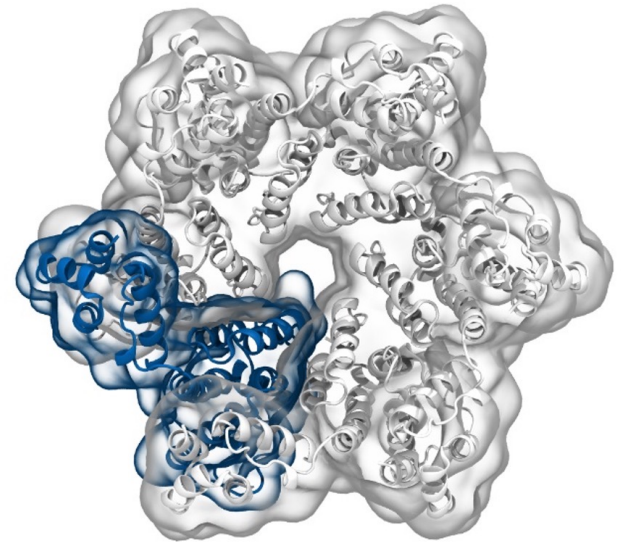
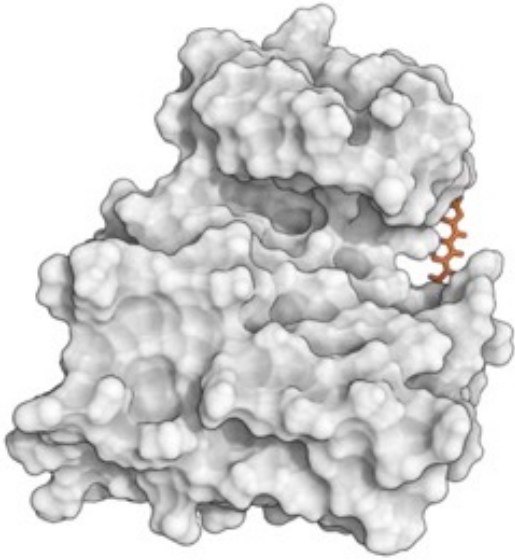


# Simulation of Biomolecules

## Docking

### 2024 CCP5 Summer School



Dr Matteo Degiacomi  
Durham University

[matteo.t.degiacomini@durham.ac.uk](mailto:matteo.t.degiacomini@durham.ac.uk)

Dr Antonia Mey  
University of Edinburgh

[antonia.mey@ed.ac.uk](mailto:antonia.mey@ed.ac.uk)

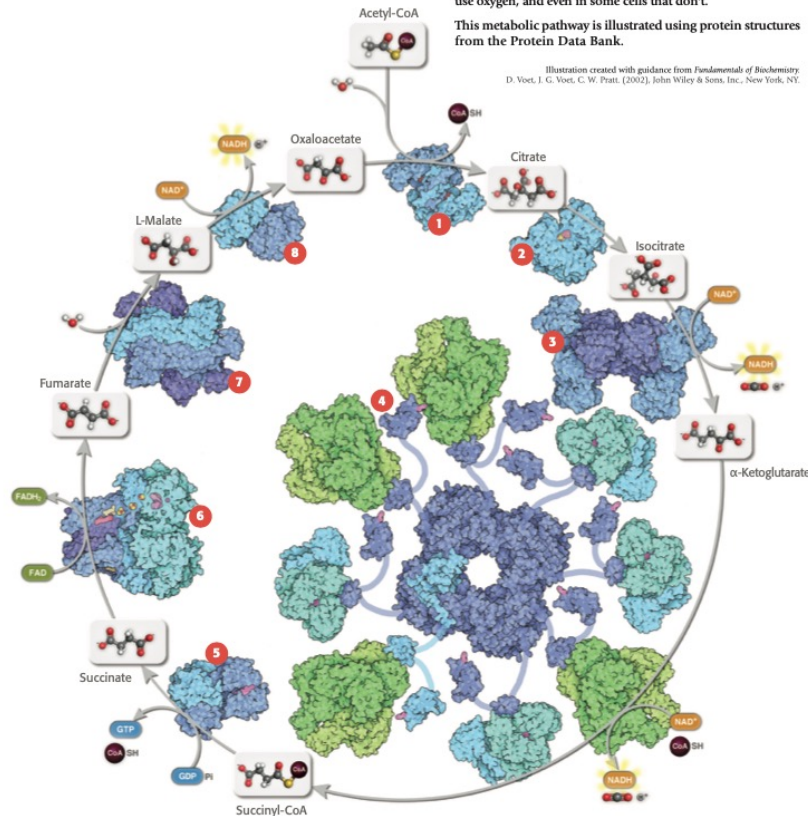
# Life is built on protein and small molecule interactions

## The Structures of the Citric Acid Cycle

Also known as the Krebs cycle or the tricarboxylic acid cycle, the *citric acid cycle* is at the center of cellular metabolism. It plays a starring role in both the process of energy production and biosynthesis. The cycle finishes the sugar-breaking job started in glycolysis and fuels the production of ATP in the process. It is also a central hub in biosynthetic reactions, providing intermediates that are used to build amino acids and other molecules. Citric acid cycle enzymes are found in all cells that use oxygen, and even in some cells that don't.

This metabolic pathway is illustrated using protein structures from the Protein Data Bank.

Illustration created with guidance from *Fundamentals of Biochemistry*  
D. Voet, J. G. Voet, C. W. Pratt, [2005], John Wiley & Sons, Inc., New York, NY



### Eight Reactions

The eight reactions of the citric acid cycle use the small molecule *oxaloacetate* as a catalyst. The cycle starts by addition of an acetyl group to oxaloacetate, then, over the course of eight steps, the acetyl group is completely broken apart, finally restoring the oxaloacetate molecule for another round.



Small molecules are:

- substrates of enzymes
- Inhibitors or activators
- Co-factors

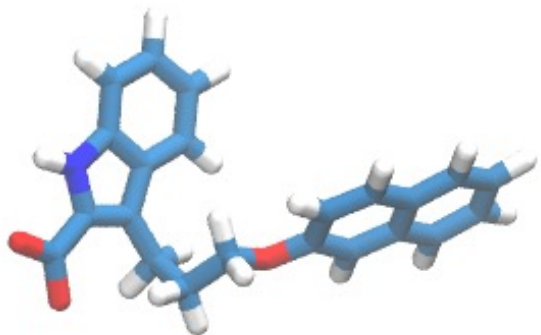
And play an important role in life.

Accurate interaction prediction is essential.

# What is docking?

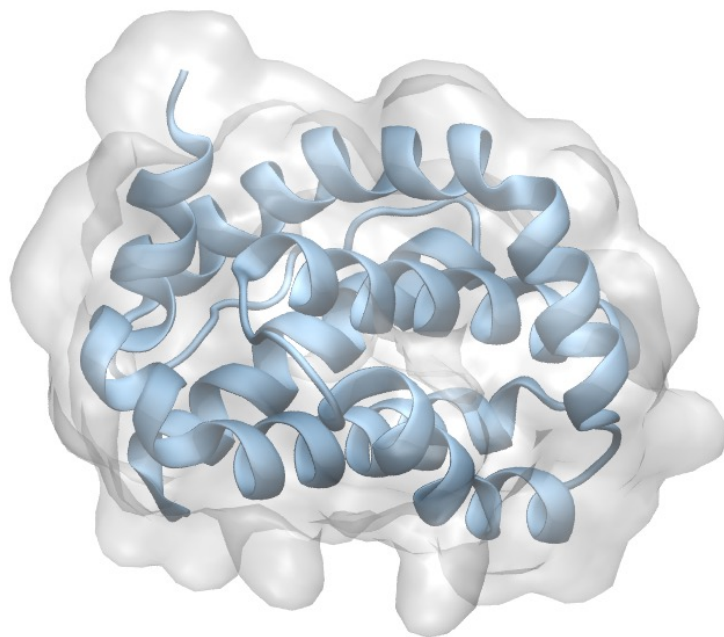
The process of predicting a stable 3D geometry of an interacting pair of molecules – a **binding mode/pose**.

Ligand:



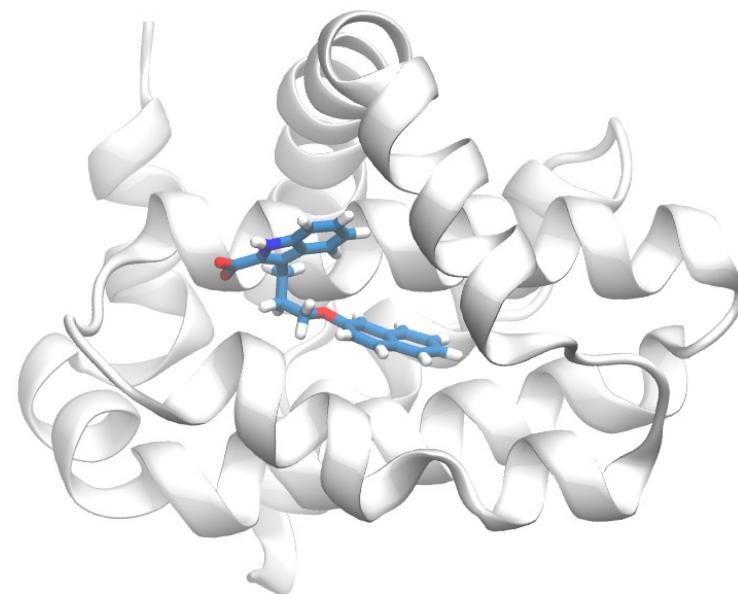
+

Receptor:



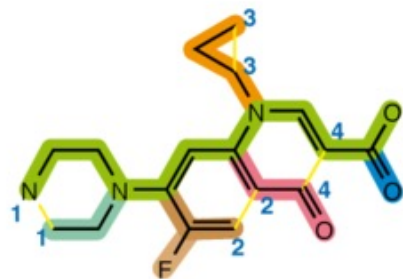
=

Binding pose:



# Typical workflow

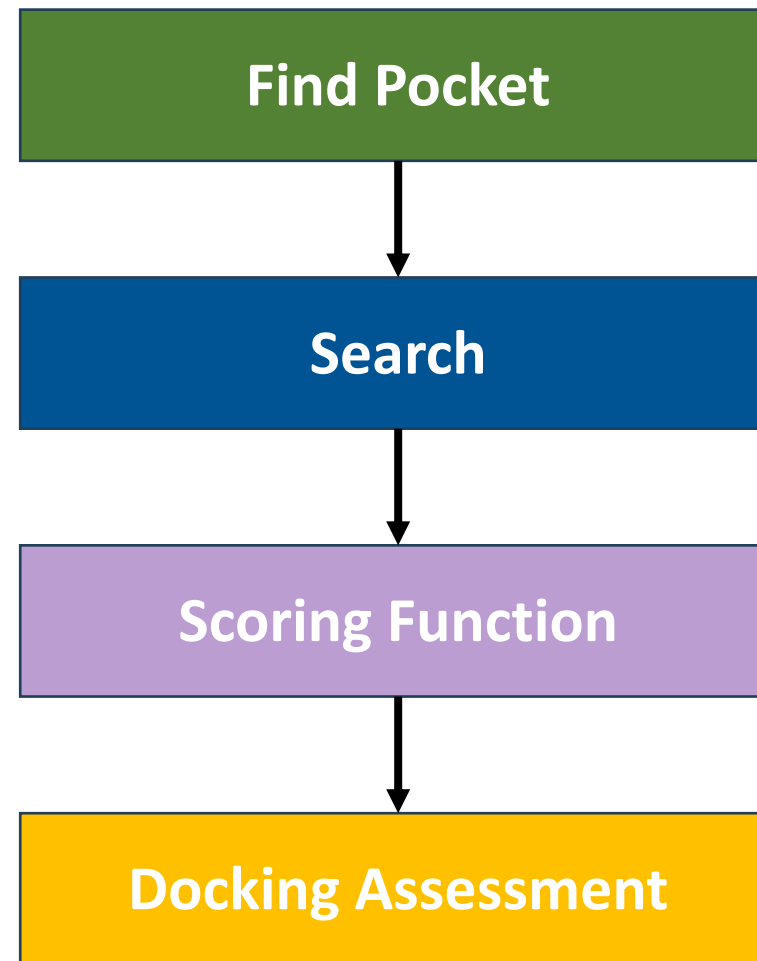
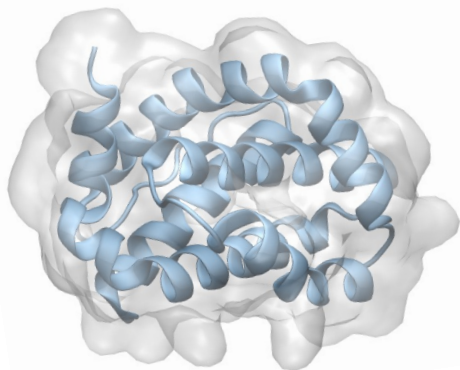
1-D or 2-D ligand structure



N1CCN(CC1)C(C(F)=C2)=CC(=C2C4=O)N(C3CC3)C=C4C(=O)O

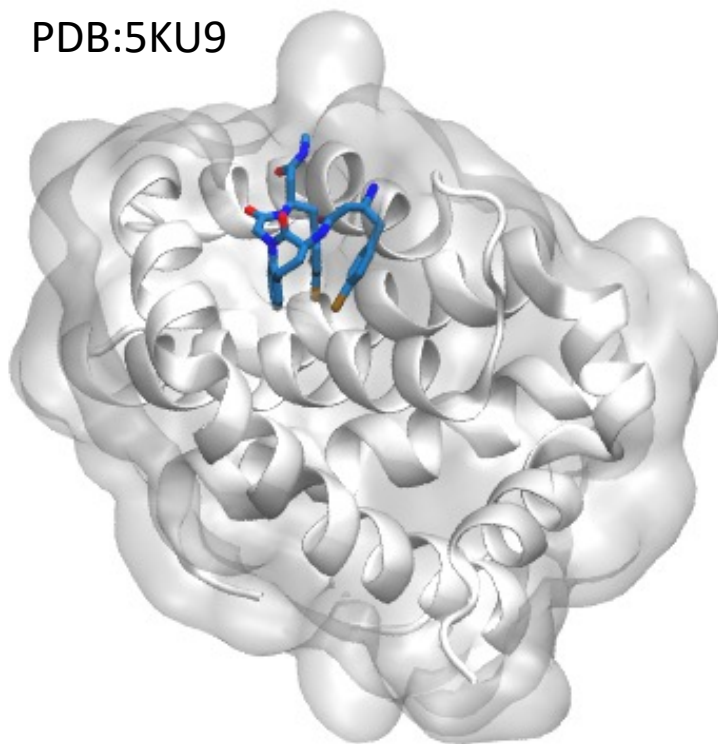
+

Protein structure



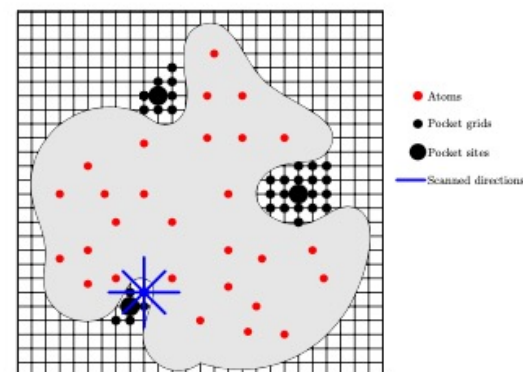
# Finding a pocket

PDB:5KU9

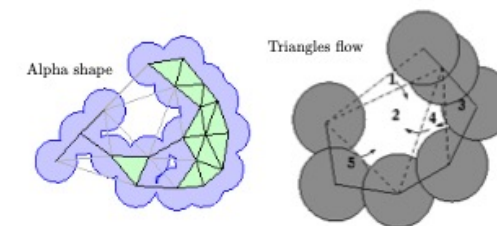


Using a reference atomic  
structure with an existing  
molecule bound

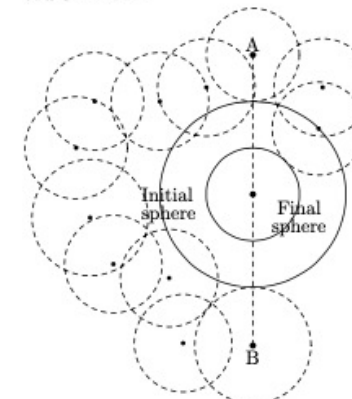
a. POCKET, LIGSITE, LIGSITE<sup>csc</sup>



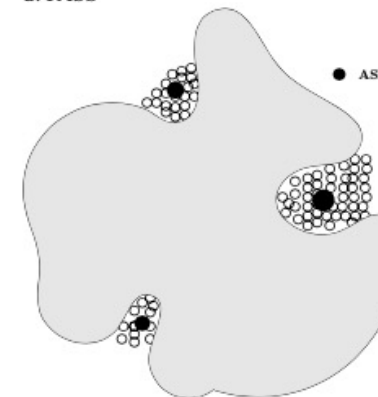
c. CAST



b. SURFNET



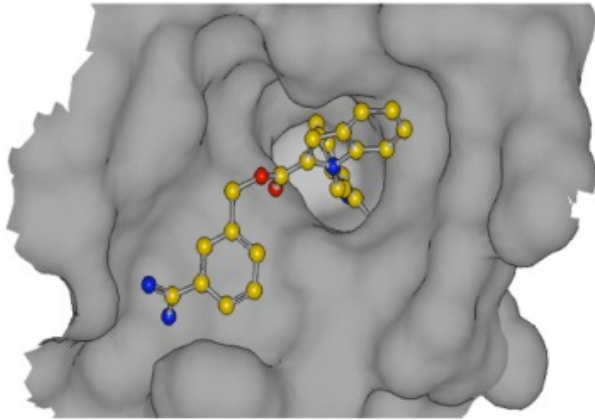
d. PASS



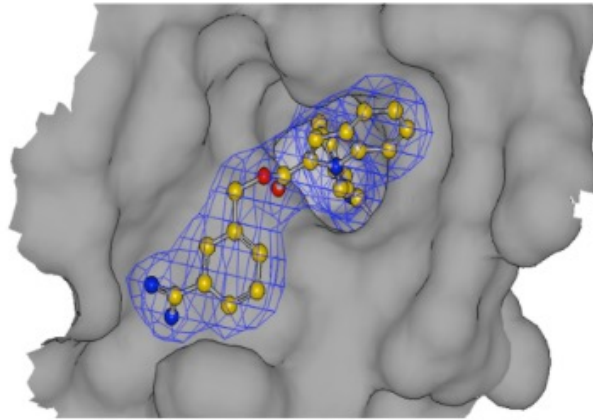
Using a pocket finding algorithm



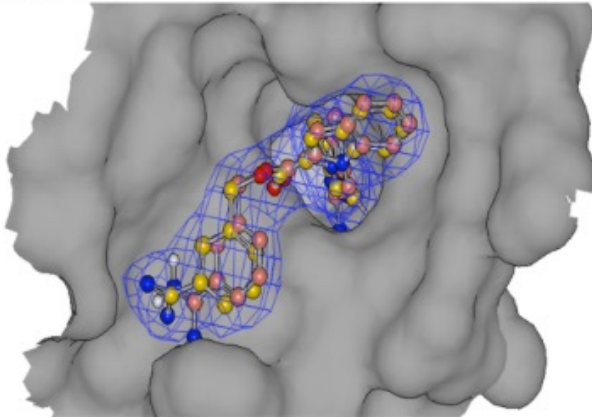
# Shape based methods



Find Bound Ligand



Identify Shape Constraint



Use Shape Constraint to Optimize Overlay

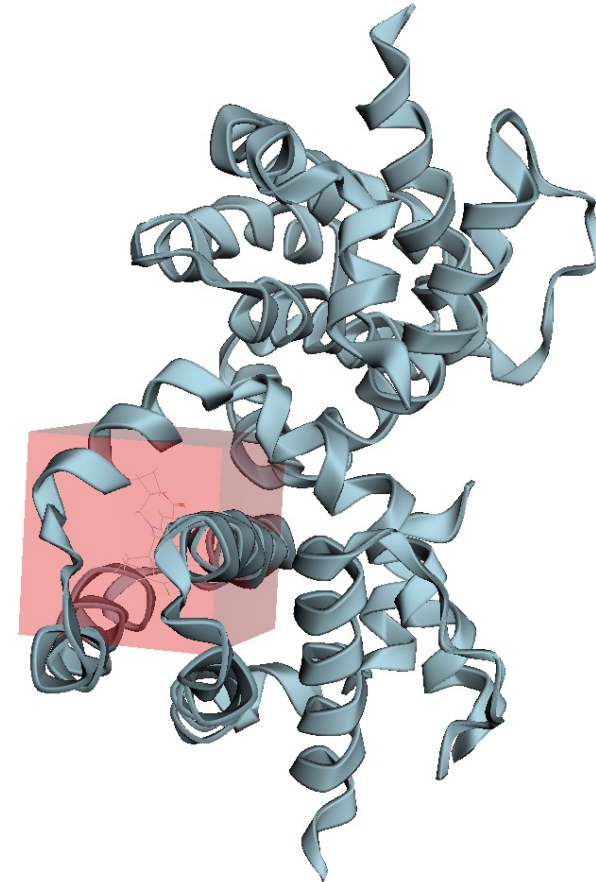
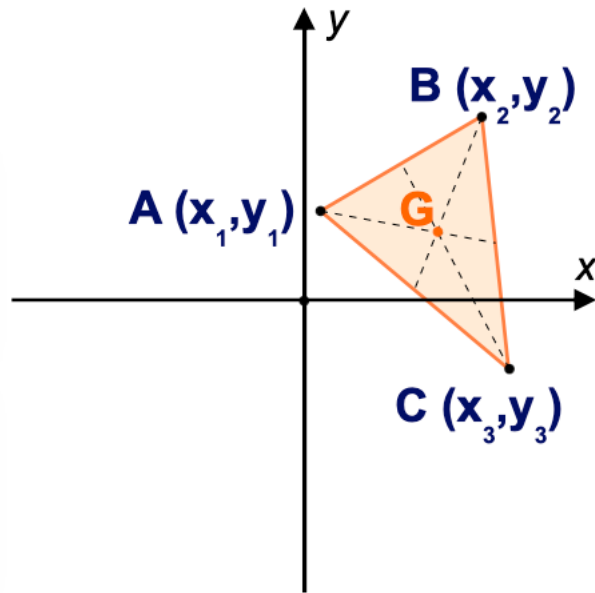
With an existing ligand it is possible to match the shape and optimize the overlay

**Fast and robust**

**Ligand changes are not taken into account**

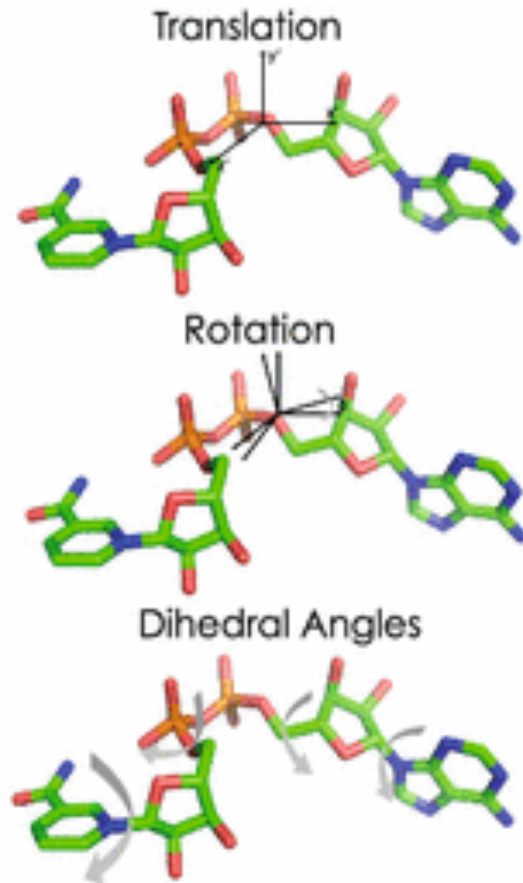
# Finding the docking grid area

Often you have a ligand template or binding site residue to help with designing the docking grid

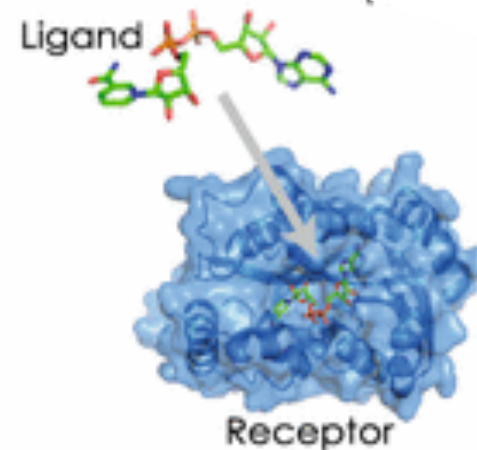


Use centre of geometry of a molecule from X-ray

# Genetic algorithm for ligand conformers



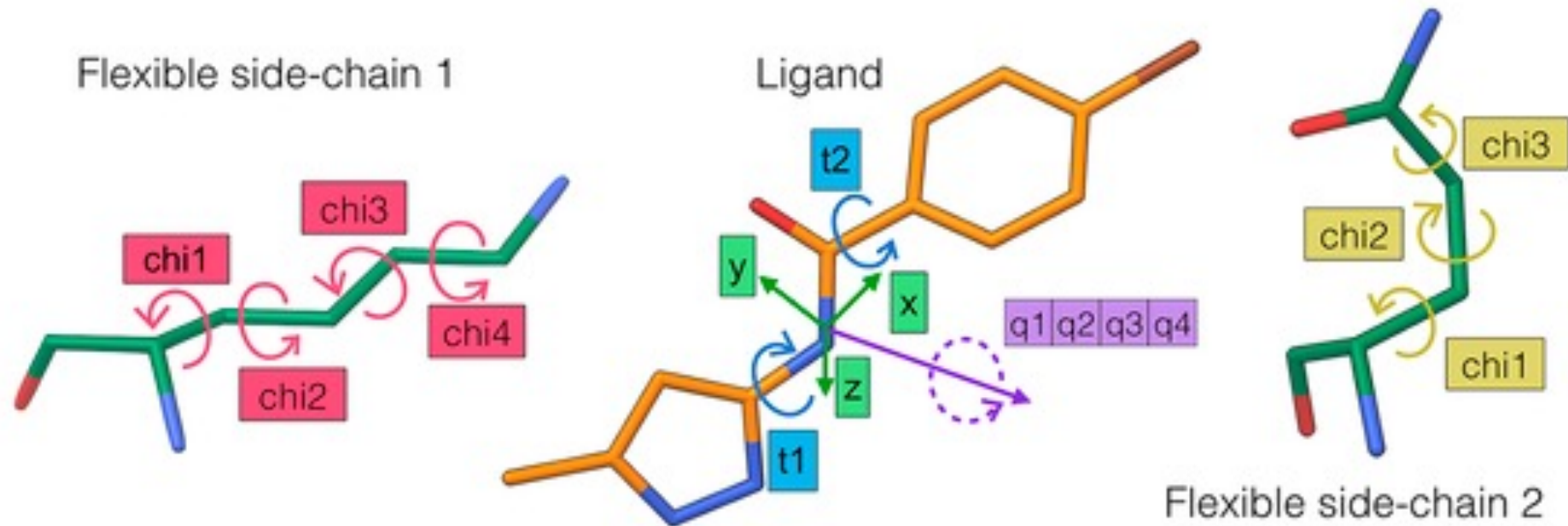
Testing different arrangements of the ligand in the rigid binding site of the protein



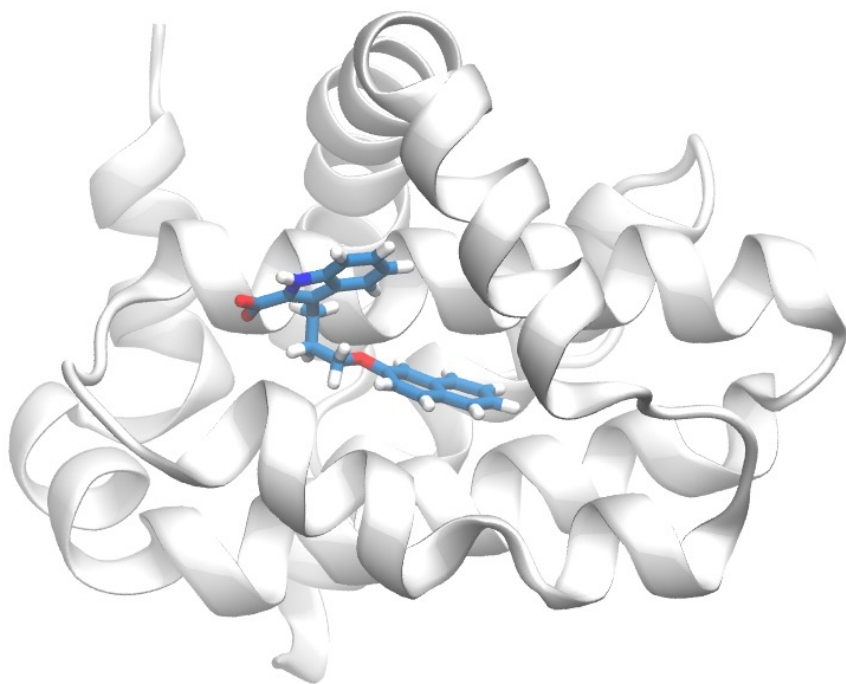


# Allowing protein and ligand flexibility is often better

Genes		Genome														
		Ligand								Receptor						
		Translation			Rotation				Torsion 1	Torsion 2	Flexible side-chain 1				Flexible side-chain 2	
Variables	x	y	z	q1	q2	q3	q4	t1	t2	chi1	chi2	chi3	chi4	chi1	chi2	chi3



# Flexibility increases compute time



$$N = T360/i$$

N: number of conformations

T: number of rotatable bonds

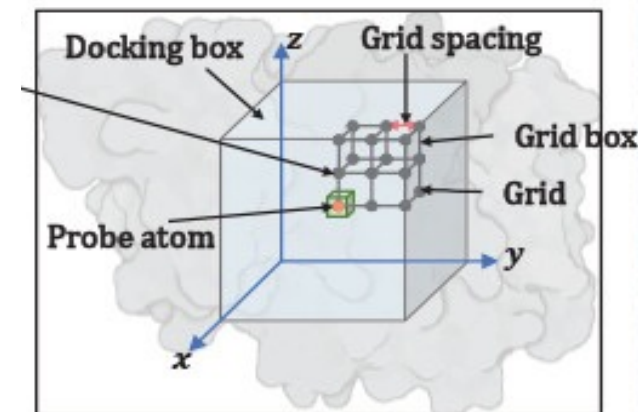
i: incremental degrees

**Typical drug molecule**

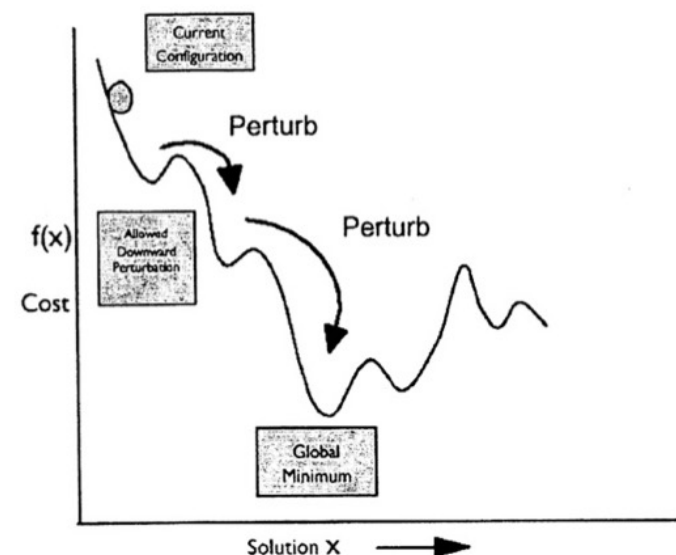
10 rotatable bonds

30° increments (discrete)

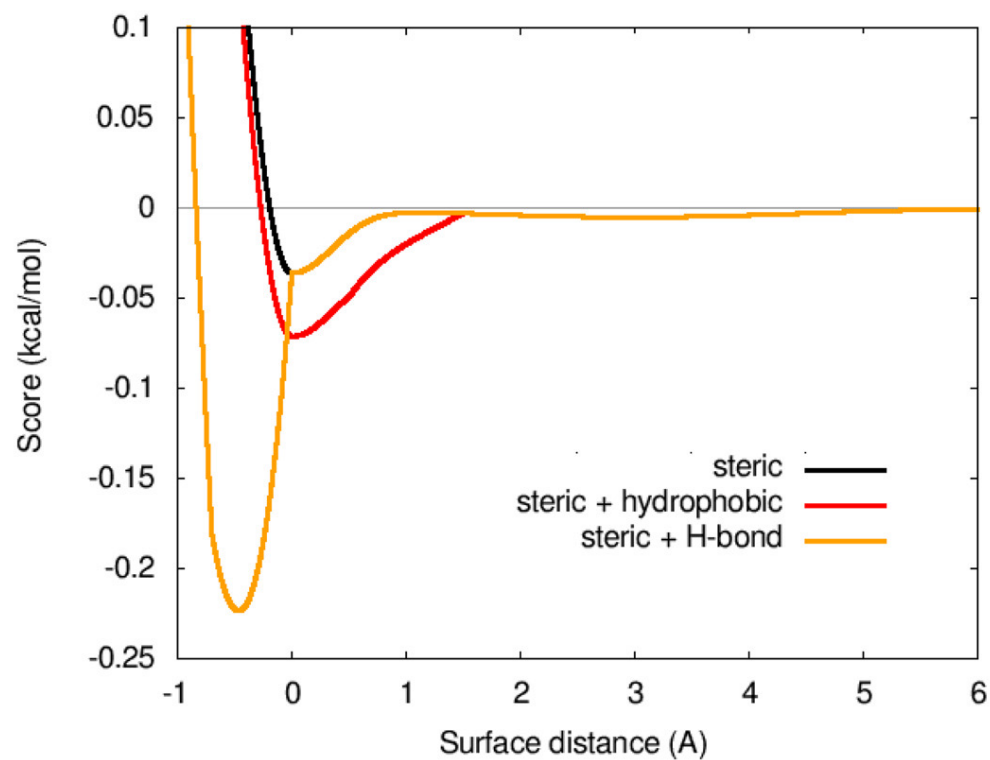
***10<sup>12</sup> plausible conformations!***



Simulated annealing



# Scoring functions



Scoring functions can be used beyond shape optimization to optimize ligand and protein interactions

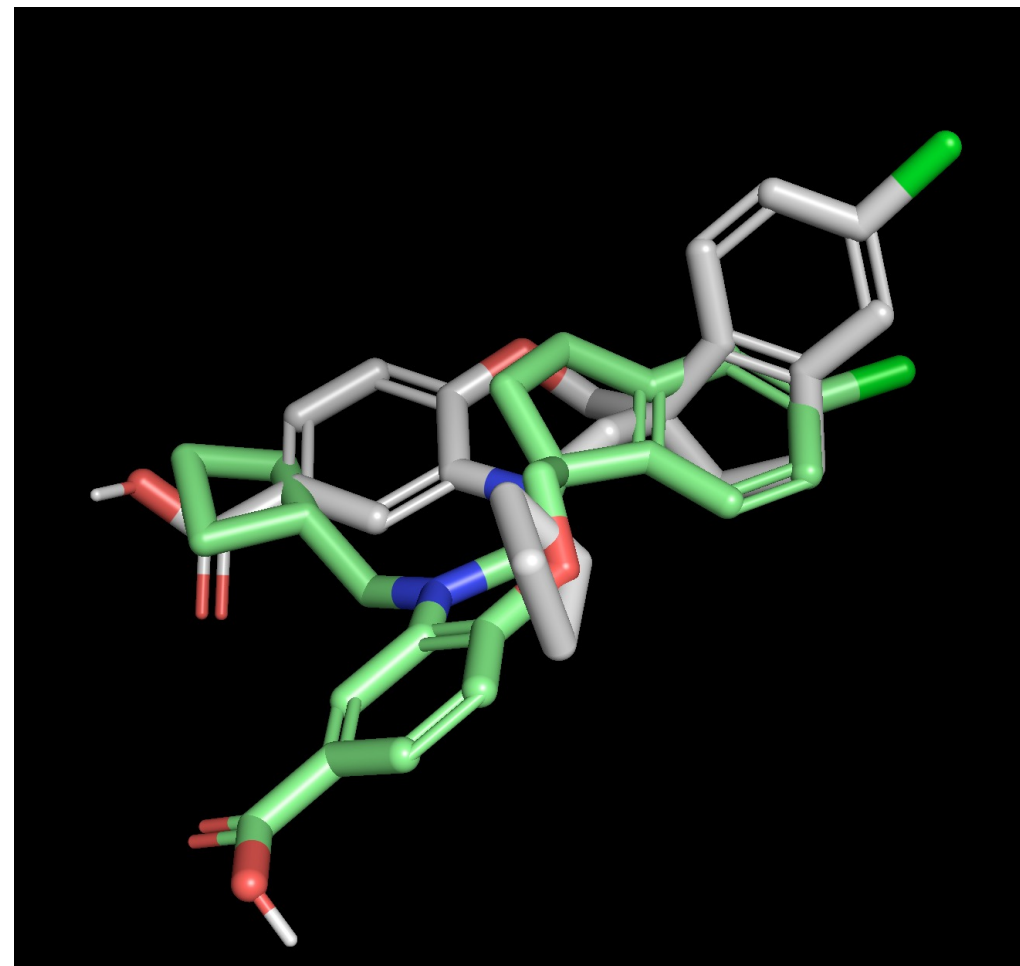
$$\Delta G = (V_{bonded}^{L-L} - V_{unbonded}^{L-L}) + (V_{bonded}^{R-R} - V_{unbonded}^{R-R}) + (V_{bonded}^{R-L} - V_{unbonded}^{R-L} + \Delta G_{conf})$$

$$V = W_{vdw} \sum_{i,j} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + W_{hbond} \sum_{i,j} E(t) \left( \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{elec} \sum_{i,j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} + W_{sol} \sum_{i,j} (S_i V_j + S_j V_i) e^{\frac{-r_{ij}^2}{2\sigma^2}}$$

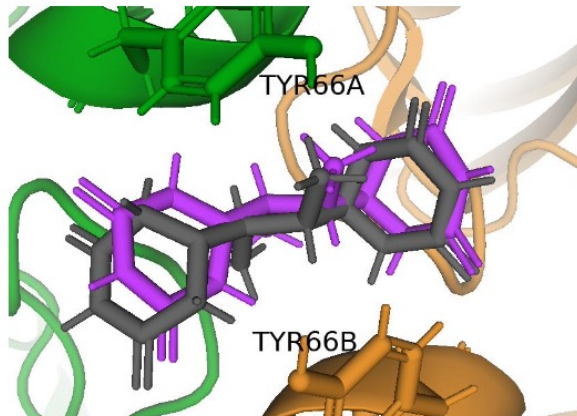
$$\Delta G_{conf} = W_{conf} N_{tors}$$

# Typical docking output generates multiple poses

mode	affinity (kcal/mol)	dist from best mode	
		rmsd l.b.	rmsd u.b.
1	-8.36	0	0
2	-8.08	2.899	6.789
3	-7.985	3.643	7.852
4	-7.914	3.415	5.21
5	-7.765	2.167	2.826

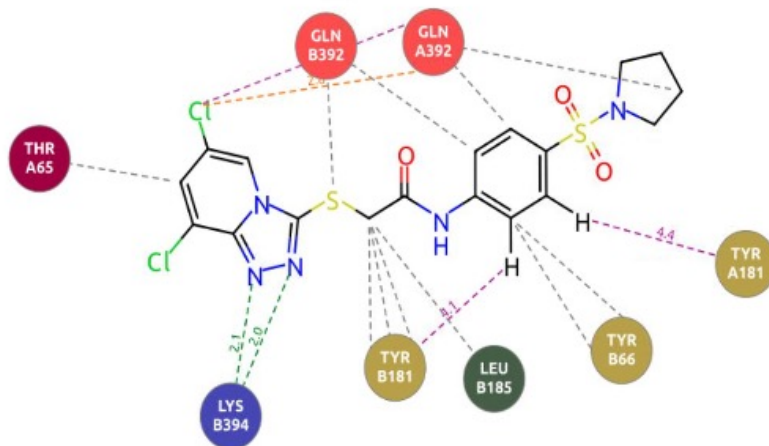
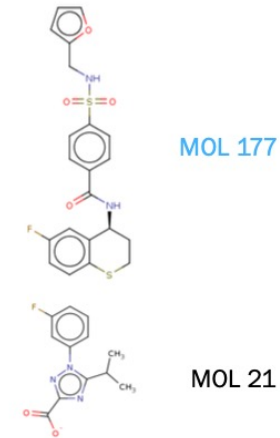
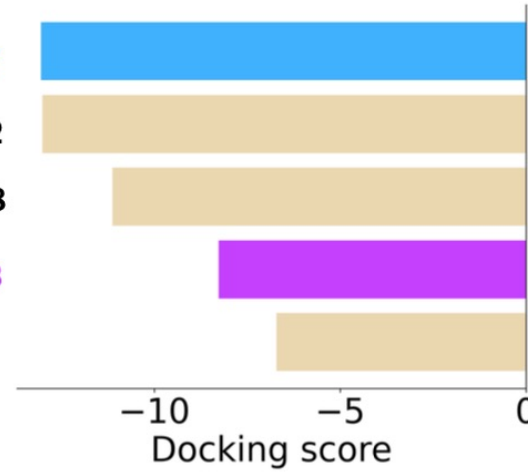


# Evaluating Docked structures

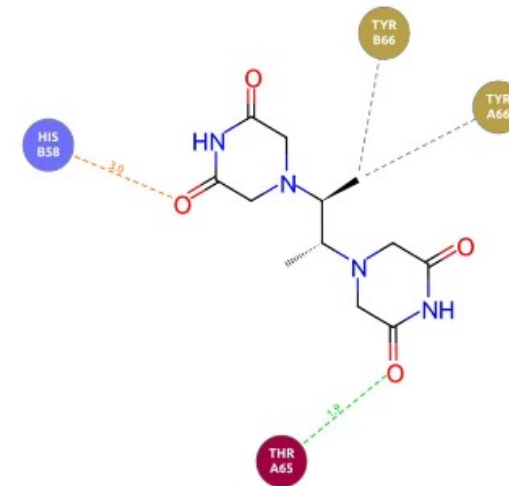


Crystal structure ICRF-193

MOL 177  
MOL 122  
MOL 108  
ICRF-193  
MOL 21



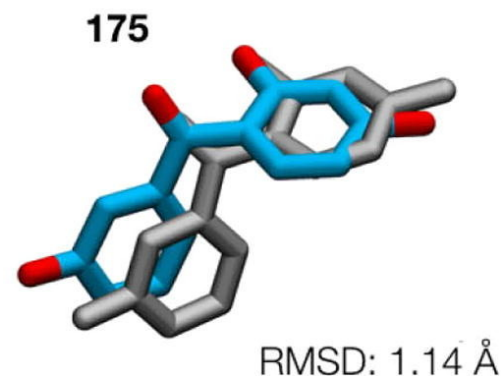
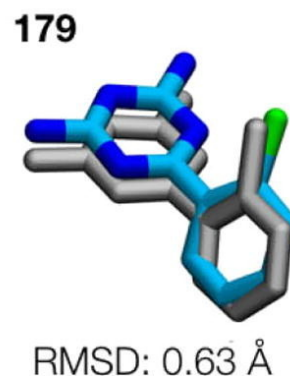
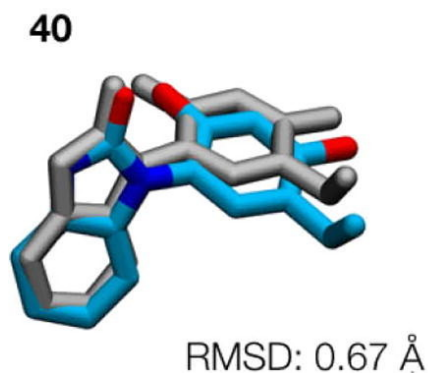
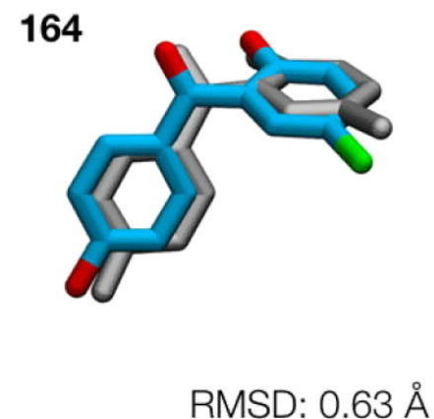
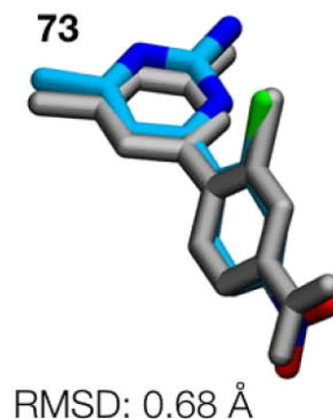
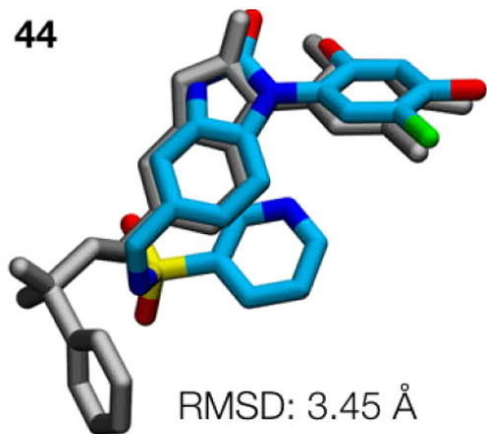
Contact map MOL-177



Contact map ICRF-193

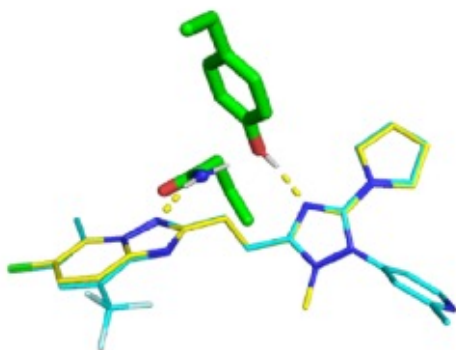


# Evaluating the binding mode/pose

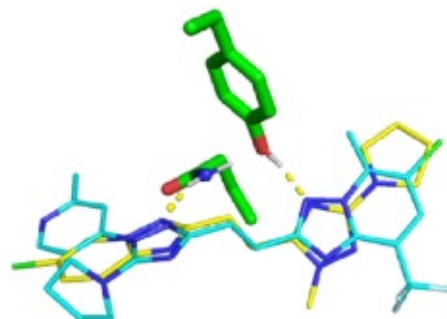


# Template docking and cross docking improves docking

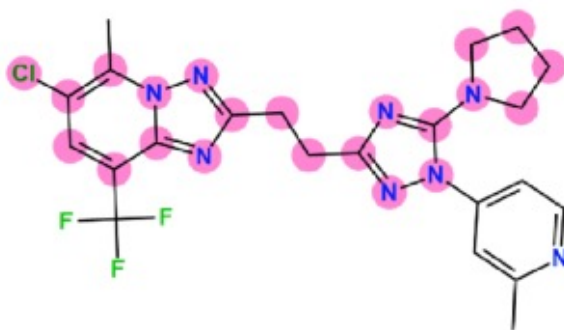
(a)



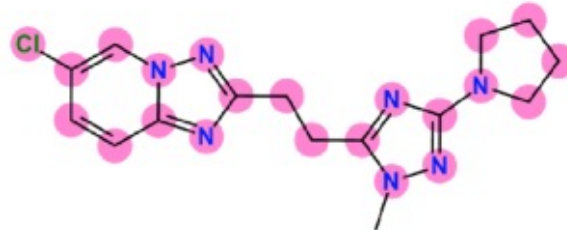
(b)



(c) Ligand compound 40:



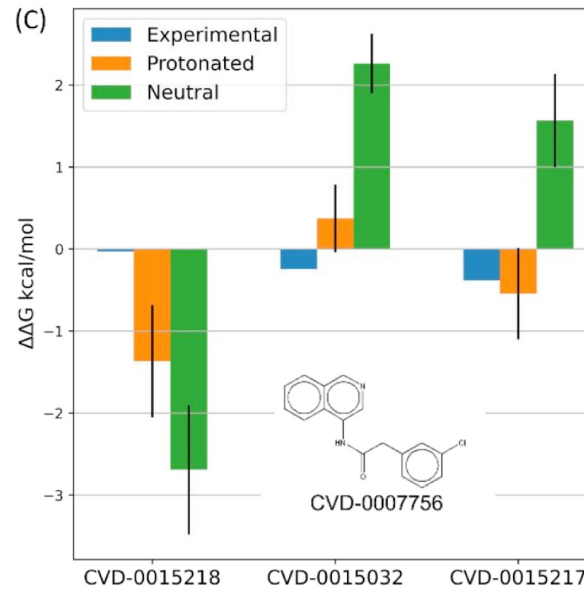
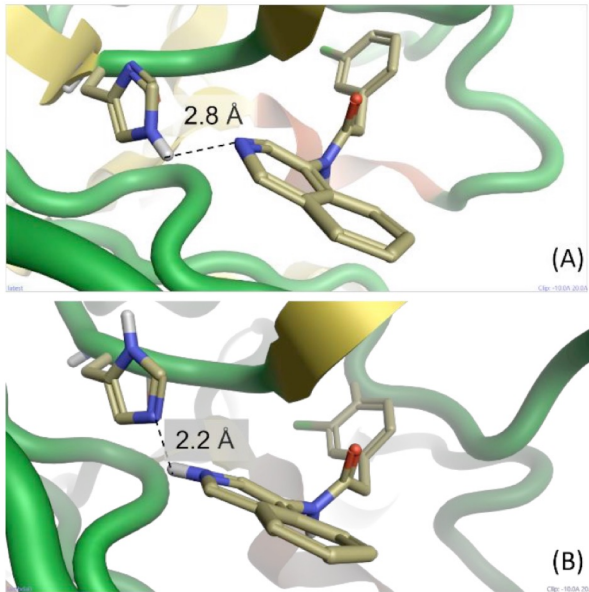
(d) Template compound PDB code 5sej:



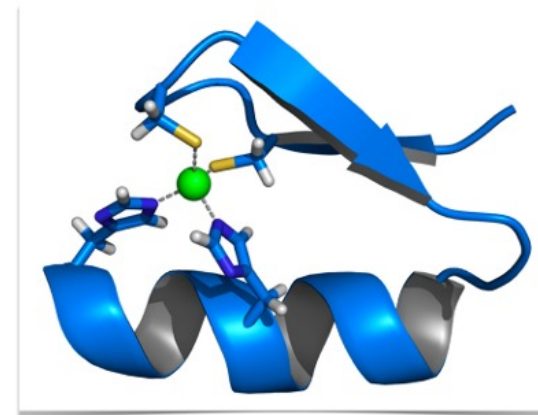
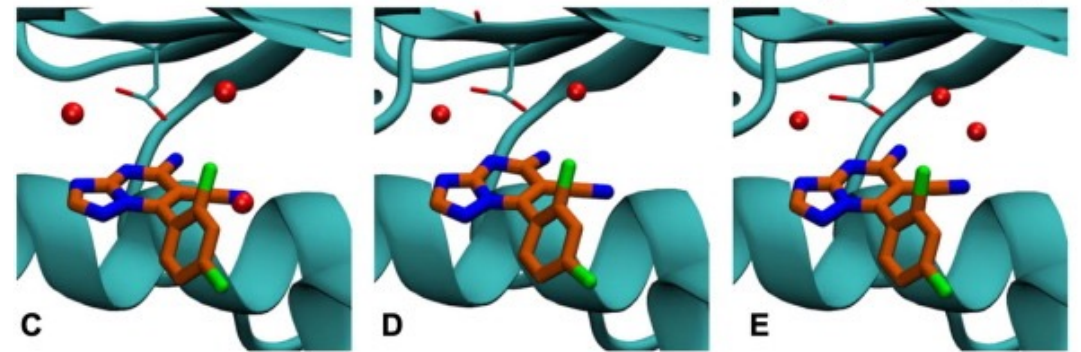
- Dock the same ligand into multiple protein structures (X-ray, MD)
- Generate multiple ligand conformers and dock into multiple structures
- Use existing ligand data as a template or guide e.g. through Maximum common substructure

# Things to worry about

pKa of ligands and binding site  
protonation need to be considered

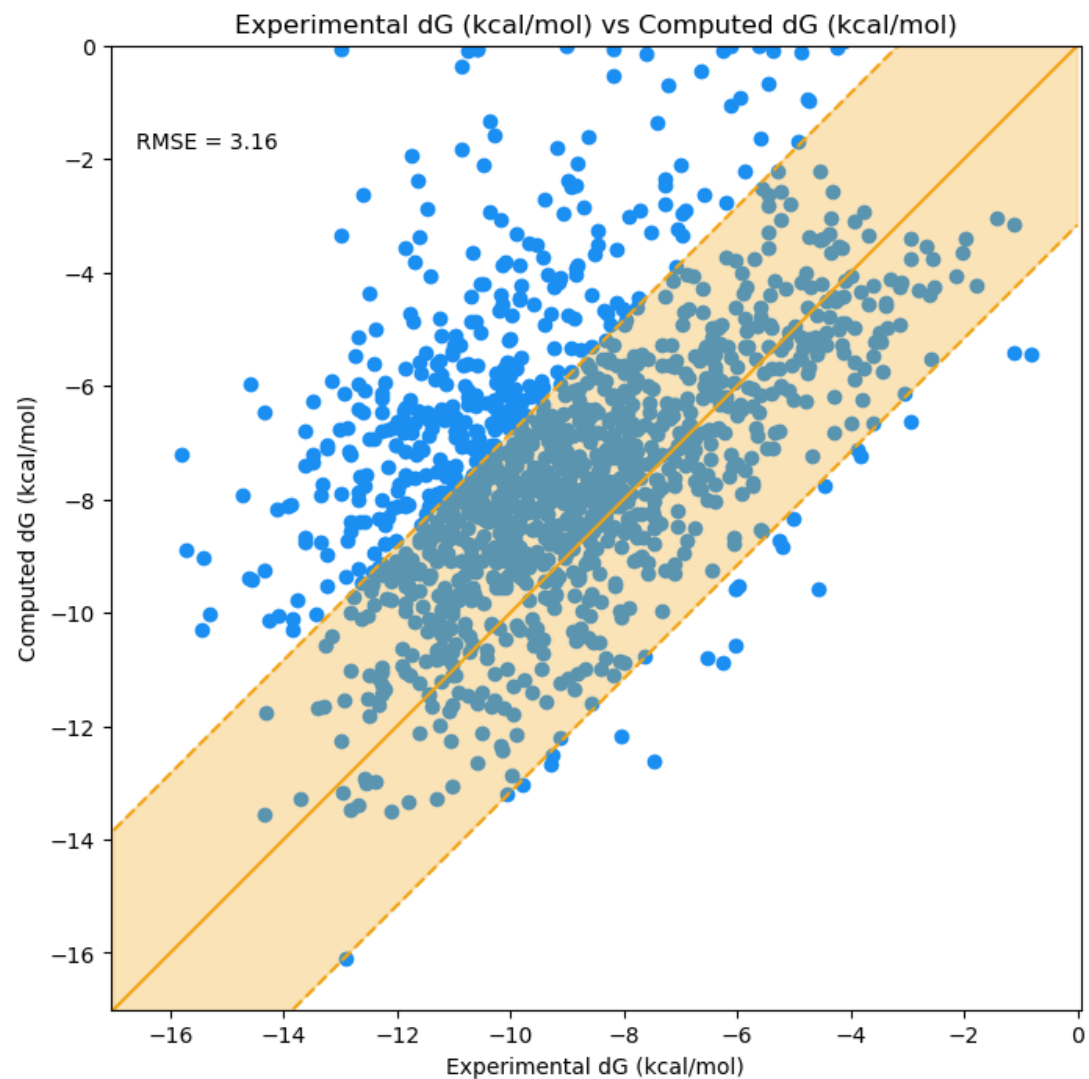


Structural waters are important

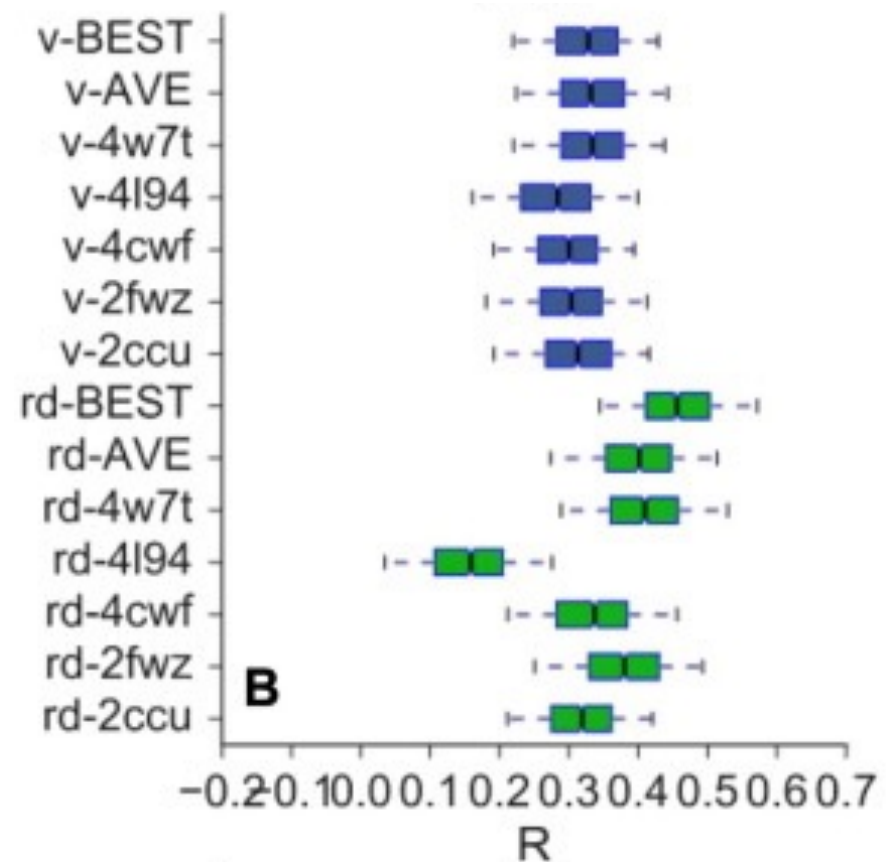


Co-factors such  
as ions and  
other molecules  
are important

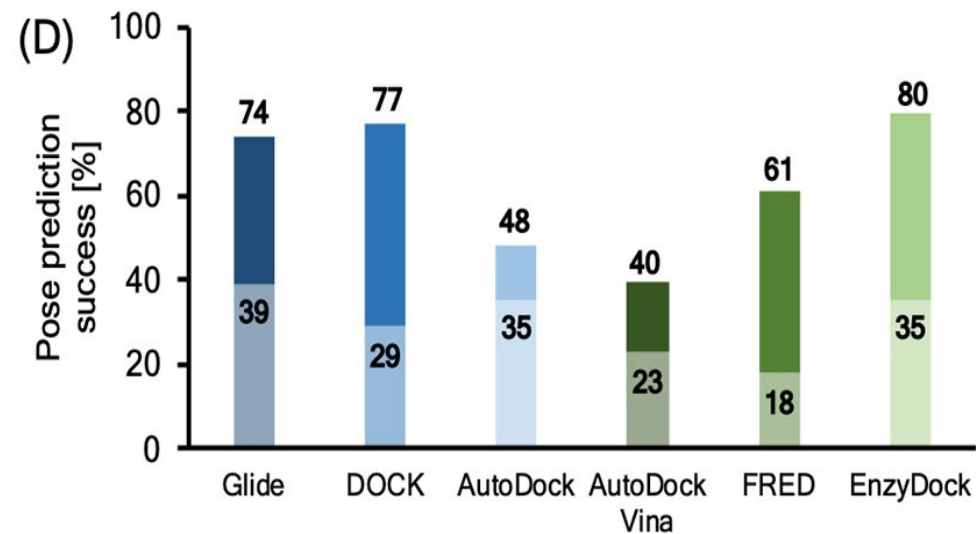
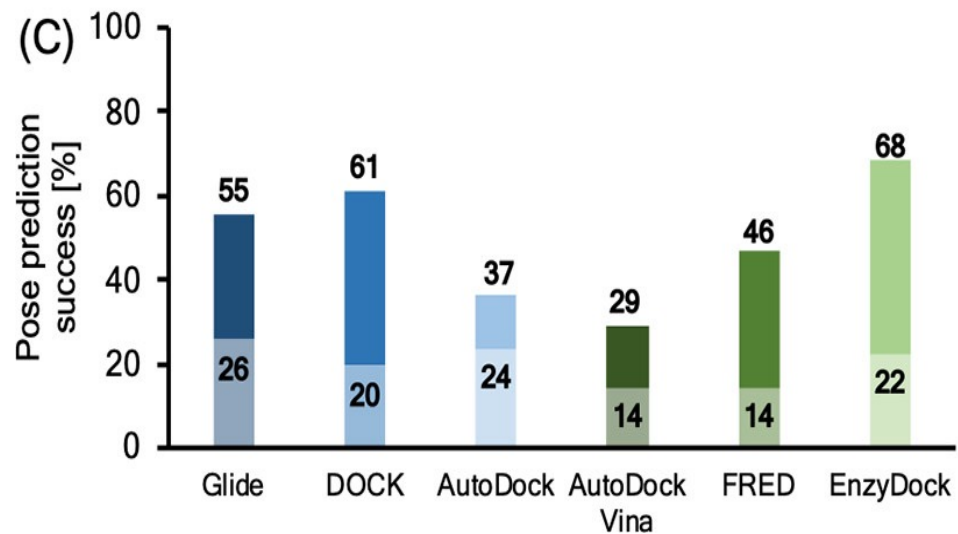
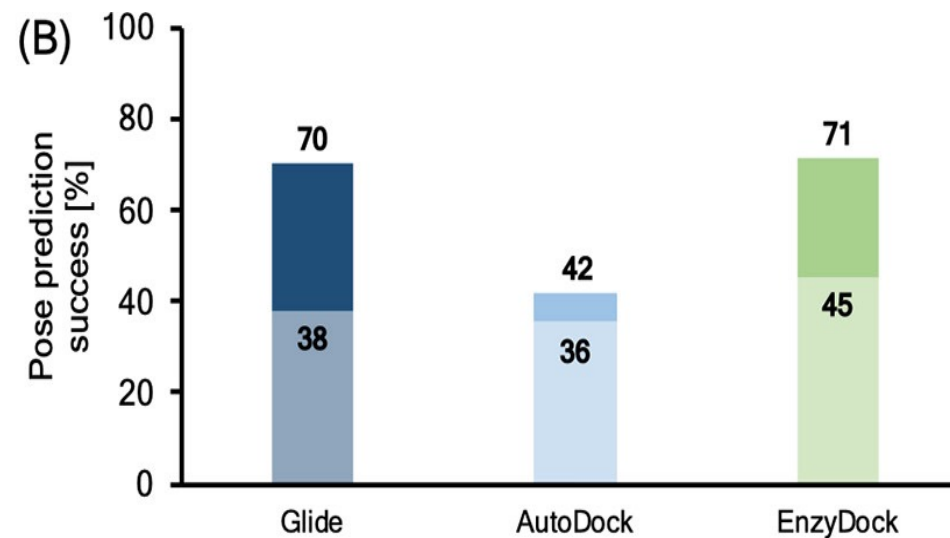
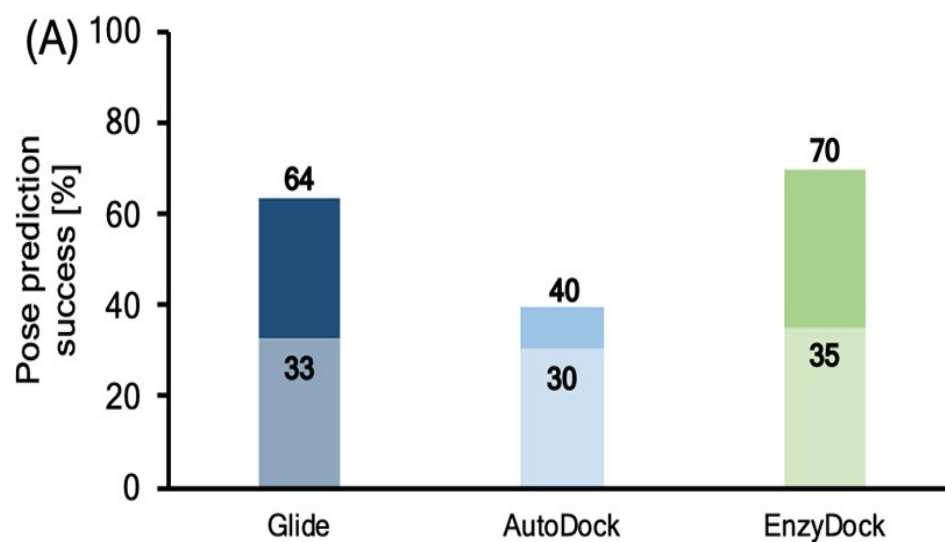
# Comparing against experimental $\Delta G$



Bulk assessment of correlation coefficient

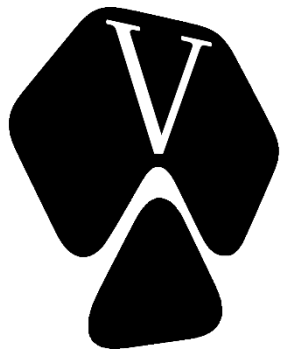


# Recent docking benchmark





# What tools exist for molecular docking?



Autodock Vina



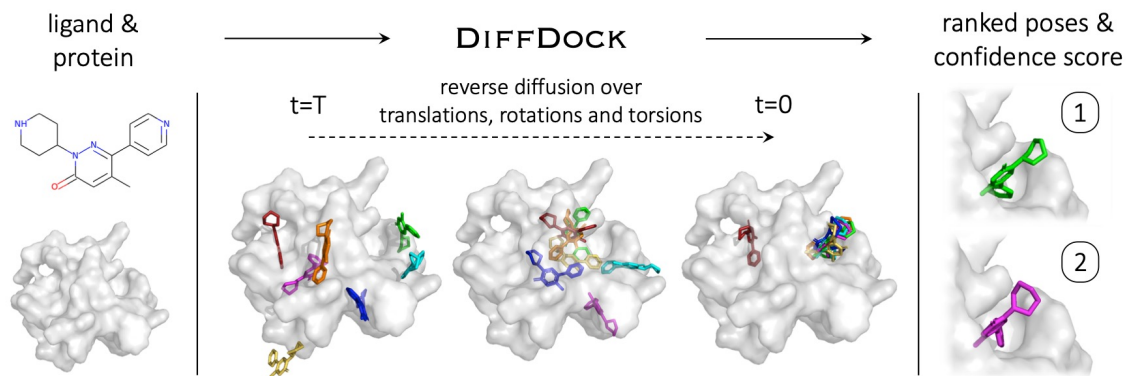
Glide



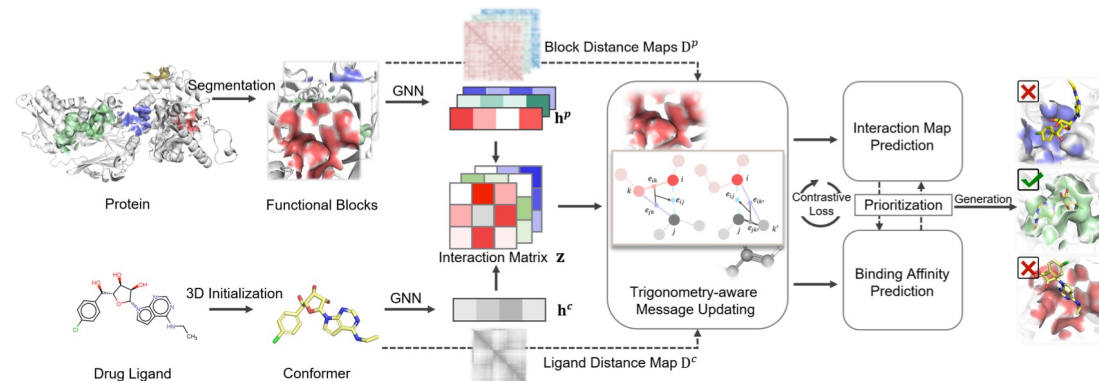
And others....

# ML-based docking

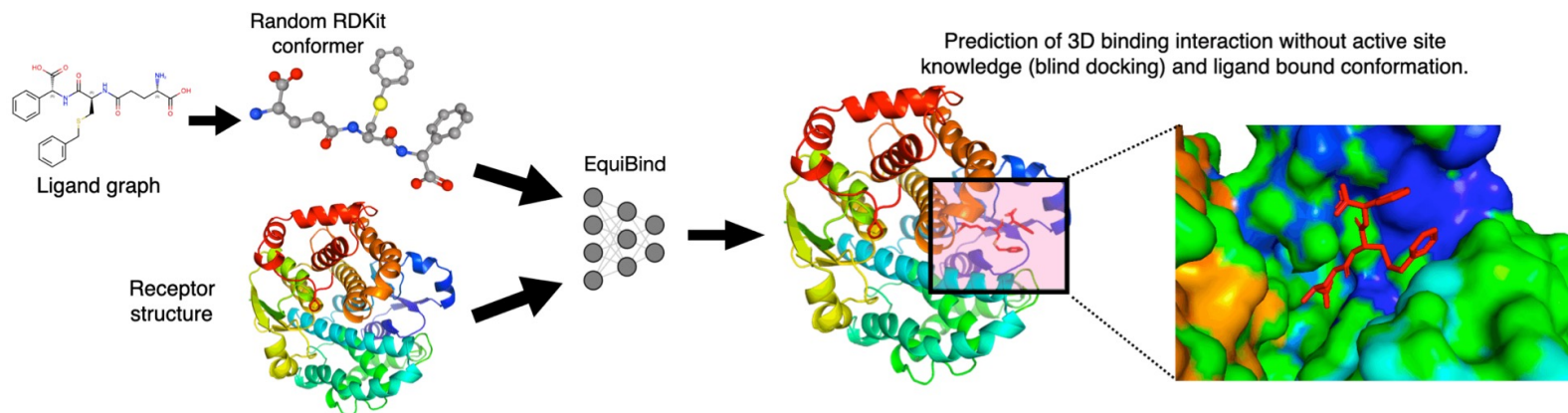
## Diffdock



## TankBind



## Equibind



## Gnina

## Deepdock

....

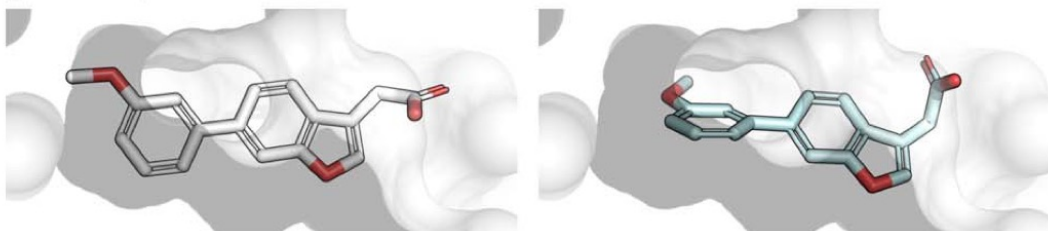
# Evaluating the binding mode/pose of ML tools



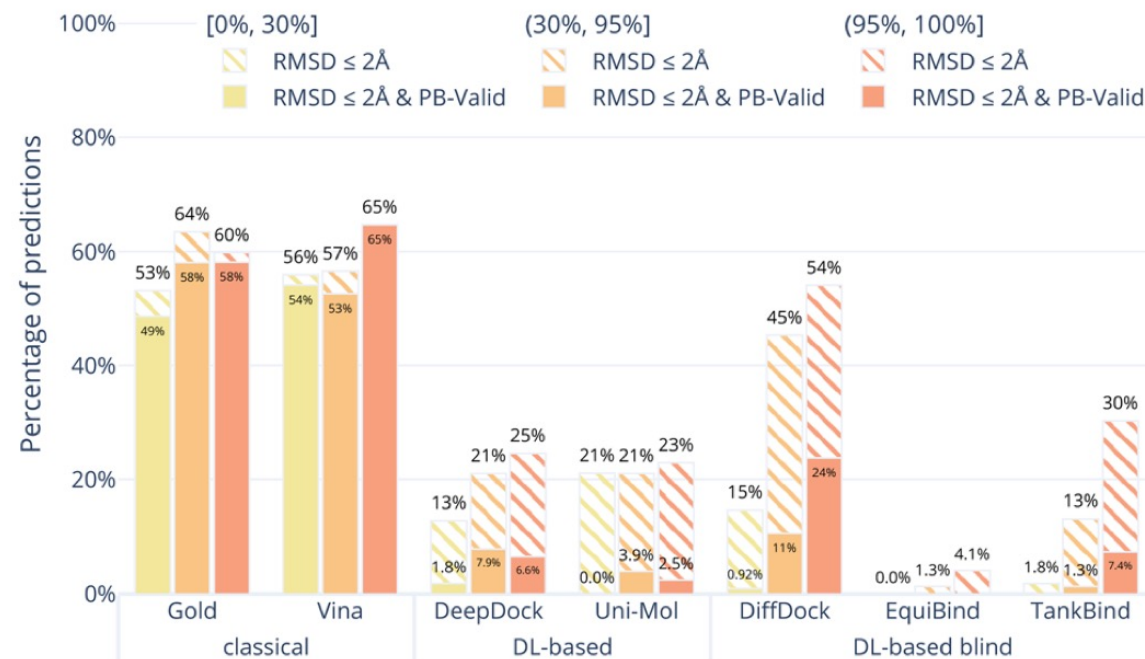
(f) Double bond not flat. TankBind prediction for ligand DBQ of protein-ligand complex 1U4D. RMSD 1.7 Å.



(g) Energy ratio too high. AutoDock Vina prediction for ligand IFM of protein-ligand complex 7LOU. RMSD 1.9 Å.



(h) Clash with protein. DiffDock prediction for ligand XQ1 of protein-ligand complex 7L7C. RMSD 1.6 Å.



Break Time!

