p	-0.132 ± 0.011	-0.105 ± 0.011
	r_s	-0.221 ± 0.012	-0.192 ± 0.012
Glide (SP)	r_p	0.444 ± 0.008	0.402 ± 0.009
	r_s	0.473 ± 0.009	0.419 ± 0.010
Glide (XP)	r_p	0.367 ± 0.010	0.356 ± 0.010
	r_s	0.389 ± 0.010	0.374 ± 0.010
GOLD	r_p	-0.500 ± 0.008	-0.494 ± 0.008
	r_s	-0.515 ± 0.008	-0.513 ± 0.008
MOE Dock	r_p	0.564 ± 0.008	0.411 ± 0.009
	r_s	0.589 ± 0.009	0.457 ± 0.009
Surflex-Dock	r_p	-0.340 ± 0.009	-0.350 ± 0.009
	r_s	-0.370 ± 0.009	-0.382 ± 0.009

^a r_p represents Pearson's correlation coefficient. ^b r_s represents Spearman's ranking coefficient. ^c The standard error was estimated by randomly sampling 80% of the tested dataset 100 repeats.

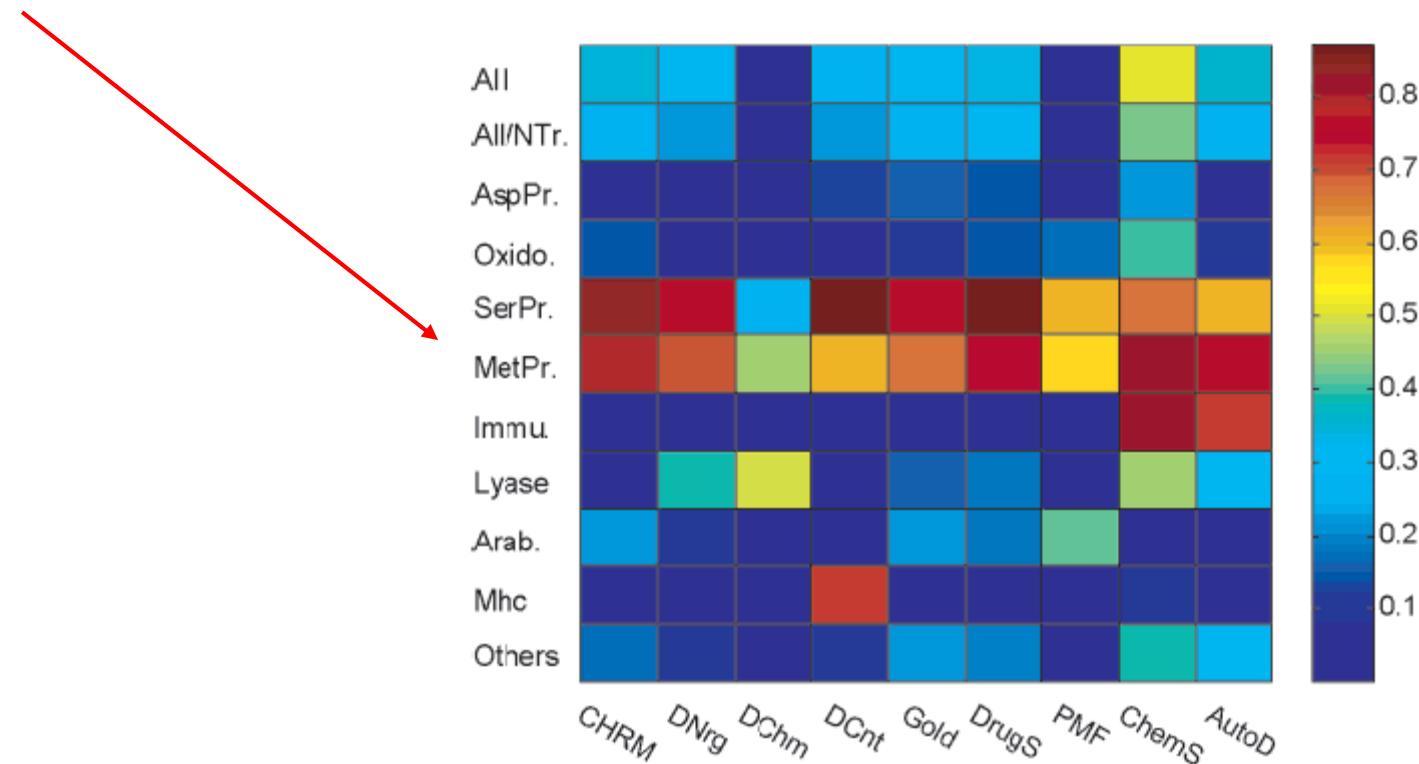


Prediction of binding affinities

In general: Regression coefficient r^2 usually well below 0.5.

P. Ferrara *J. Med. Chem.* 2004, 47, 3032:

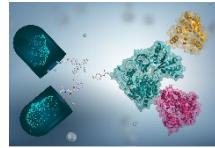
Exception: Serine and metallo-proteases



BUT: Correlation between $\log(\text{MW})$ and affinity:

SerPr: 0.81

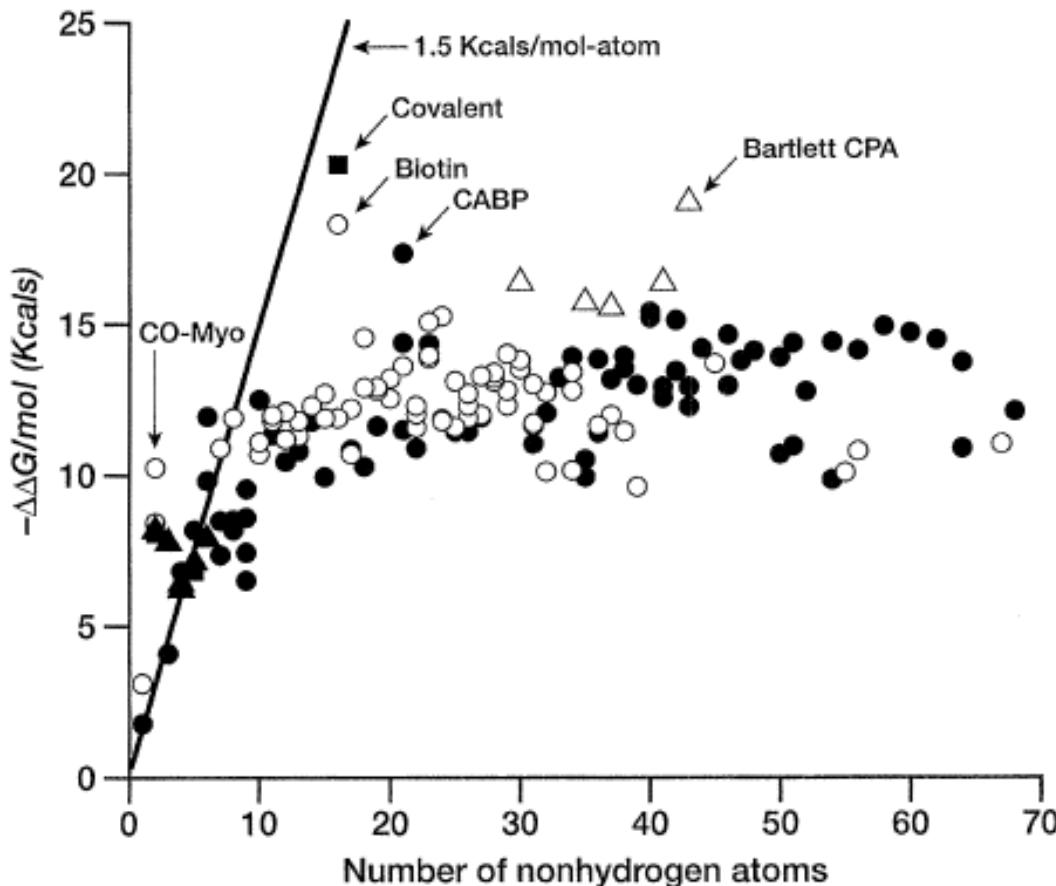
MetPr: 0.58

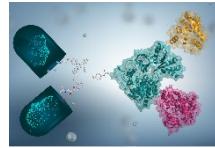


Prediction of binding affinities

I.D. Kuntz *Proc. Natl. Acad. Sci.* **1999**, *96*, 9997

Linear correlation between experimental binding affinity and number of heavy atoms in ligand (up to 15 heavy atoms):





What do to with noisy results?

Reasons for inaccuracy:

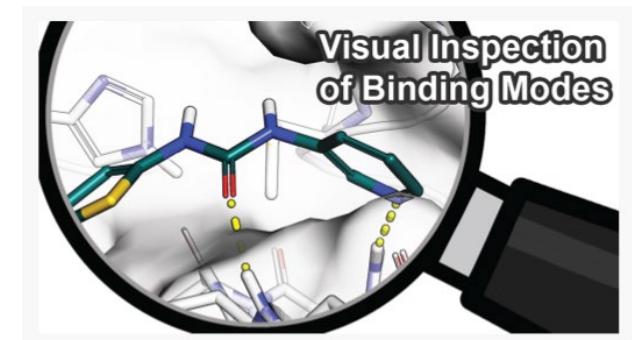
- Evaluation of binding pose based on single static protein-ligand complex (even if flexibility included in docking) → Neglect of protein-ligand thermodynamics including entropy

Approaches to deal with noisy, error-prone predictions:

- Visual inspection

Decision Making in Structure-Based Drug Discovery: Visual Inspection of Docking Results

André Fischer, Martin Smieško, Manuel Sellner, and Markus A. Lill*



- Post-processing
 - Physicochemical methods
 - AI



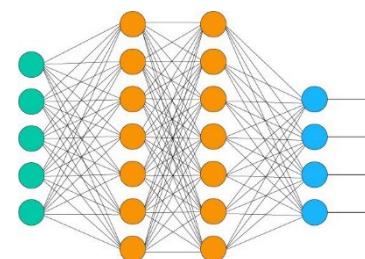
Post-processing

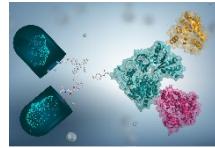
Store best N poses (for M best compounds in enrichment experiments) from docking simulation

Refine poses (e.g. minimization)
→ protein flexibility

Generate ensemble of configurations (e.g. MD)

Re-rank poses (and compounds) using higher-accuracy free energy methods based on MD or novel artificial intelligence methods





Take home message

Don't believe every result from the docking simulation.

Docking might give you good ideas about binding poses and active lead compounds,

BUT always use your medicinal chemistry knowledge:

- Check if similar compounds are all predicted as actives or in the same binding mode [Still be careful: Algorithm might predict pose or activity wrong for all similar compounds]
- Test known compounds with known binding modes first → Estimate accuracy of selected system:
Docking/Scoring method & Target protein
- Do experimental validation: X-ray, NMR, site-directed mutagenesis etc.



References

Further references:

J. C. Cole, C. W. Murray, J. W. M. Nissink, R. D. Taylor, R. Taylor (2005) Comparing Protein–Ligand Docking Programs Is Difficult. *Proteins* **60**, 325–332.

A.N. Jain (2008) Bias, reporting, and sharing: computational evaluations of docking methods. *J. Comput. Aided Mol. Des.* **22**, 201-212.

Protein flexibility



Protein flexibility

In about half of all protein systems protein flexibility plays an important role in ligand binding → rigid-protein assumption results in deleterious effect on docking results.

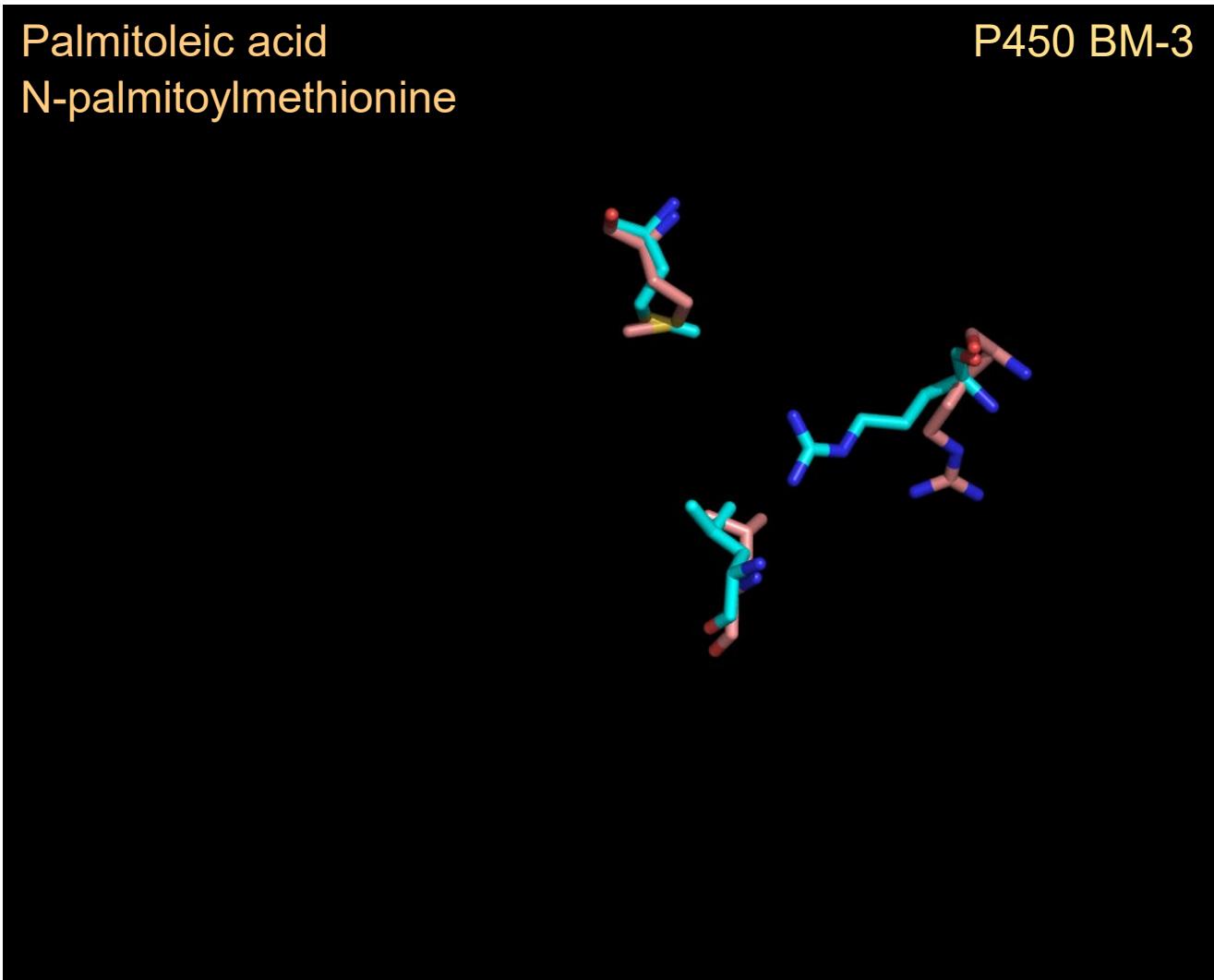
Types of protein flexibility:

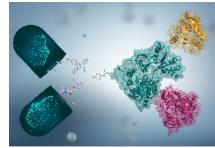
- side-chain motion
- small backbone motion (“breathing motion”)
- loop flexibility
- large conformational changes



Protein flexibility

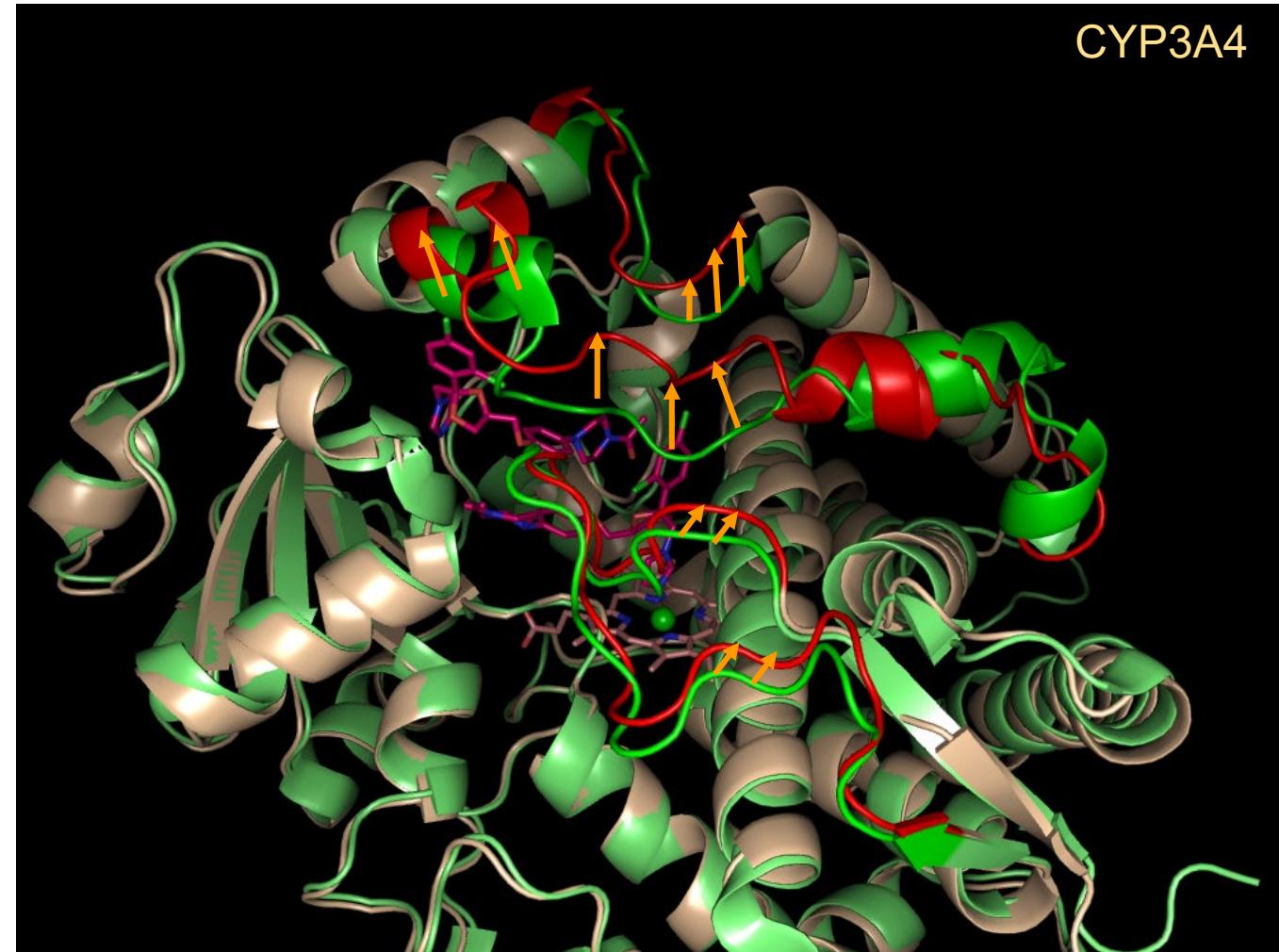
- Side-chain flexibility





Protein flexibility

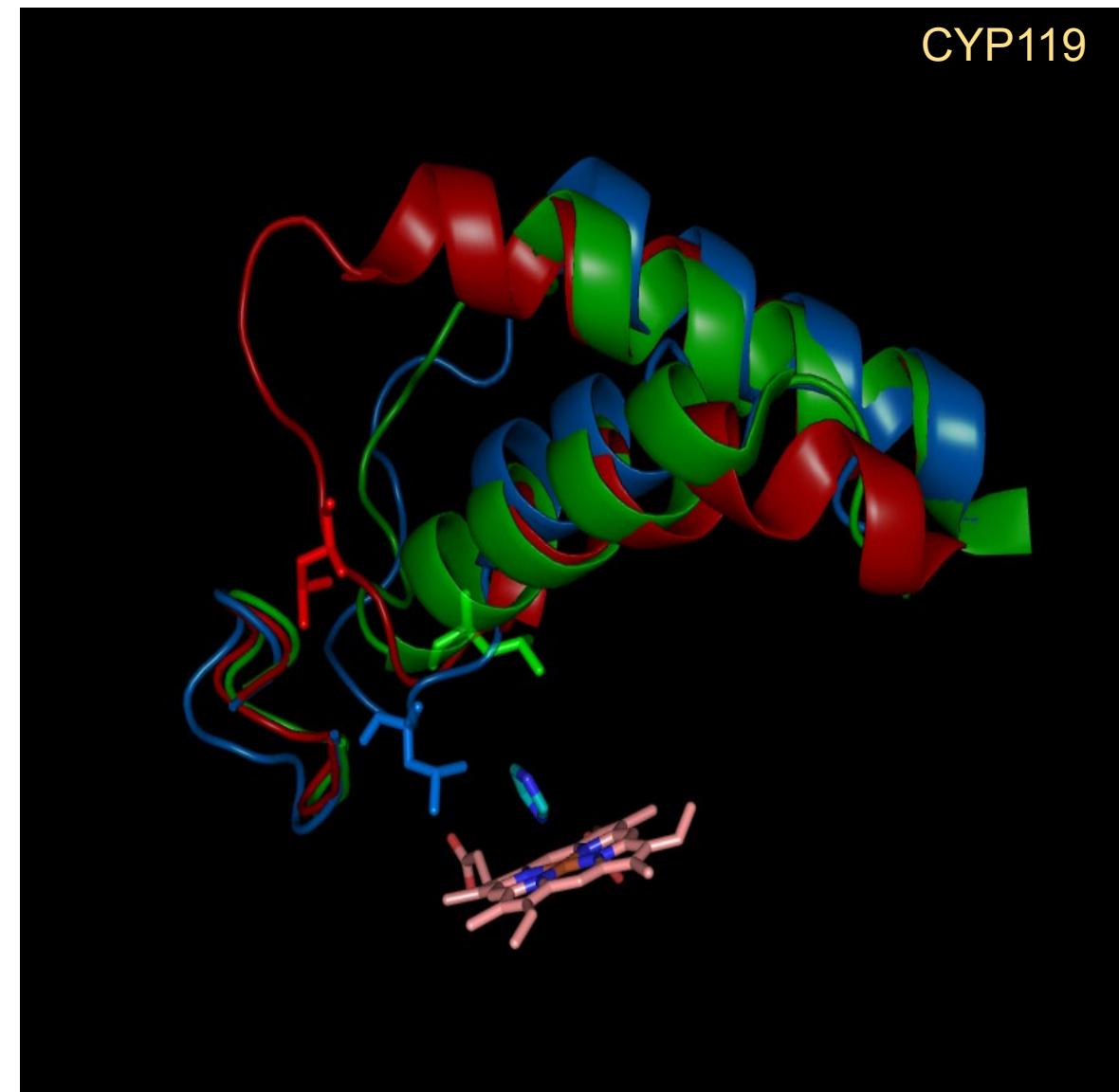
- Side-chain flexibility
- Breathing motion





Protein flexibility

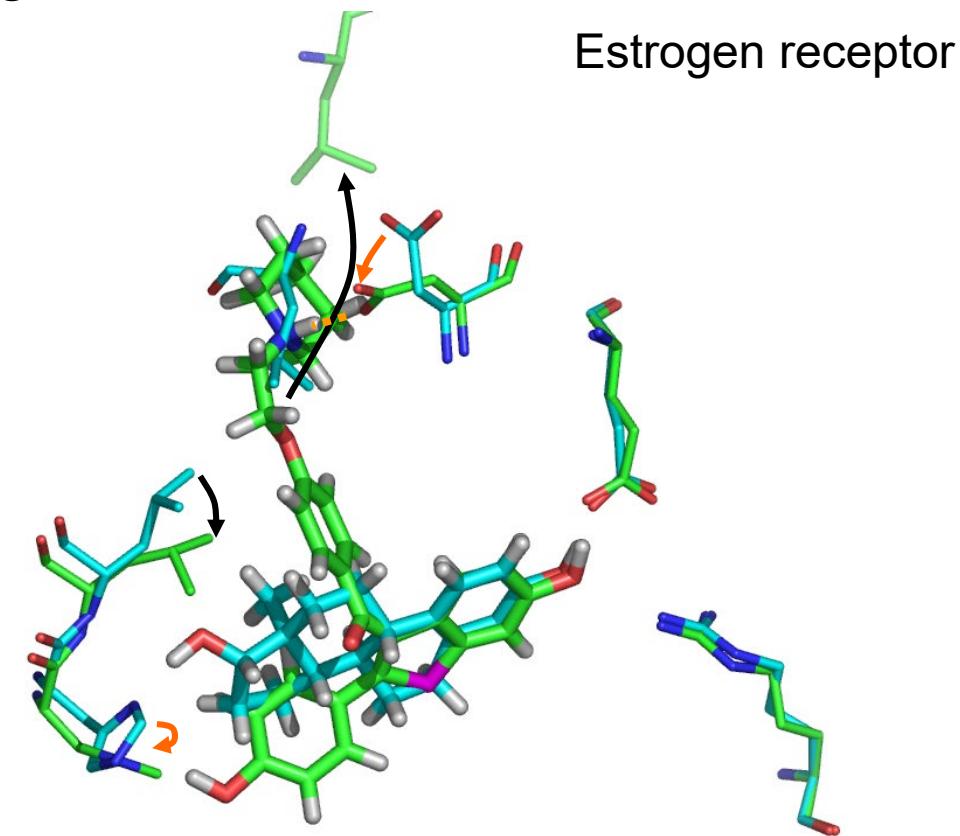
- Side-chain flexibility
- Breathing motion
- Loop flexibility





Protein flexibility

- Side-chain flexibility
- Breathing motion
- Loop flexibility
- Large conformational changes





Computational approaches

Side-chain flexibility:

- **Rotamer libraries**

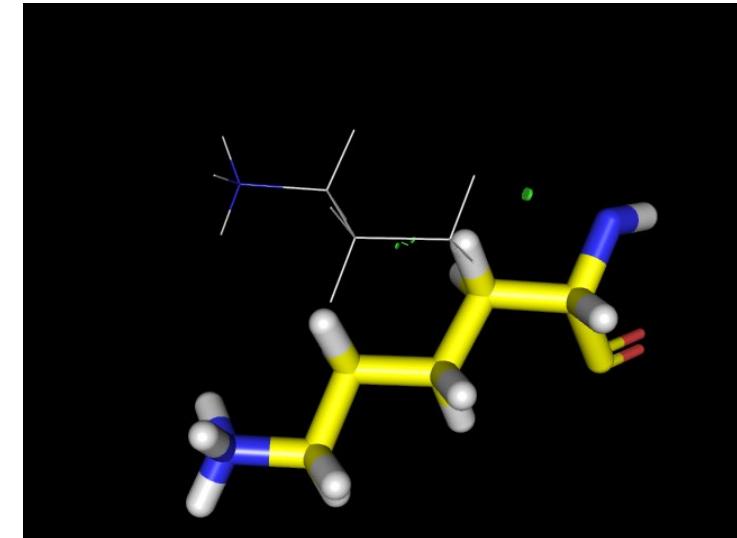
Large set of PBD structures

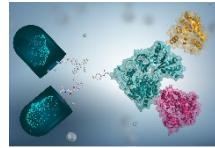
→ bin statistically preferred rotamer states

During docking search for optimal rotamer states for each docking pose.

LYS	-177.1	177.0	-176.6	-67.4
LYS	179.6	-178.0	-75.8	70.6
LYS	-175.0	-177.3	-70.2	-176.8
LYS	-175.4	-178.7	-71.0	-65.1
LYS	-162.4	-81.9	82.2	74.2
LYS	-163.0	-83.1	95.7	-178.9
LYS	-164.5	-82.8	88.3	-80.2
LYS	-161.2	-77.8	-177.5	66.4
LYS	-165.2	-87.2	-177.2	-178.9
LYS	-166.1	-83.0	-177.8	-67.3
LYS	-160.1	-90.5	-80.9	68.3
LYS	-166.8	-96.6	-69.4	178.4
LYS	-168.0	-93.8	-76.0	-65.7
LYS	-88.0	81.2	68.0	69.3
LYS	-80.9	81.3	66.9	-179.6
LYS	-86.8	86.0	84.3	-83.2
LYS	-80.7	75.7	176.2	62.3
LYS	-81.0	74.0	-178.3	177.9
LYS	-87.6	74.5	178.0	-61.8
LYS	-86.6	83.2	-85.6	68.2
LYS	-82.6	71.2	-96.3	-172.0

Example for list of rotamers [Lys:
4 rotatable bonds
= torsions]:





Computational approaches

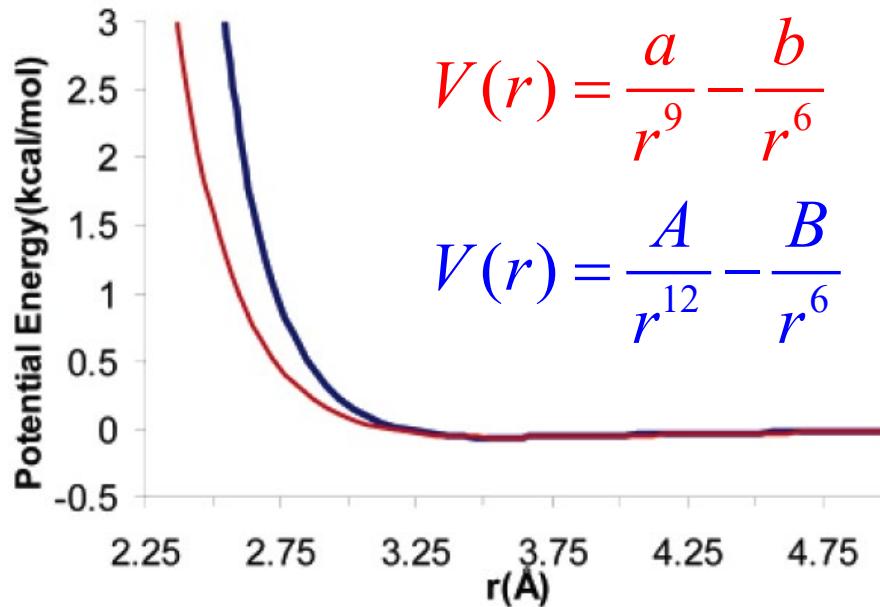
- **Ensemble docking:**

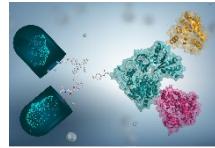
Dock ligand into each member of ensemble of protein structures individually

Source: X-ray, NMR, MD/MC simulations

- **Soft docking:**

Reduce repulsive van der Waals interactions





Computational approaches

A.M. Ferrari *J.Med.Chem.* 2004, 47, 5076.

Ensemble docking versus soft docking versus rigid protein docking:

Enrichment for two target systems:

(red: rigid protein

blue: soft docking

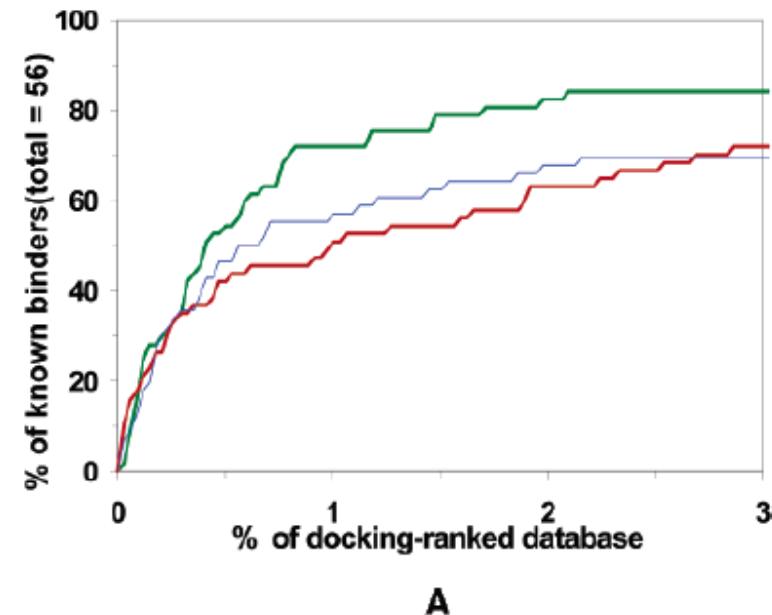
green: ensemble of X-ray structures)

Details for ensemble docking:

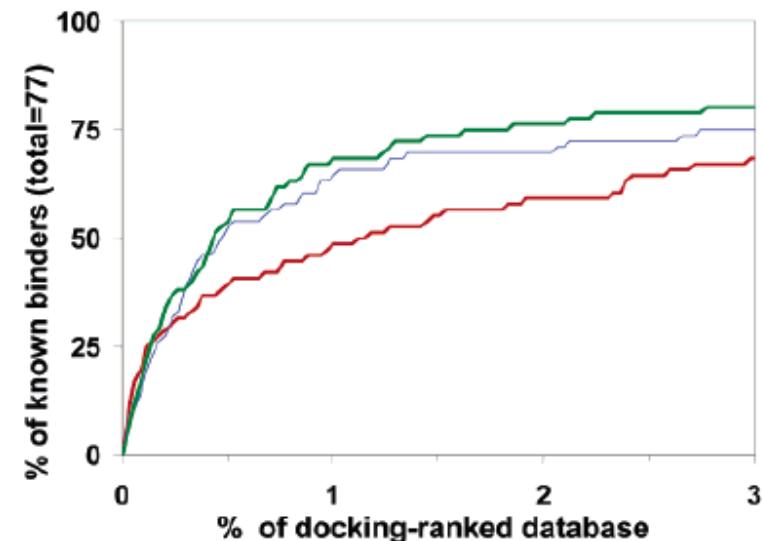
Overlay several structures of same protein
with different co-crystallized ligands.

Identify flexible regions. Overall template:

Rigid core of protein plus flexible parts in
alternative conformations.



A



B



Computational approaches

Ensemble docking:

If not multiple X-ray protein structures are available → run MD/MC simulations to generate ensemble of target structures.

Simulation on apo structure:

Often the protein has to accommodate its shape (and properties) to the bound ligand. Apo form might be in different conformation than necessary to form energetically favorable complex with ligands → Docking to trajectory of halo form is often (but not always) more successful.

Simulation on a single ligand-bound structure:

Docking into the ligand bound trajectory might bias docking towards the ligand used in the MD simulation → Docking of structurally diverse compounds might give worse results than apo form.

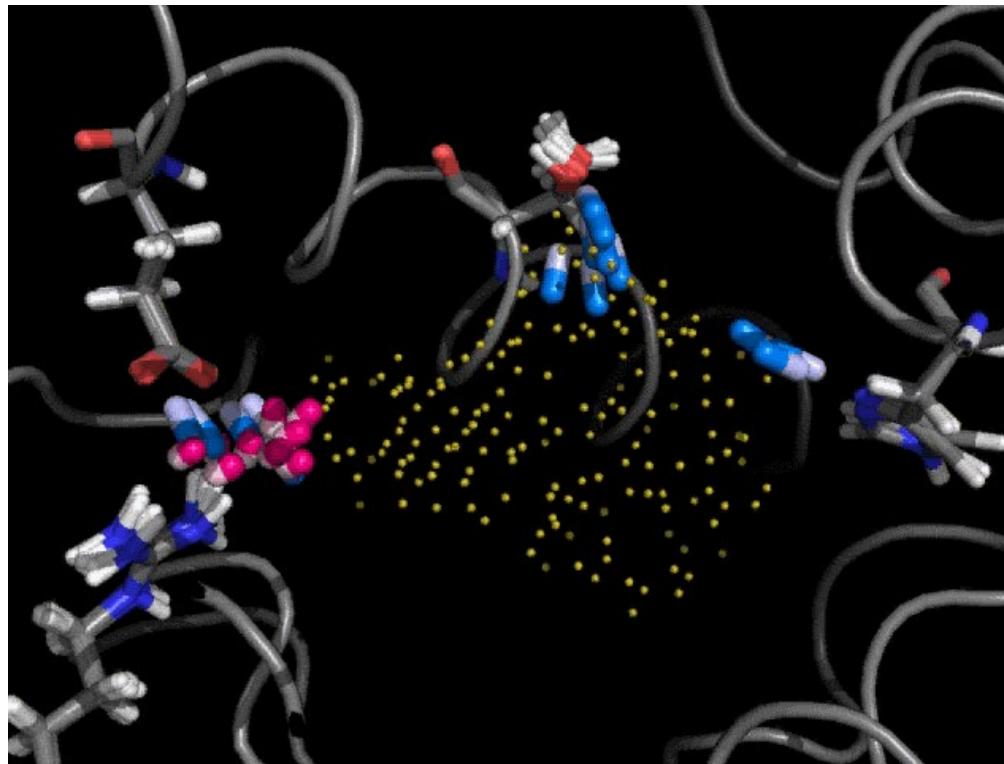
“Holo form too specialized for specific class of ligands”



Computational approaches

Ensemble docking:

MD simulation in frequently changing ligand model (Xu, Lill
J.Chem.Inf.Model. 2011)

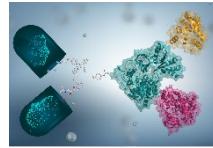




Induced-fit docking

Schrodinger

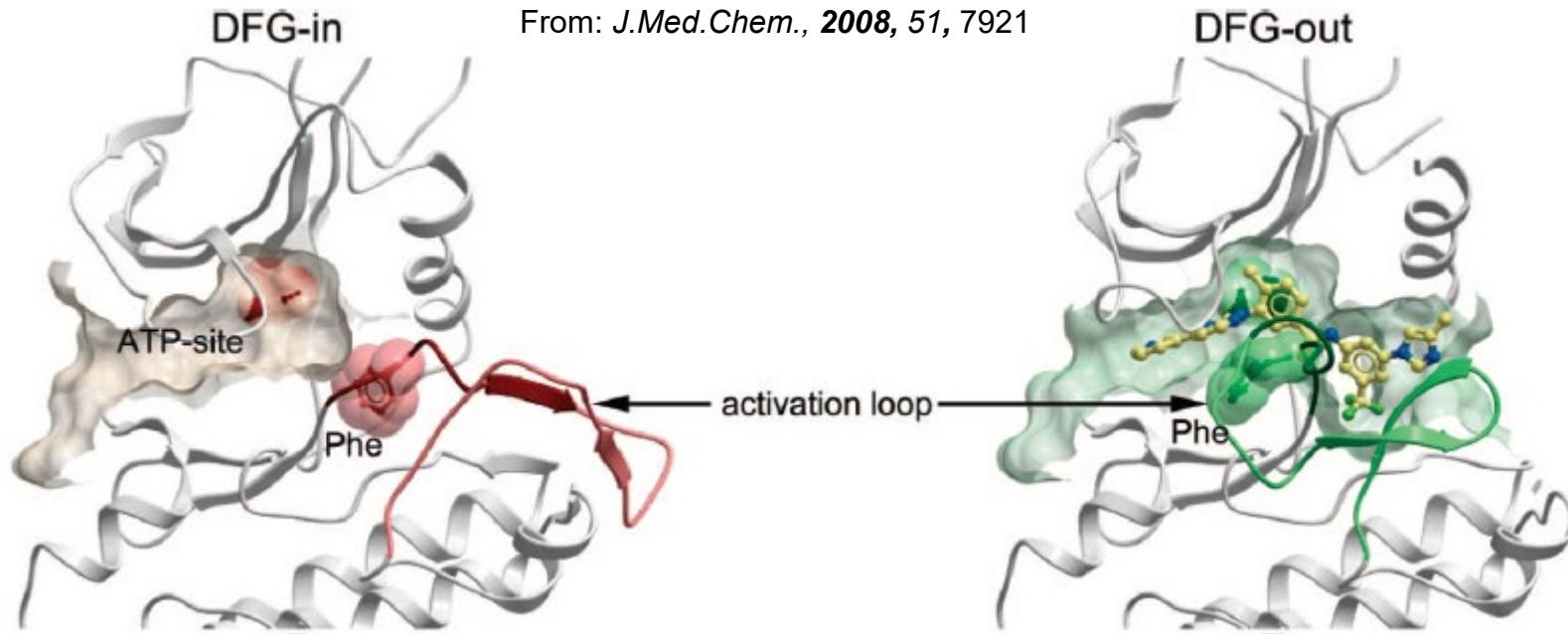
1. Docking with reduced vdW radii (and flexible residues removed) using Glide
2. Find optimal side chain conformations (using Prime)
3. Minimize protein-ligand structure
4. Re-dock and score



Loop flexibility

Loop regions are usually significantly less stable than α -helices or β -sheets \rightarrow regularly observed conformational changes for different ligands bound.

e.g. kinases:



Possible approach:

Predict alternative loop conformations (e.g. with PLOP, LOOPY) \rightarrow use alternative templates for docking (similar to ensemble docking)

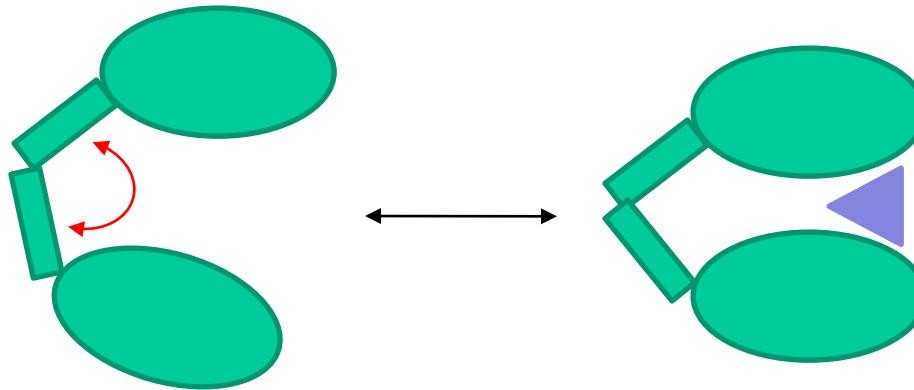


Large scale conformational changes

Possible approaches:

- Identify hinge regions in protein

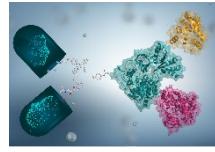
Use reaction coordinates for hinge-bending motion as additional degrees of freedom



- Run MD simulation

Do principal component analysis (PCA) → essential modes of protein dynamics (= reaction coordinates for main global conformational changes)

Use essential modes as additional degrees of freedom in docking



References

- Totrov, M., and Abagyan, R. Flexible ligand docking to multiple receptor conformations: A practical alternative. *Curr. Opin. Struct. Biol.* **2008**, 18, 178 – 184.
- Lill, M.A. *Biochemistry* **2011**, 50, 6157 – 6169