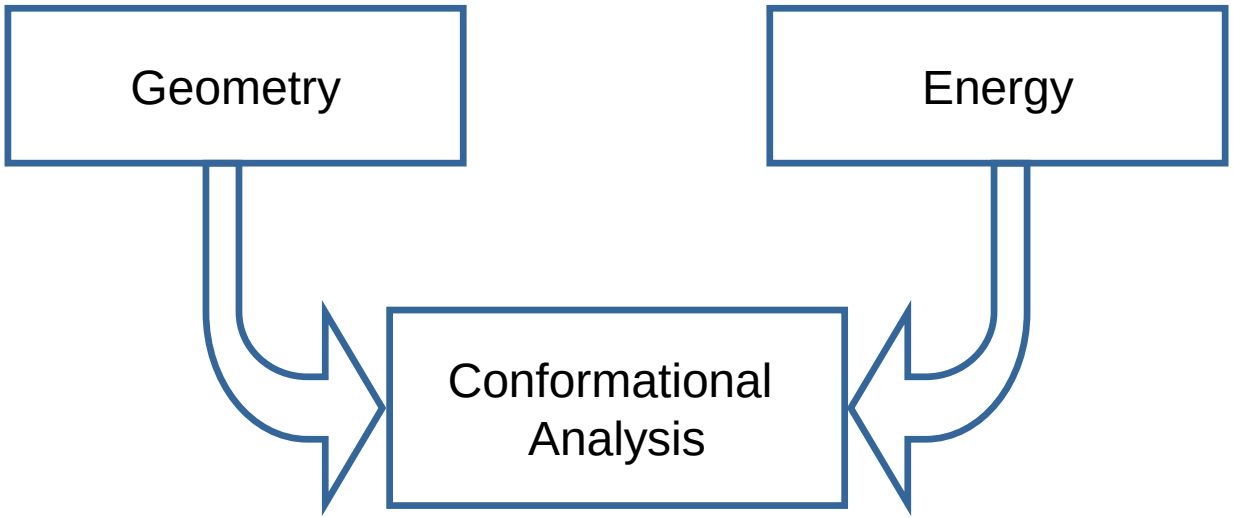
Conformational Search

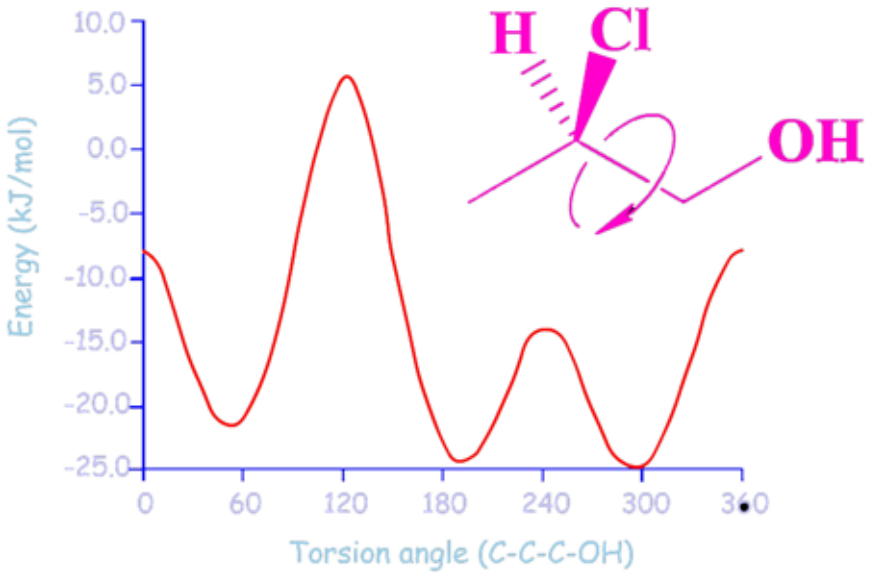
Conformational Analysis: Definition

Molecules can change their geometry by processes involving changes in torsion angles. The study of these geometries and their associated energies is called “conformational analysis”.



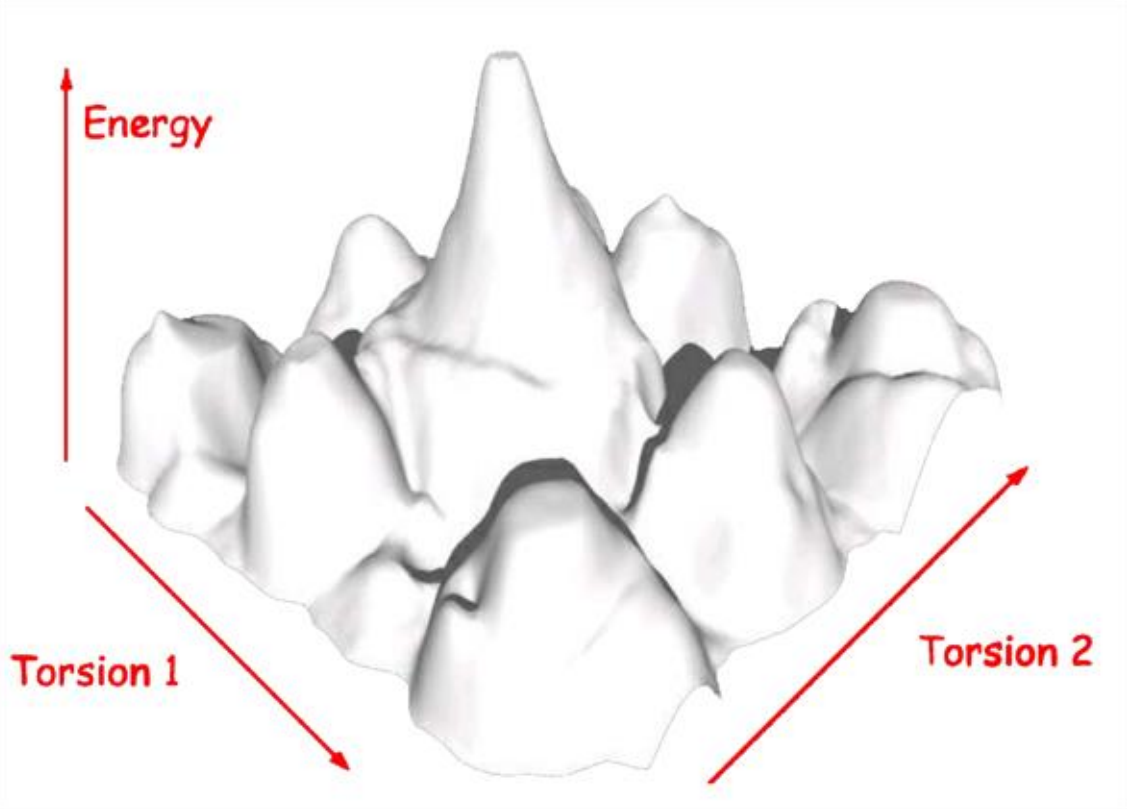
Conformational Potential Surface

In the case of a molecule with one rotatable bond, the "conformational potential surface" consists of the curve representing the molecular energy as a function of the torsion angle. The minima of the curve correspond to low energy conformations.



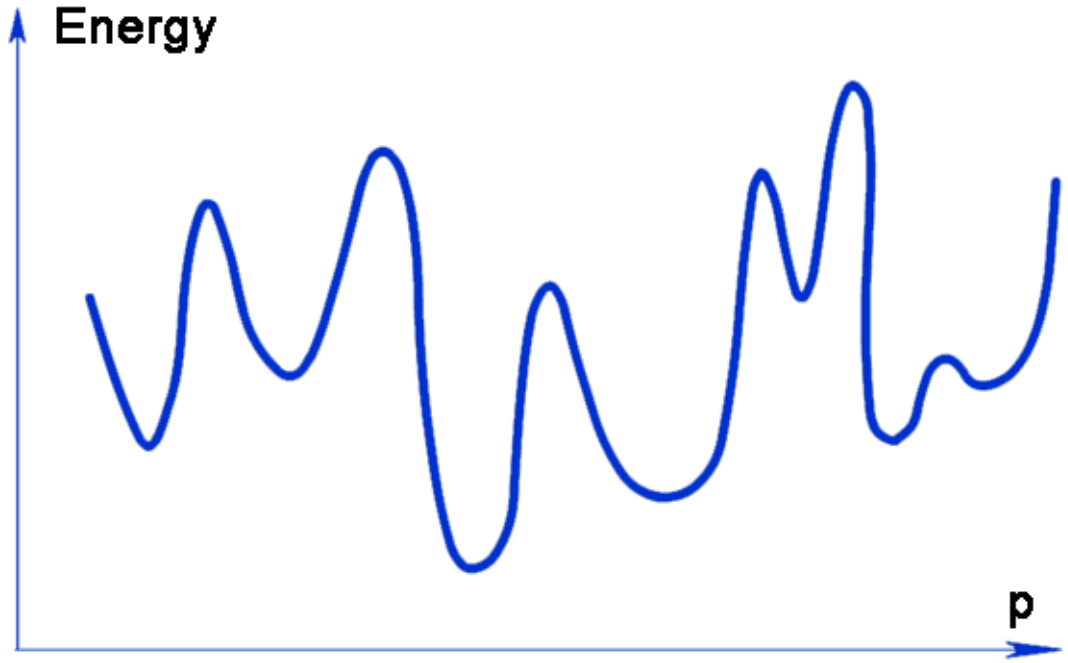
Conformational Potential Surface

In the case of a molecule with two rotatable bonds, the total energy is represented as a function of the two variable torsion angles.



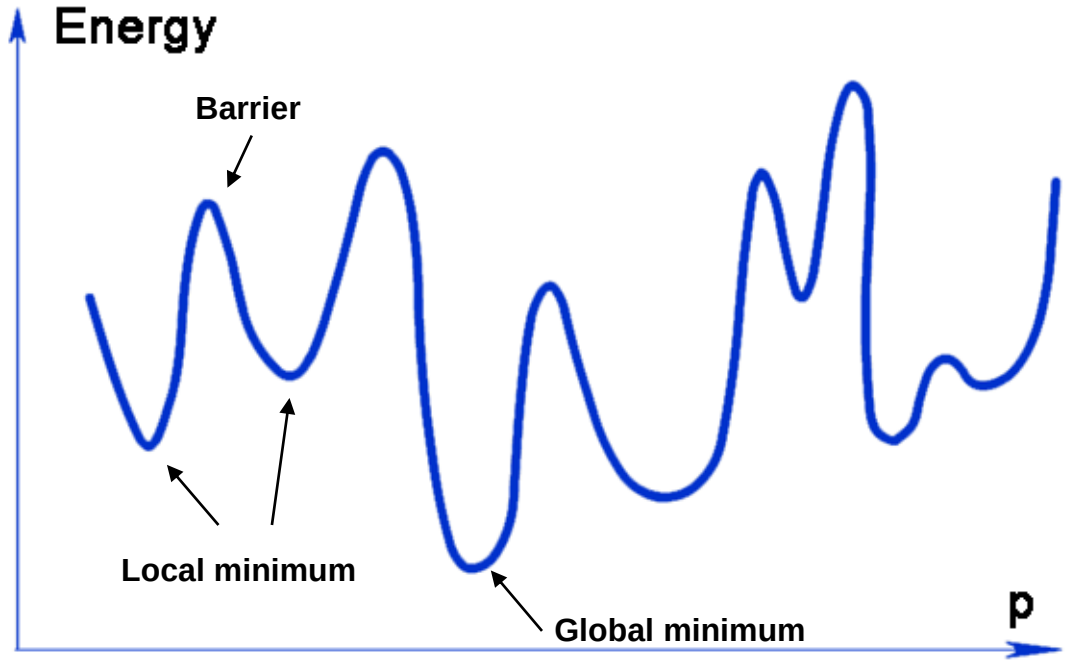
Conformational Potential Surface

Most molecule have more than two variable torsion angles. Therefore, their conformational potential surface cannot be visualized, because the surface is defined in a space with more than 3 dimensions. This surface can be represented schematically as a curve displaying the energy as a function of a parameter (p) that has no real geometrical meaning.



Conformational Potential Surface

The conformational potential surface contains “local minima” (low energy conformers), “barriers” (high energy conformers), and a “global minimum” (the conformation of lowest energy). When the barrier is high, the interconversion is either slow or impossible. For values greater than 80 kJ/mol the interconversion is not possible at room temperature.

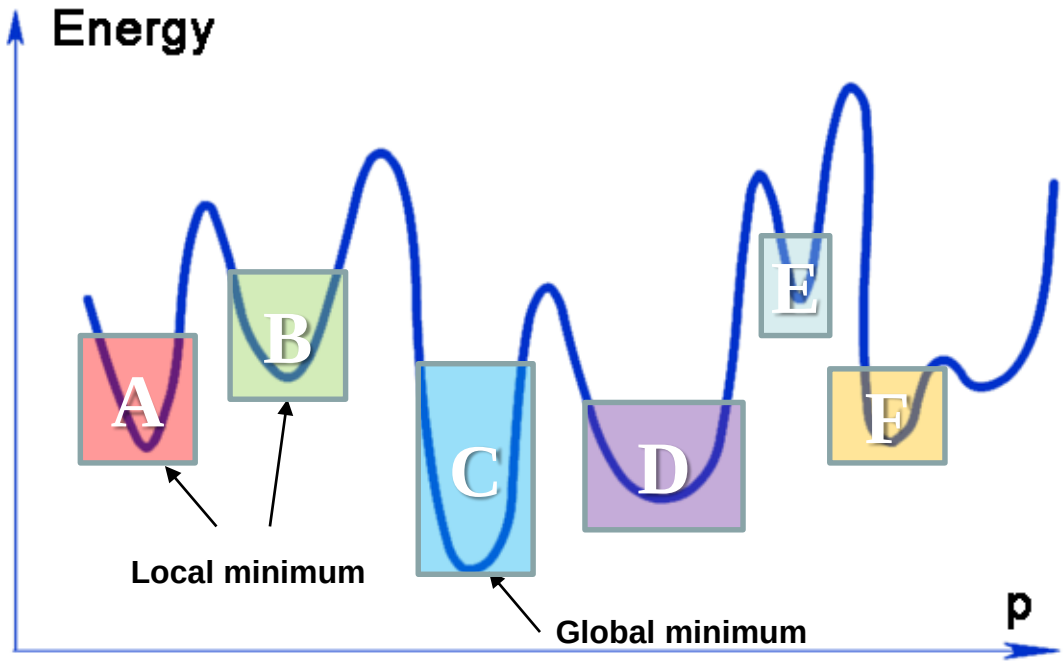


Conformational Analysis Principles

A simple method for exploring the conformational potential surface can be accomplished by a **systematic scanning** of all geometries of the molecule. However, this is impractical. For a molecule with 7 rotatable bonds and one non-planar ring, considering a 10° resolution for each rotatable bond (36 conformers per bond) and 4 conformations for the ring, the number of conformers is $36^7 \cdot 4$ (more than 300 billions).

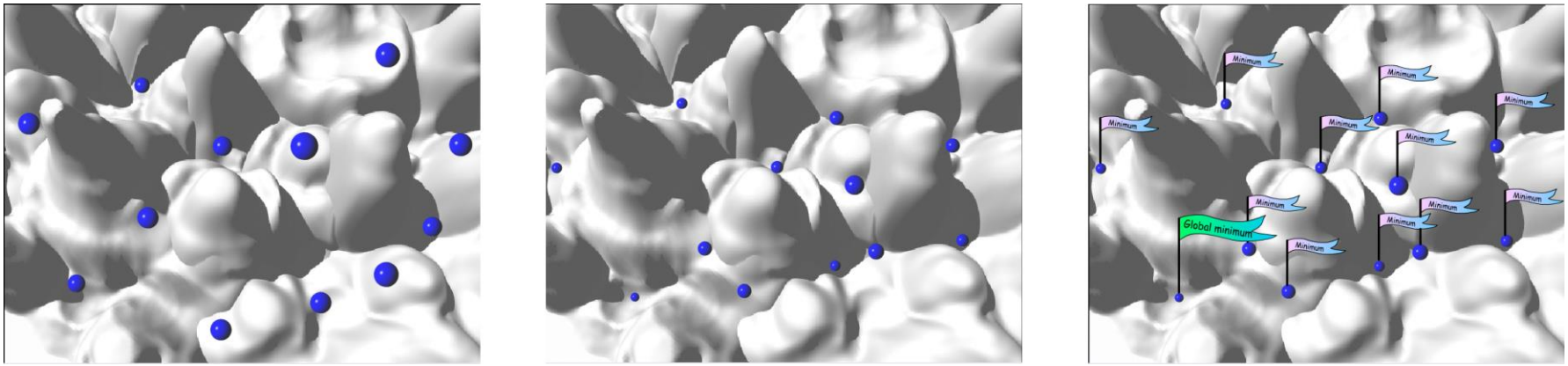
All geometries inside a given well correspond to a same family of conformers; therefore, conformational analysis can be reduced to the identification of all the families of conformers for the molecule object of the study (identifying the local minima by a subsequent minimization). With this rule, all the "typically different" conformers of the molecule are represented.

Generation of Representative Conformers



Conformational Analysis Principles: Summary

Usually it is not possible to systematically generate all the conformers of a molecule. It is necessary to consider a method that produces tipically different conformers. They can be generated either by the analysis of their fragments (sistematic search) or by random methods (e.g. Monte Carlo). Minimization procedures then have to be used on the generated conformers for obtaining meaningful energy values.



Conformational Analysis → **Minimization** → **Results**

Conformational Search Algorithms

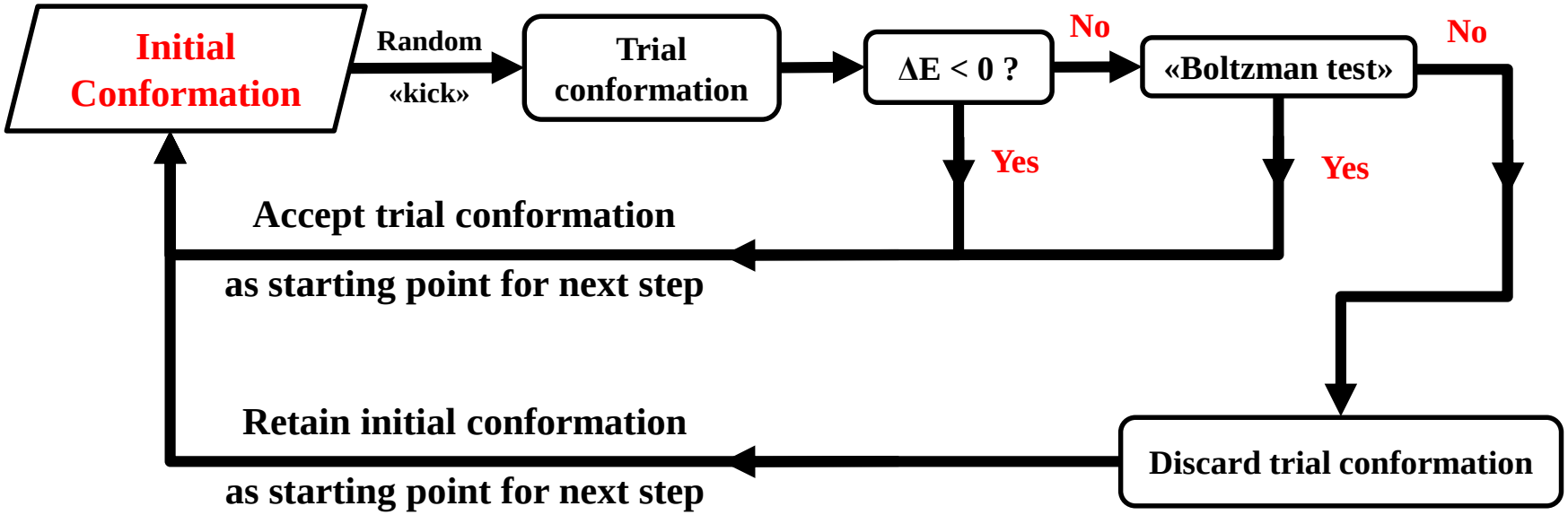
SYSTEMATIC. It uses a systematic method to explore the majority of conformations of a molecule. For acyclic molecules, the algorithm rotates each bond by a specified angle (usually 120 degrees) and searches for minima. This typically spans the conformational space effectively enough to find all conformations of small molecules. For cycles, a similar rotation is used to sequentially bend rings within the molecule. For large molecules, this method is time consuming, as the number of conformers searched grows exponentially. For this reason, SYSTEMATIC search should only be used for small, preferably acyclic, molecules.

Conformational Search Algorithms

MONTE-CARLO. The name "Monte Carlo" comes from the random-chance nature of the simulations, akin to the games of chance at Monaco's gambling resort. Rather than being a deterministic method, Monte Carlo algorithms are stochastic and use random numbers. An extremely important Monte Carlo algorithm for molecular systems was developed by Metropolis *et al.* The method starts from a given geometry, and new configurations are generated by adding a random “kick” to one or more atoms. The new geometry is accepted as a starting point for the next perturbing step if it is lower in energy than the current. Otherwise, the Boltzmann factor is calculated and compared to a random number R between 0 and 1. If the factor is higher than this number, the new geometry is accepted, otherwise the next step is taken from the old geometry. This generates a sequence of configurations from which geometries may be selected for subsequent minimization.

Conformational Search Algorithms

Illustrative scheme of Metropolis-Hastings MC Algorithm



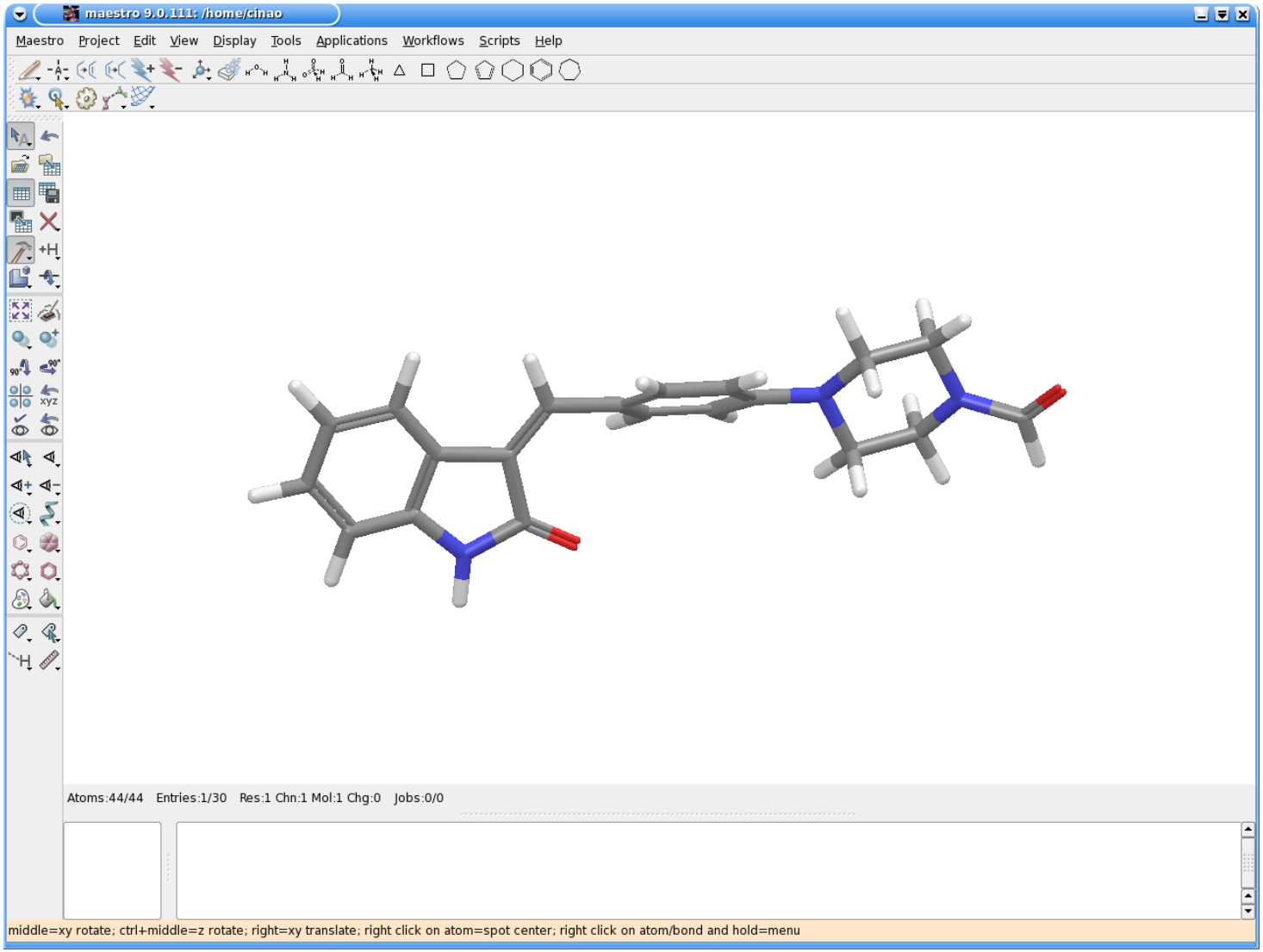
«Boltzmann test»: for $R_{[0,1]}$, if $e^{-\Delta E/kT} > R$ the conformation is accepted

Bioactive Conformation

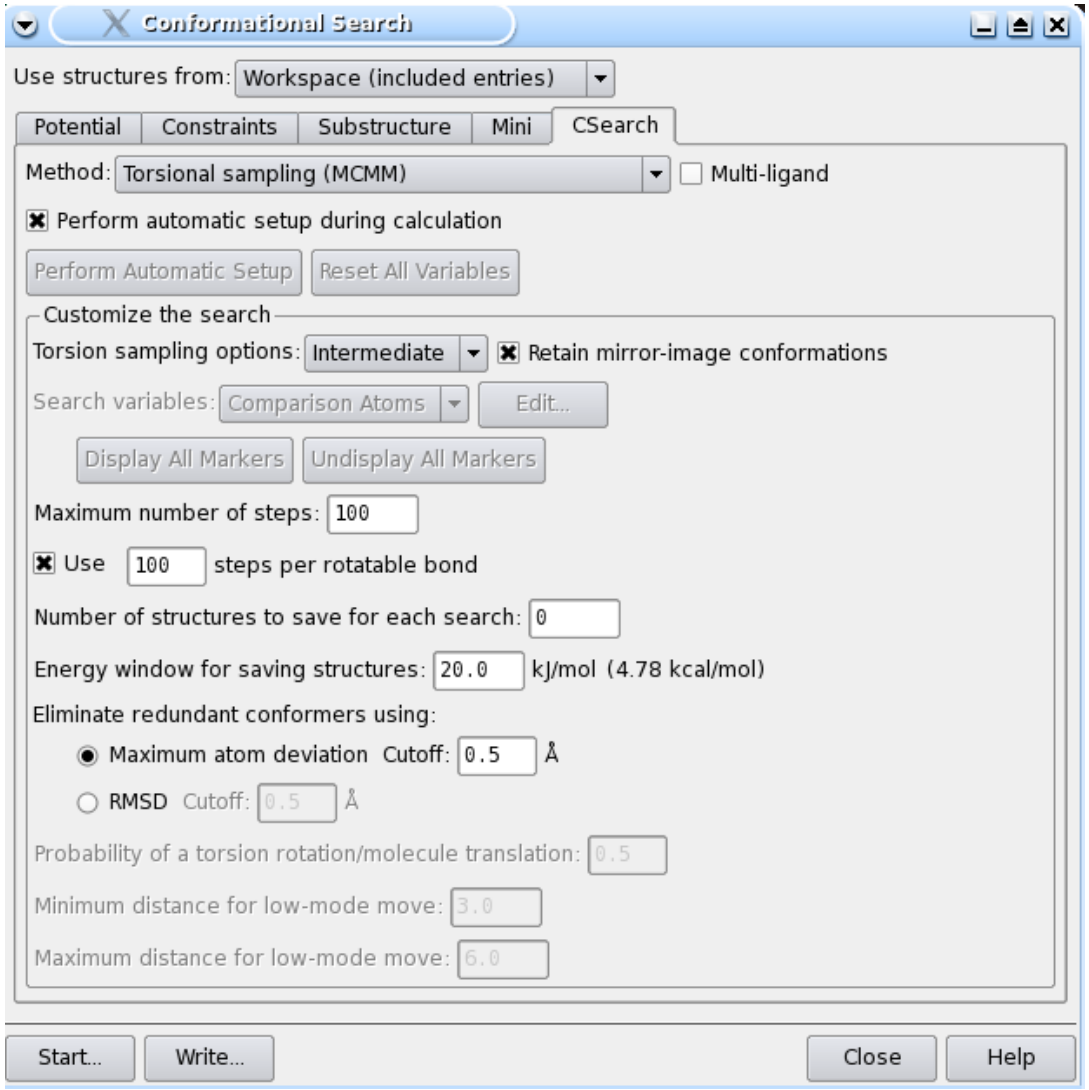
The geometry adopted by a molecule when it binds to its biological target is called «bioactive conformation». It does not necessary correspond to the global minimum of a molecule (it can even be very different). In fact, a conformation less stable than that associated to the global minimum can be more suitable for establishing intermolecular interactions with the target receptor. In this case, the energetic gain required for assuming a more strained conformation is compensated by the energy released through favourable interactions.

However, in order to be able to assess the population of the bioactive form, it is necessary to known the energy of the global minimum of the molecule concerned, so that populations can be derived based on Boltzmann's distribution. The difference of energy between the bioactive conformation and the global minimum is generally less than 13 kJ/mol.

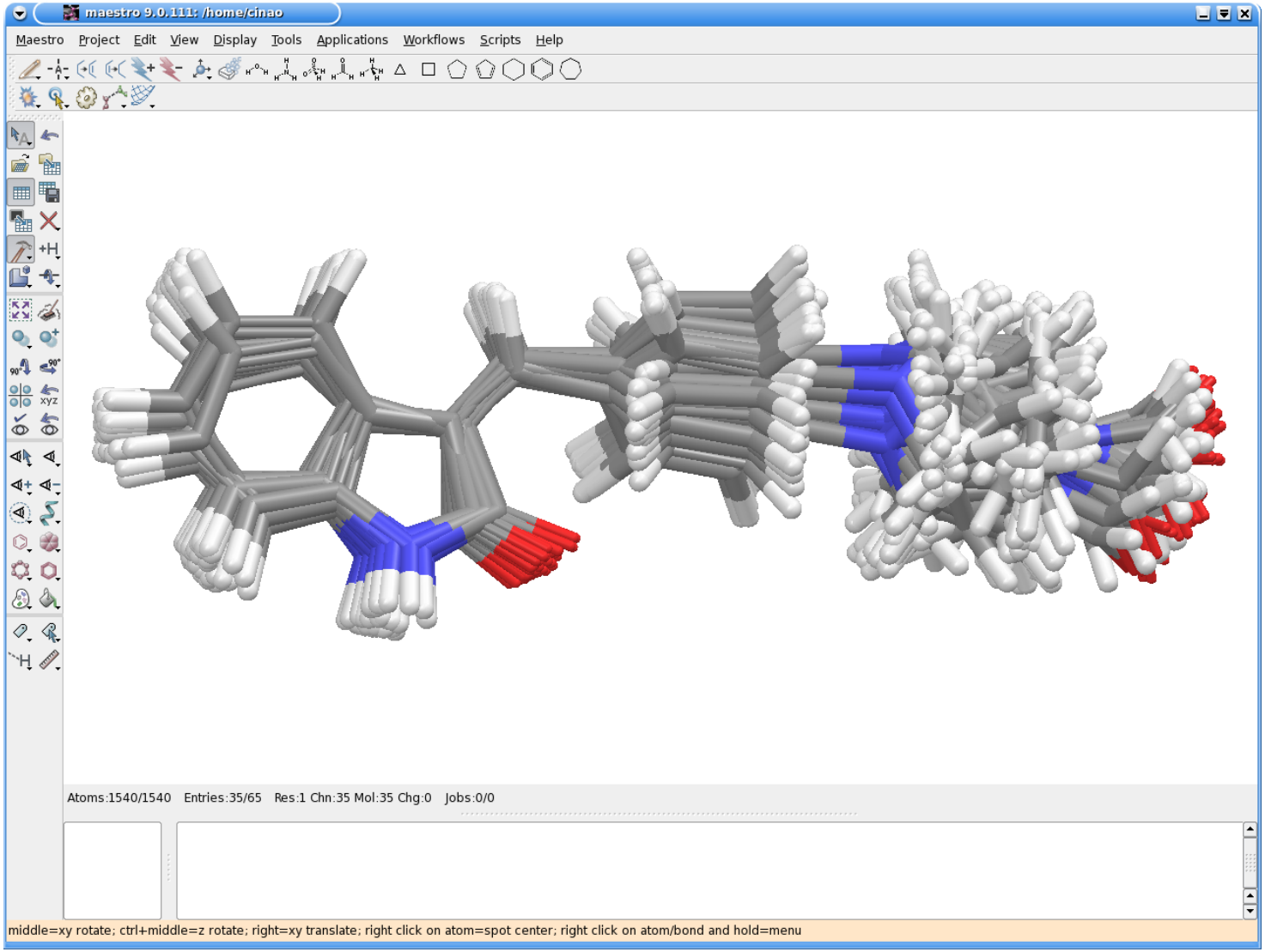
Conformational Search example



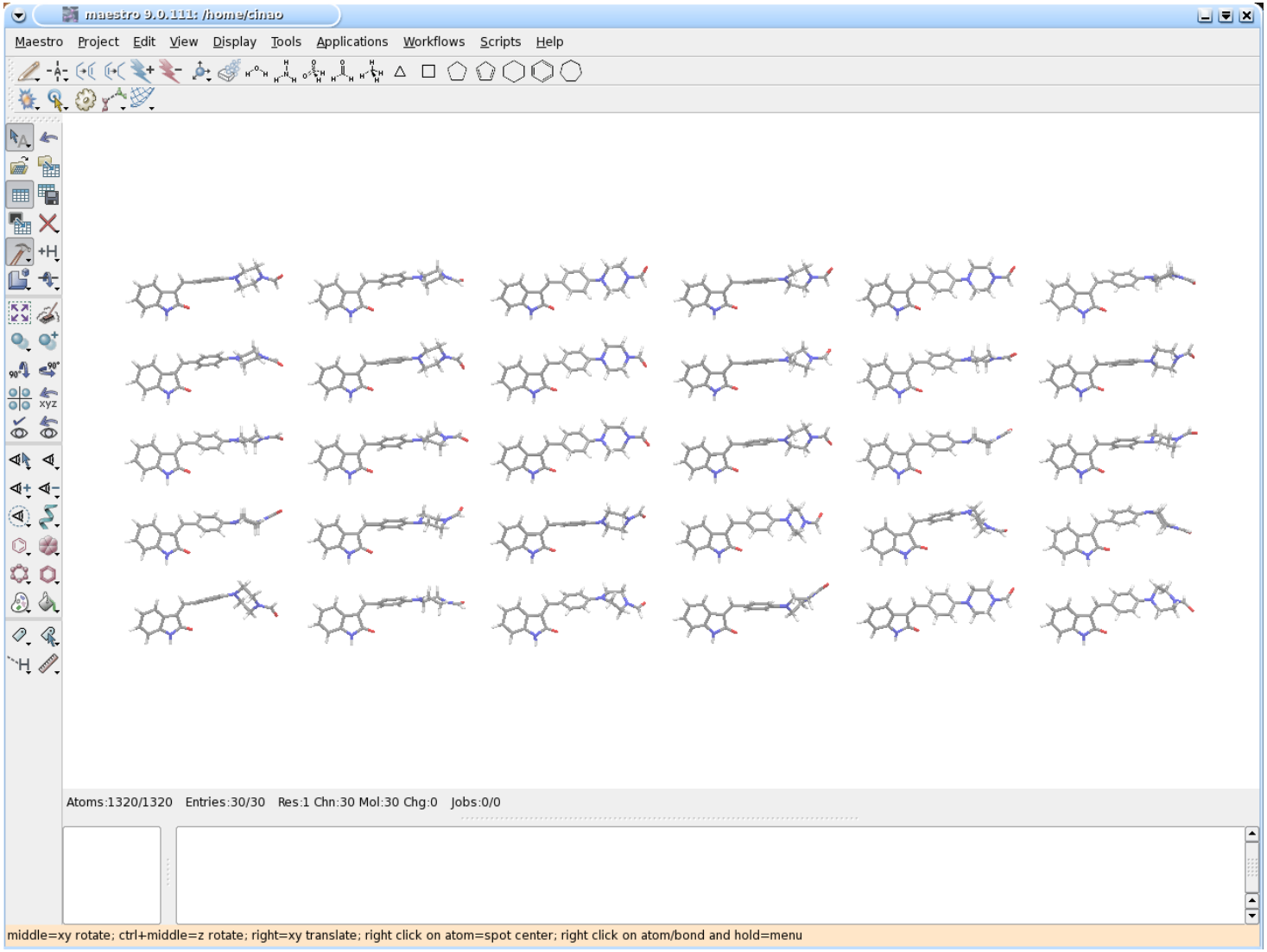
Conformational Search example



Conformational Search example



Conformational Search example



Receptor Based Drug Design

Receptor Based Drug Design

Molecular recognition in biological systems relies on the existence of specific attractive interactions between two partner molecules. Receptor-based drug design seeks to identify and optimize such interactions between ligands and their host molecules, typically proteins, given their 3D structures.

The aim of the Receptor Based Drug Design is to create or identify new molecules with high affinity and/or selectivity for a certain macromolecular receptor whose 3D structure is known. The molecules are conceived based on their 3D complementary features with the macromolecular target, whose structure is thus exploited for the design. In addition, one can exploit the molecular interactions to strengthen the binding of a known prototype ligand (inhibitor, agonist, or antagonist) into the active site of the receptor.

3D Receptor data

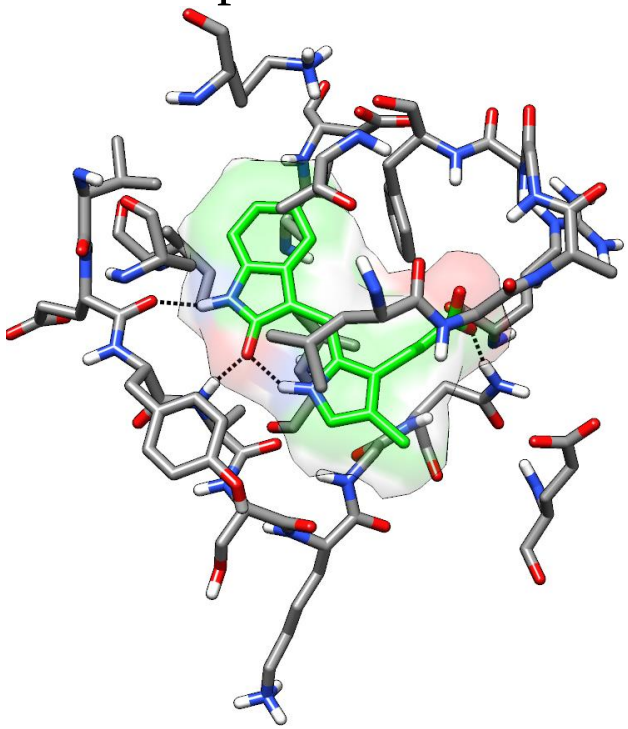
The most common method to determine the 3D structure of macromolecules is X-ray crystallography. This technique requires the isolation, purification and crystallization of the protein, which is not always feasible. The results are usually deposited in the Brookhaven Protein databank (PDB), which has more than 150 thousand protein X-ray structures. NMR techniques and electron diffraction have recently been used.

In the absence of a detailed 3D experimental structure, models of the target may be used as surrogates. For example, if the primary sequence of a certain protein is known and it shares a certain degree of sequence homology with one or more proteins for which a detailed structural information is available, it is possible to construct a model of the target protein using homology modeling techniques.

In absence of a good template, *ab initio* methods can also be exploited.

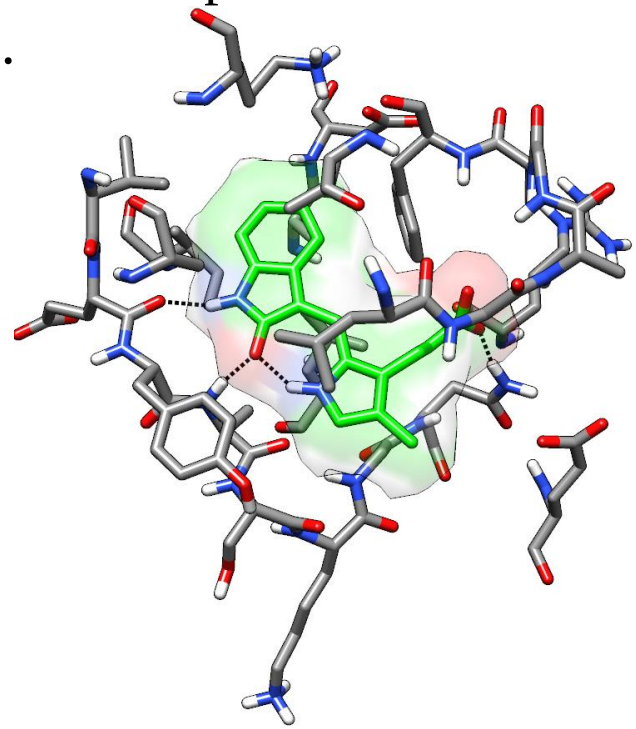
Definition of Docking

Docking is an energy-based operation for exploring the binding modes of two interacting molecules. A docking procedure is used as a guide to identify the preferred orientation of a ligand interacting with a macromolecule. The ligand can accomodate small conformational changes to avoid steric repulsions and produce favorable interactions with the receptor.



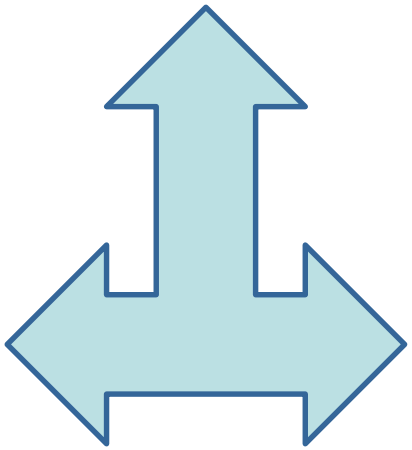
Definition of Docking

Docking procedures are very useful for assessing the quality of candidate ligand prototypes. If the ligand is a known active molecule, docking simulations help us identify its bioactive conformation. If the ligand is a candidate molecule, the docking allows us to analyze how it fits to the receptor. The best-docked compounds can be used as leads for further design and optimization.



Definition of Docking

DOCKING



MANUAL

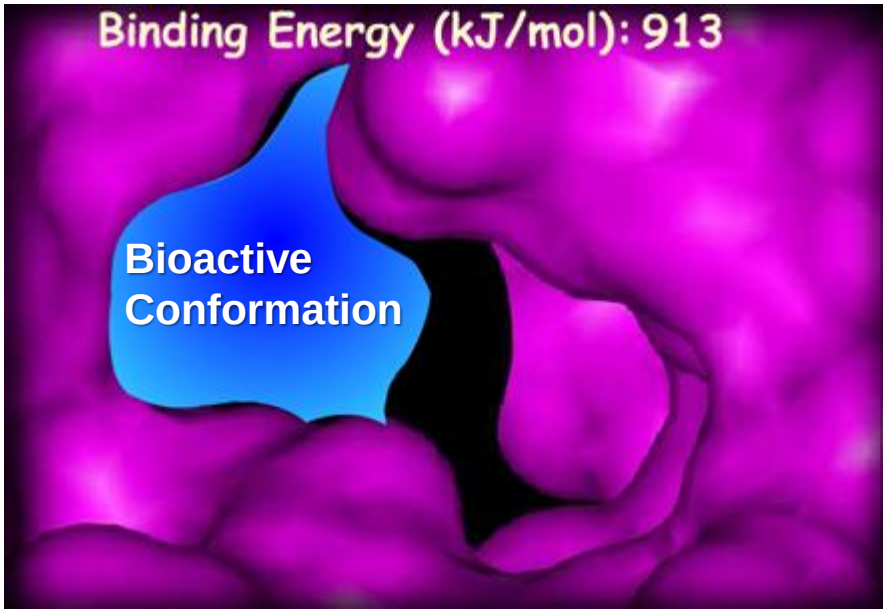
- It is necessary to have a binding hypothesis.
- It needs molecular mechanics calculations.
- It is possible only for hundreds of compounds.

AUTOMATED

- The ligand is positioned in an automatic way.
- It is possible for millions of compounds.
- It is necessary to evaluate the results.

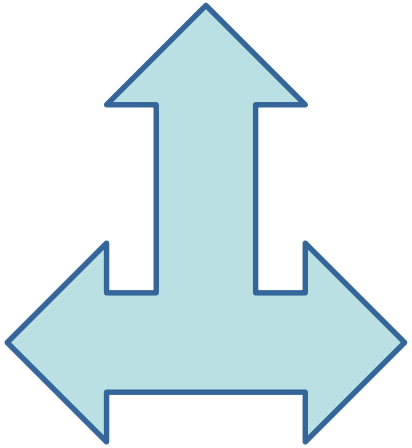
Definition of Automated Docking

Given the 3D structure of a protein target, compounds can be designed to fit into a cavity, which is called «docking». Starting from an approximate orientation of a ligand into the active site, the docking process modifies its position and conformation in order to have a maximum of favorable intermolecular interactions. The treatment ends when a minimum of energy is obtained for the complex.



Definition of Docking

AUTOMATED DOCKING



EXHAUSTIVE

- It needs a large number of ligand conformers.
- Es. FRED

STOCHASTIC

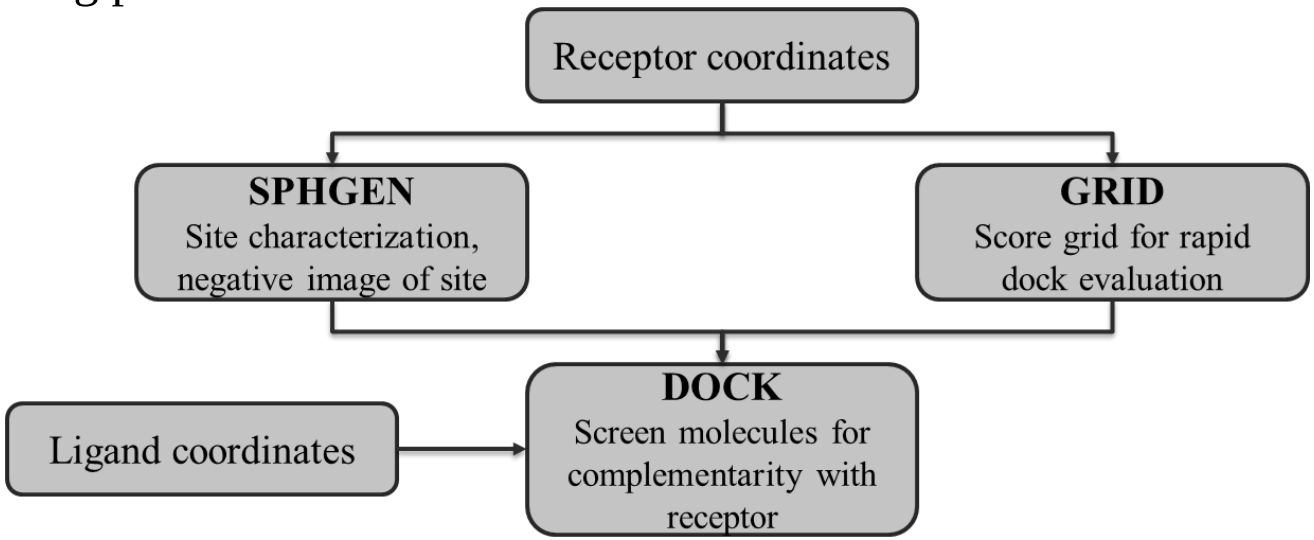
- Only some solutions are evaluated.
- The results may not be repeatable.
- Es. DOCK, GOLD, FlexX, AUTODOCK.

Main Docking Software

The main docking software performs flexible ligand docking, while treating the receptor as a rigid entity.

DOCK and FlexX

The DOCK algorithm automatically generates many possible orientations and conformations of a putative ligand within a receptor pocket. The shape of the receptor pocket is described by spheres, and the centers of the spheres are considered as potential locations for ligand atoms. A scoring function is used to evaluate the quality of the docking poses.



DOCK

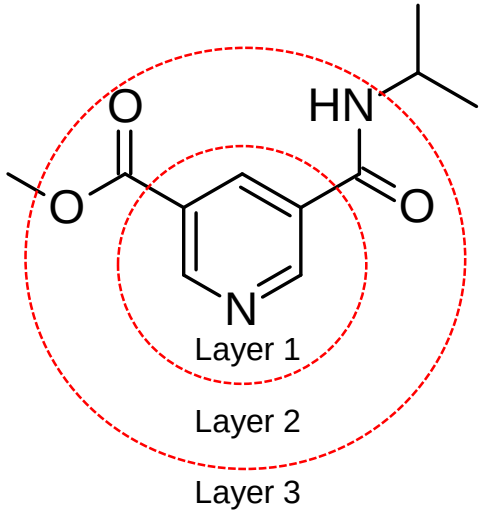
The ligand flexibility is considered by applying the anchor-first strategy: the largest rigid substructure of the ligand is identified and rigidly oriented in the active site by matching its heavy atom centers to the receptor sphere centers.

Various anchor orientations are generated and optimized using the scoring function and the energy minimizer. The orientations are then ranked according to their score, spatially clustered by their heavy atom root mean squared deviation (RMSD), and pruned.

Next, the remaining flexible portions of the ligand are built onto the best anchor orientations within the context of the receptor. It is assumed that the shape of the binding site will help restrict the sampling of ligand conformations to those that are most relevant for the receptor geometry.

DOCK

When an anchor has been selected, then the molecule is divided into non-overlapping segments, which are then arranged concentrically around the anchor segment. Segments are reattached to the anchor according to the innermost layer first and, within a layer, to the largest segment first. The remaining segments are subsequently re-attached during the conformation search.

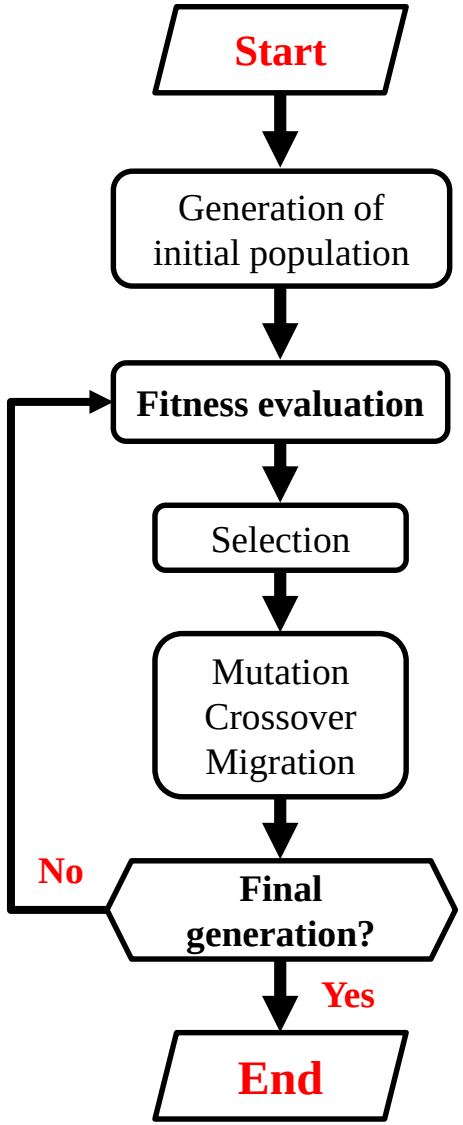


FlexX also uses an incremental construction method for ligand flexibility, but it differs from DOCK in that the positioning of the largest rigid substructure and the substituents are developed on the basis of geometrically restrictive interactions using pose clustering techniques.

Main Docking Software

AUTODOCK and GOLD

In order to explore the conformational states of the flexible ligands, Autodock and Gold employ a genetic algorithm, which is basically characterized by the evolution of a population of possible solutions *via* genetic operators (mutations, crossovers and migrations) to give a final population that optimizes a predefined fitness function. Degrees of freedom are encoded into **genes,** or binary strings, and the collection of genes, or **chromosome**, is assigned a value for fitness based on a scoring function. The **mutation** operator randomly changes the value of a gene, while **crossover** exchanges a set of genes from one parent chromosome to another, and **migration** moves individual genes from one subpopulation to another.

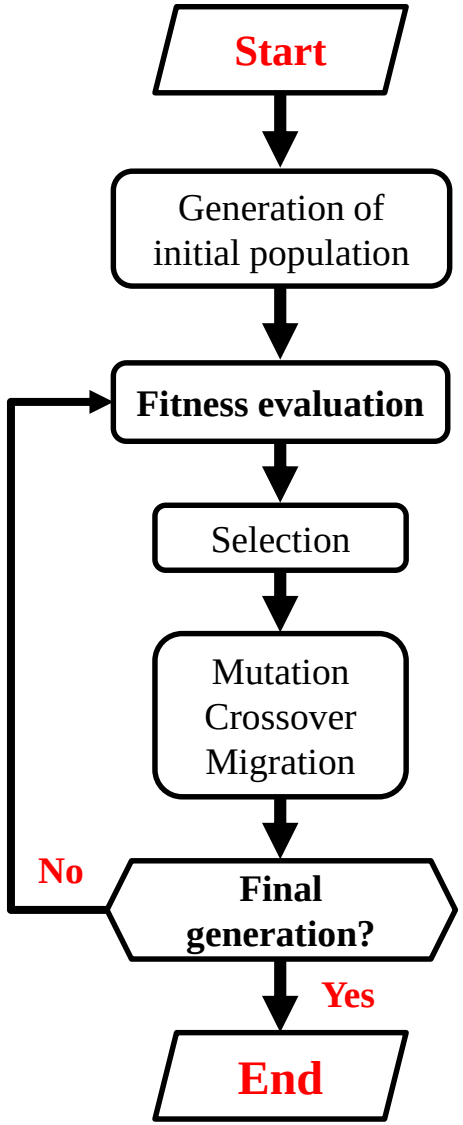


Genetic Algorithm

Initialization: Random generation of a population of possible solutions.

Selection: During each successive generation, a proportion of the existing population is selected to breed a new generation. Individual solutions are selected through a *fitness-based* process (based on the energy interaction of the individual solutions).

Reproduction: This step generates a second generation (population of solutions) from those selected above. For each new solution to be produced, a pair of "parent" solutions is selected for breeding from the pool selected previously. By producing a "child" solution using the above methods of crossover and mutation, a new solution is created which typically shares many of the characteristics of its "parents". New parents are selected for each new child, and the process continues until a new population of solutions of appropriate size is generated.



Main Docking Software

FRED

F.R.E.D. (Fast Rigid Exhaustive Docking) is a protein-ligand docking program, which takes a multiconformer library/database and a receptor file as inputs and outputs molecules of the input database most likely to bind to the receptor.

FRED's docking strategy is based on exhaustively score all possible positions of each ligand in the active site. The exhaustive search is based on rigid rotations and translations of each conformer.

This novel strategy completely avoids the sampling issues associated with stochastic methods used by many other docking programs. However, a proper conformational sampling analysis of the compounds to be docked is necessary for the generation of the multiconformer library used by FRED.

Scoring Functions

Scoring functions are based on the assumption that binding affinity can be represented as the sum of independent terms. The scoring functions are needed from both qualitative and quantitative points of view.

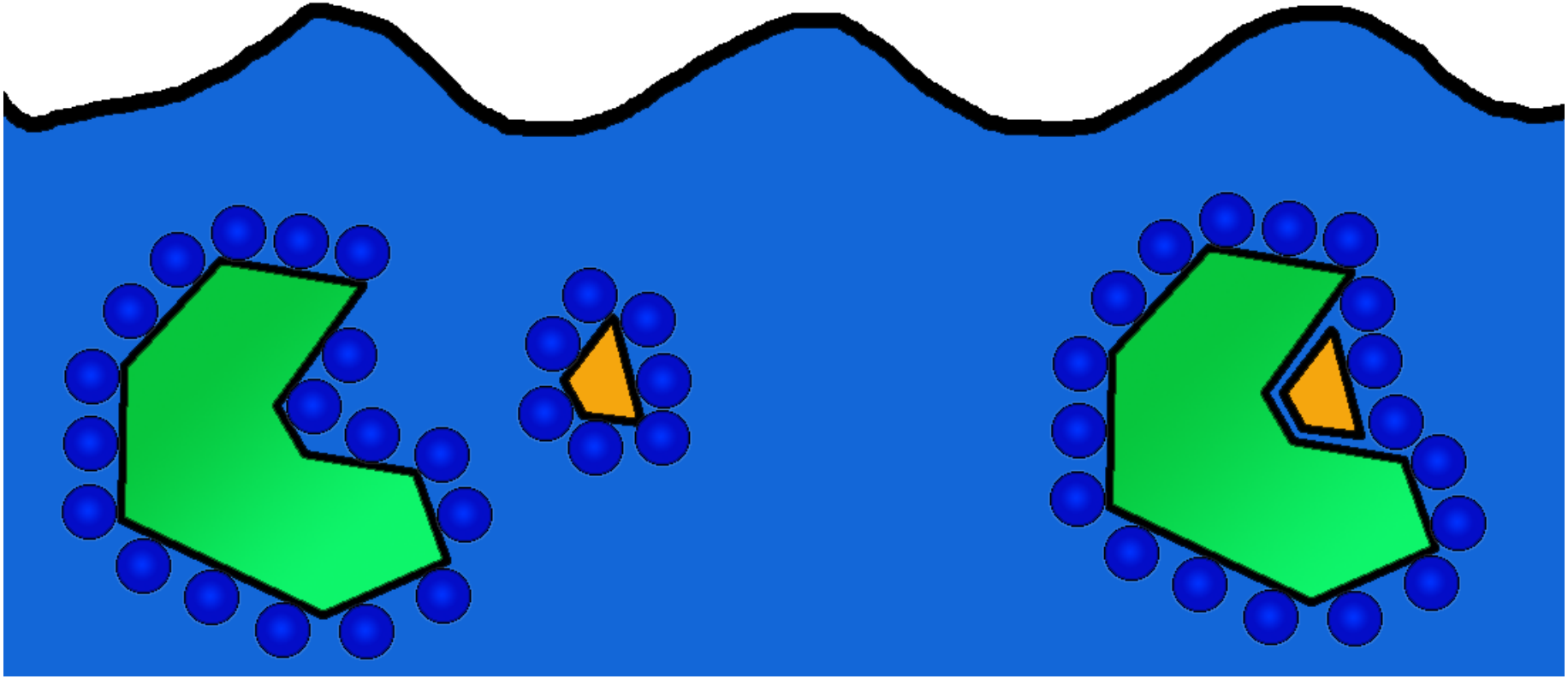
For each ligand, a docking algorithm usually produces a large number of solutions, and the first role of a scoring function is to rank the quality of the solutions generated for each ligand, to select the best one (top-scored pose).

The second role is to rank the docking results generated for different ligands according to their relative binding affinities. Basically, they evaluate the forces that contribute to the ligand binding. The three main forces involved upon the binding of a ligand are hydrophobic interactions, hydrogen bonding and electrostatic forces.

Hydrophobic Interactions

The driving force of hydrophobic interactions between a ligand and a receptor is to reduce the exposure of hydrophobic moieties of the ligand to water. The classic concept of the hydrophobic effect is as follows: a hydrophobic solute disrupts the structure of bulk water and decreases entropy because of stronger bonding and ordering of water molecules around the solute. Such disruptions can be minimized if nonpolar solute molecules aggregate. Water then forms one larger “cage” structure around the combined solutes, whose surface area will be smaller than the combined surface areas of the isolated solutes. This maximizes the amount of free water and thus the entropy.

Desolvation and the Hydrophobic Effect



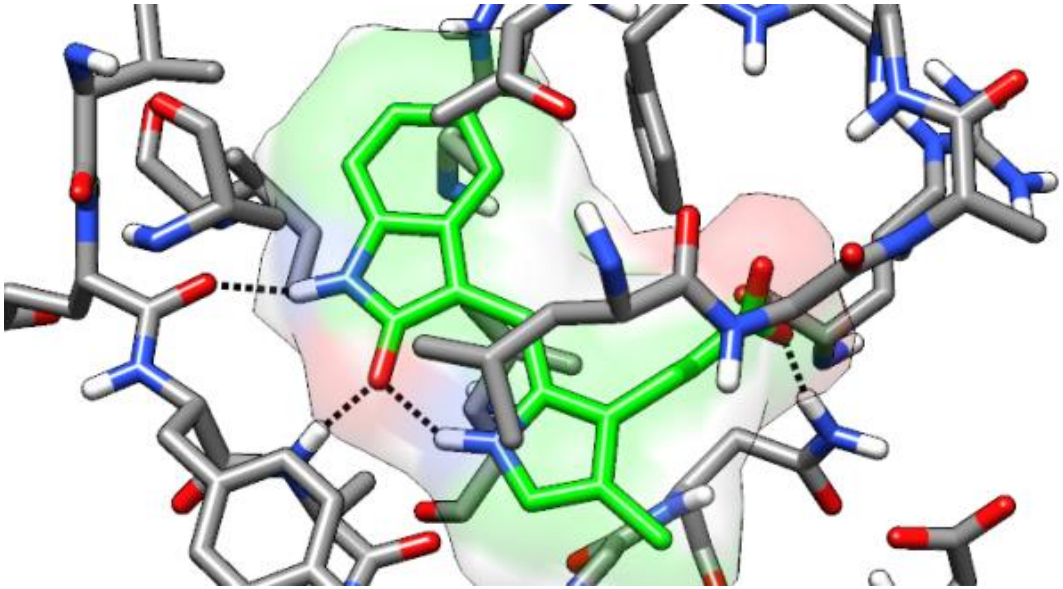
Hydrophobic Interactions

In addition, there is a favorable interaction between lipophilic groups in contact. The hydrophobic effect describes the energetic preference of nonpolar molecular surfaces to interact with other nonpolar molecular surfaces and thereby to displace water molecules from the interacting surfaces. The hydrophobic effect is due to both enthalpic and entropic effects. Hydrophobic interactions are short-range attractive interactions that make an important contribution to ligand–receptor binding affinities. They also contribute to specificity but in a less geometrically constrained way than hydrogen-bonding interactions. High occupancy of hydrophobic pockets of the receptor by a ligand leads to favorable van der Waals attractions, which results in increasing the binding affinity. Note that, although the hydrophobic interactions are weak, there are a lot of them, so the effect can be substantial.

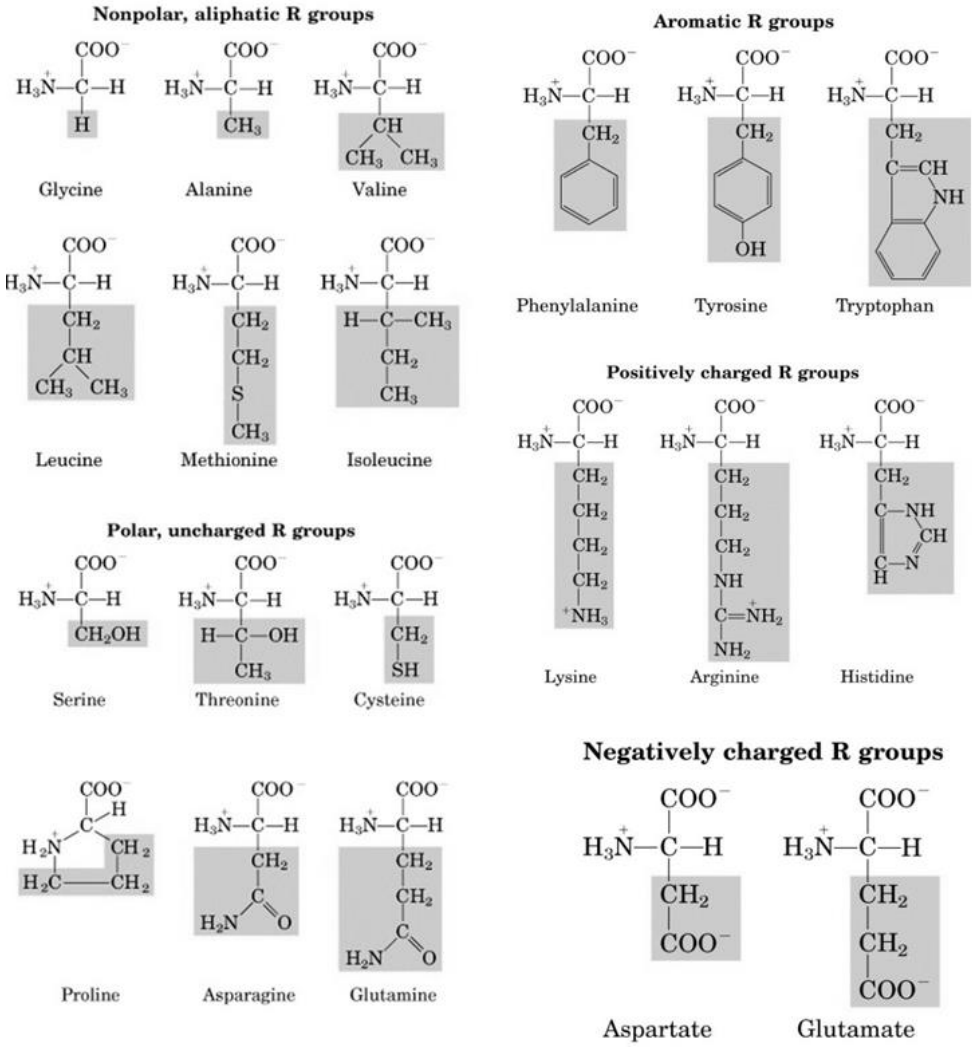
Hydrogen Bond Features

Hydrogen bonds are directional and contribute to the positioning of a molecular fragment in a precise orientation within the active site of a protein. Multiple hydrogen bonds can «zip» the ligand into place.

A protein has many hydrogen bond donor and acceptor groups. In addition to the backbone peptidic moiety that can form hydrogen bonds with the C=O or the N-H groups, hydrogen bonds can be formed with the side chains of many residues: Asn, Asp, Arg, Cys, Gln, Glu, His, Lys, Met, Ser, Thr, Trp, and Tyr.



Hydrogen Bond Features



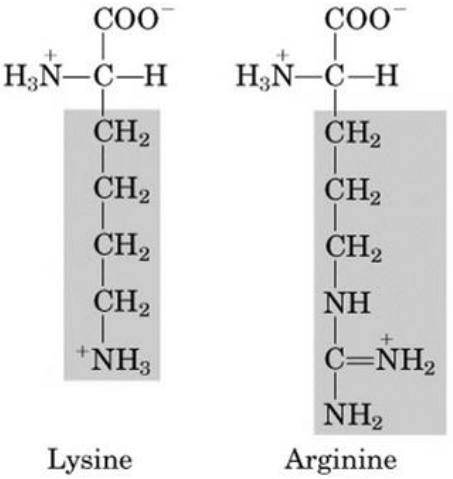
Hydrogen bonds always convey specificity to a recognition process but do not always add much binding free energy. H-bonds between ligand and receptor are formed at the expense of bonds broken between ligand and solvent, as well as between receptor and solvent, as a result of the release of water molecules. Desolvation of donor and acceptor may nearly compensate for the effects of hydrogen bond formation.

Electrostatic Interactions

Electrostatic interactions occur bewteen polar or charged groups. The following residues contribute to the formation of electrostatic (columbic or salt-bridge) interactions in proteins. Electrostatic interactions appear to be significantly stronger than neutral hydrogen bonds.

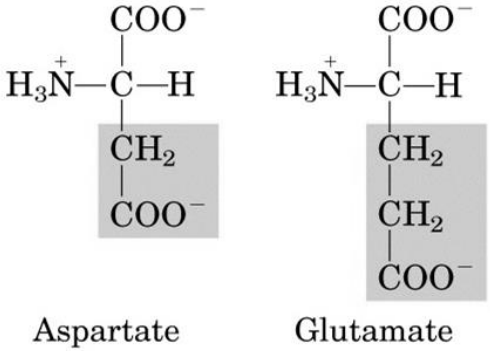
Positively charged residues:

- Arginine
- Lysine



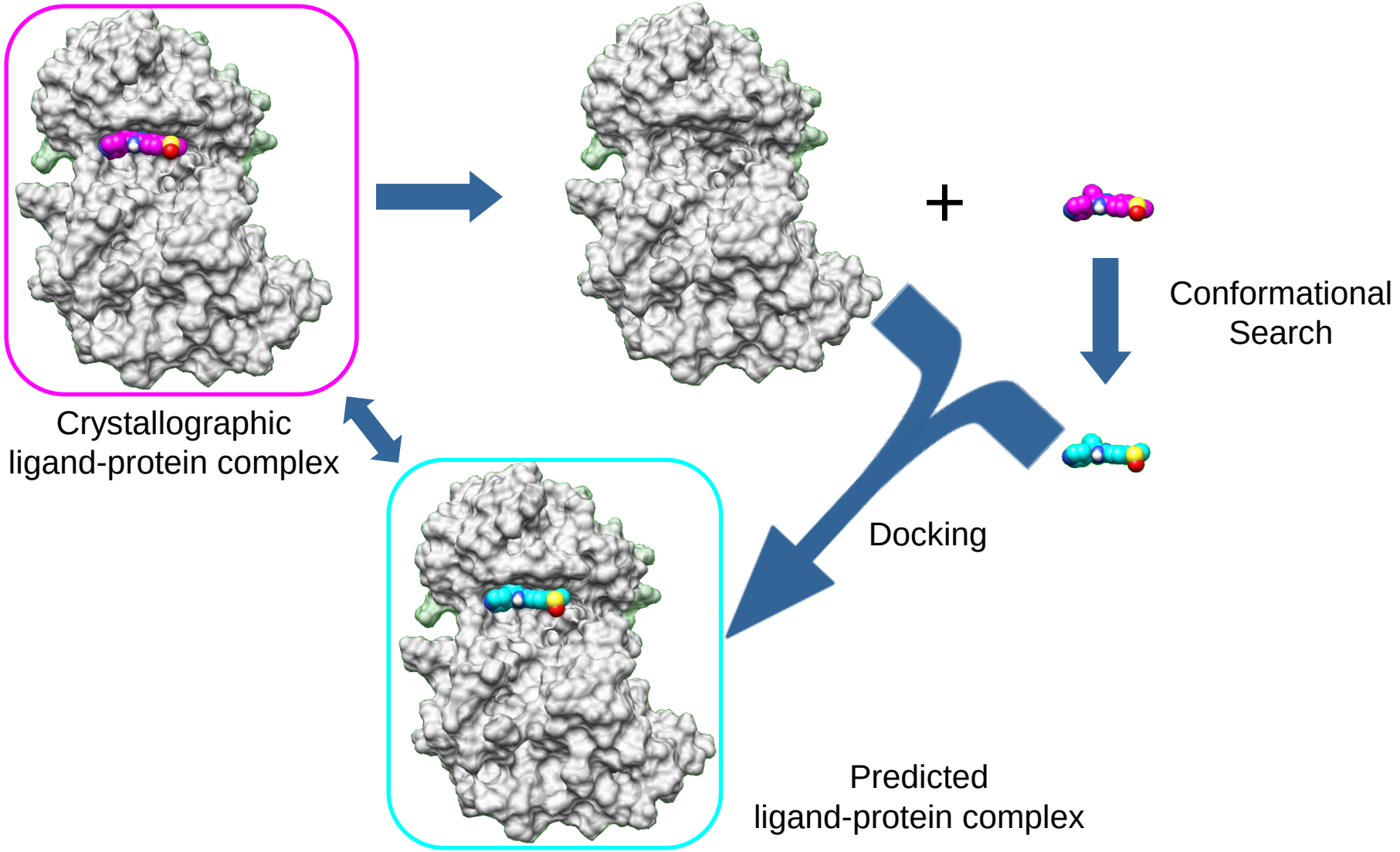
Negatively charged residues:

- Glutamic Acid
- Aspartic Acid



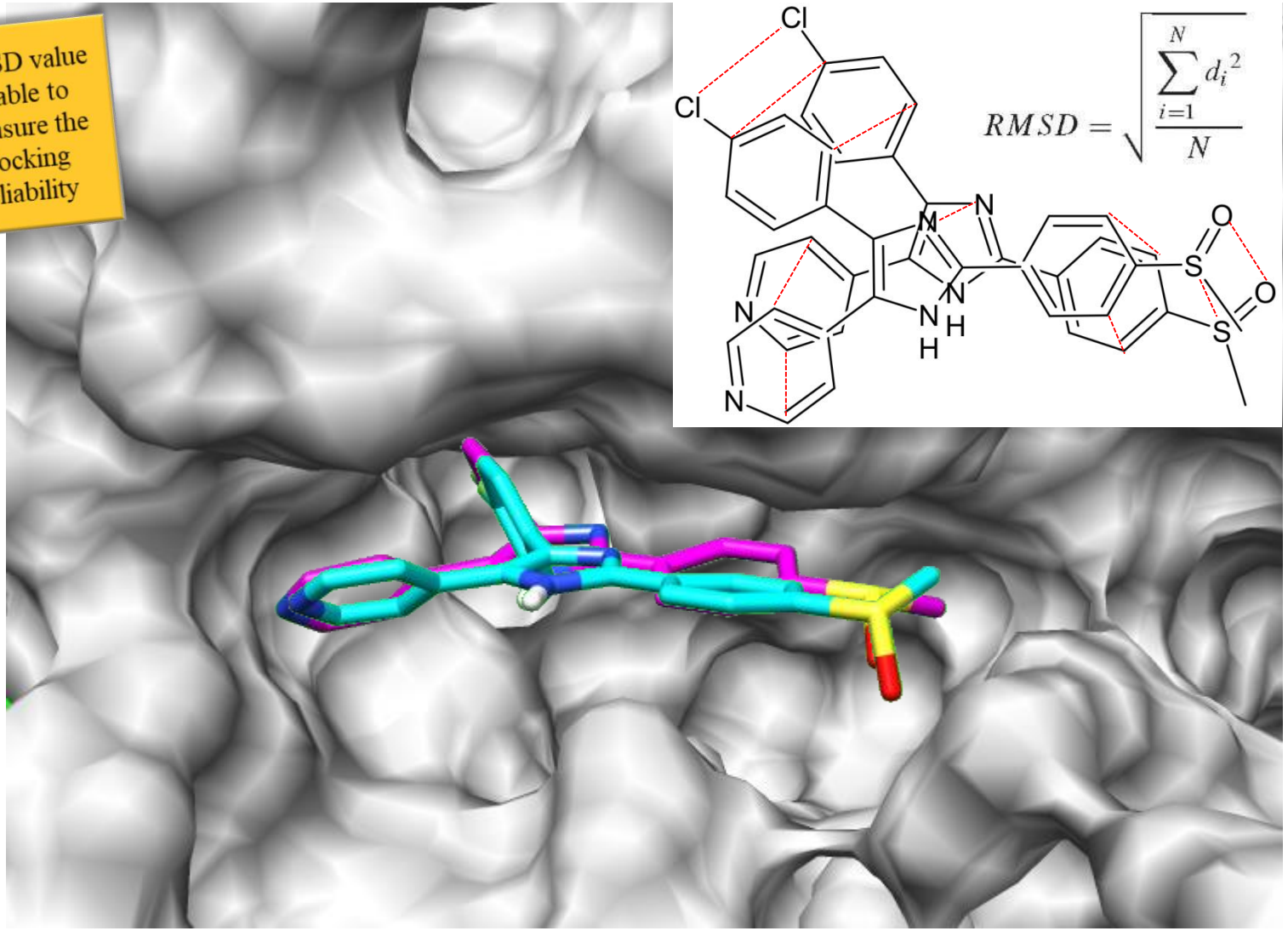
Docking reliability studies

Self-docking analysis



Self-docking analysis

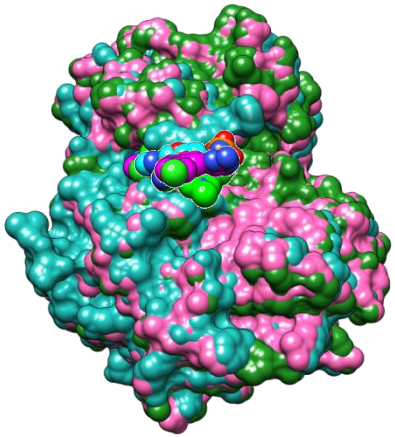
RMSD value is able to measure the docking reliability



Self-docking analysis

Selfdocking is not the best way to test docking reliability

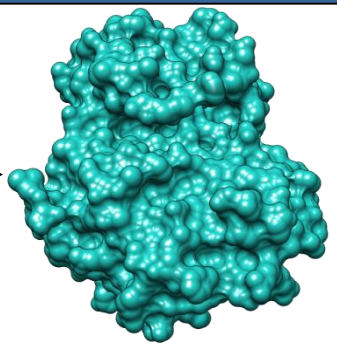
When you dock new compounds you do not have a protein that is perfectly adapted for the interaction with your ligand

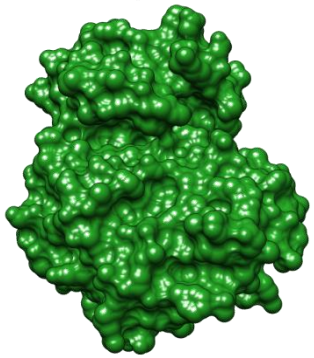


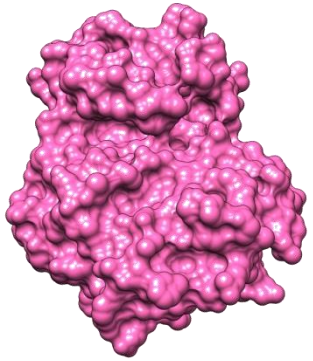
LIG A → RMSD A → Protein A

LIG B → RMSD B → Protein B

SD C → Protein C

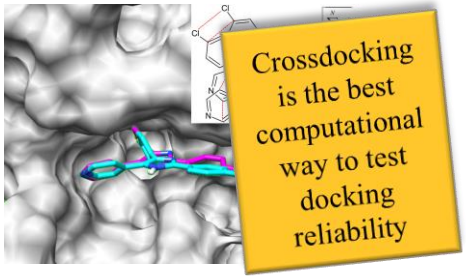




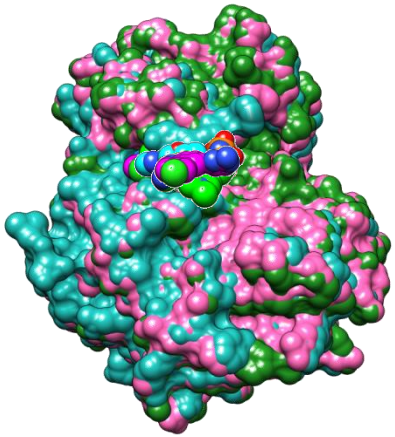


	Protein A	Protein B	Protein C
Lig A	2.5		
Lig B		1.8	
Lig C			3.5

Cross-docking analysis

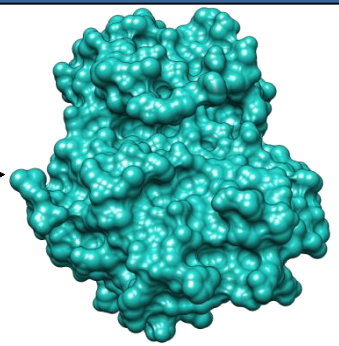


Crossdocking is the best computational way to test docking reliability

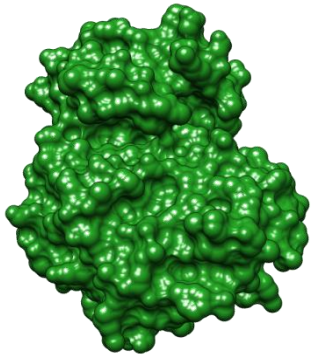


LIG A

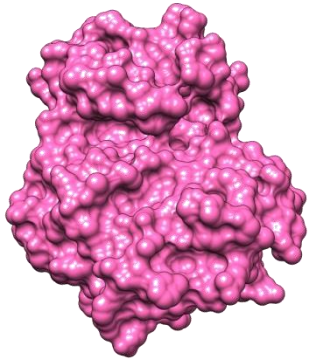
LIG B



Protein A



Protein B



Protein C

	Protein A	Protein B	Protein C
Lig A	2.5	2.1	0.5
Lig B	1.4	1.8	0.8
Lig C	2.3	3.4	3.5

Cross-docking analysis

The average RMSD provides an overall estimation of docking reliability

	1GWQ	1X7E	1ZKY	2B1V	2B1Z	2FAI	2G44	2I0J	2QE4	2QGT	2QGW	2QSE	3ERD	3HLV	3HM1	3Q95	3UU7	4IU7	4IV4	4IW6	4IWC
	1.6	3.0	1.9	1.0	0.5	0.6	0.9	1.4	1.3	0.6	2.0	6.7	0.8	0.6	6.5	0.4	2.6	0.4	1.6	1.1	2.5
	0.4	3.1	3.0	1.1	0.4	2.7	2.8	1.6	1.2	0.4	0.6	6.7	2.5	0.6	0.7	0.4	3.3	2.9	1.7	3.7	2.4
	0.5	0.5	1.9	2.7	2.7	2.9	2.9	1.5	1.3	0.6	2.1	0.7	1.2	0.6	0.5	0.3	1.6	0.6	1.6	1.8	6.9
	0.5	1.0	1.5	1.1	0.5	0.7	0.8	1.9	0.7	0.4	0.4	2.1	0.3	0.7	0.7	2.7	0.7	3.0	2.1	1.0	2.2
	0.4	0.6	1.4	0.6	0.3	0.4	0.9	1.6	0.9	0.5	0.6	6.7	2.5	2.8	2.8	2.8	1.1	0.7	1.8	2.0	2.3
2B1Z	0.4	1.2	3.2	1.9	1.3	2.7	1.1	2.8	1.8	1.2	0.6	2.1	6.6	1.2	0.6	0.6	0.4	0.9	0.7	0.7	1.0	1.3
2FAI	0.3	0.5	3.0	1.7	1.7	0.3	0.5	2.7	1.6	0.8	0.6	0.5	0.9	2.5	0.6	3.0	0.4	1.0	0.7	0.8	1.2	1.8
2G44	0.4	0.7	3.2	1.5	1.2	2.7	0.8	2.8	1.3	1.6	0.6	1.1	1.0	1.3	0.6	0.8	0.4	1.3	1.6	0.7	1.8	1.7
2I0J	0.5	7.1	3.2	1.8	1.2	0.5	0.8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.4	2.8	2.9	0.7	4.1	6.0
2QE4	0.8	0.4	0.6	1.7	1.3	2.6	0.7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.8	2.8	0.3	0.5	0.6	2.5
2QGT	0.4	1.9	3.0	1.5	0.8	2.6	0.6	2.7	0.7	0.9	0.2	0.6	[illegible]	2.6	0.9	0.8	0.5	0.6	0.6	1.7	1.7	1.1
2QGW	0.3	0.4	3.1	1.7	1.4	0.4	0.7	0.8	1.8	0.6	0.4	[illegible]	[illegible]	1.1	0.9	0.8	0.3	1.0	0.3	0.6	1.0	2.6
2QSE	0.4	0.3	3.2	1.6	1.3	0.4	1.2	2.8	1.8	0.9	0.5	0.5	6.7	1.1	0.5	0.6	0.4	2.7	0.5	0.6	1.6	3.0
3ERD	0.5	0.5	0.8	1.8	2.4	0.9	2.6	2.8	3.0	0.7	0.7	0.5	6.7	0.3	0.6	0.7	0.6	2.7	0.6	1.2	4.0	2.4
3HLV	0.3	7.1	0.8	1.5	2.4	0.8	2.8	2.0	1.5	1.4	0.7	6.8	0.5	1.1	0.7	1.2	0.2	2.8	1.5	1.2	1.5	2.6
3HM1	0.7	0.5	0.4	2.1	0.9	0.6	2.7	2.9	3.0	0.7	0.7	0.5	2.2	2.5	0.8	0.7	0.8	2.4	3.3	3.6	4.0	2.7
3Q95	0.2	1.9	3.1	1.7	0.8	0.5	0.7	1.3	0.6	0.9	0.3	2.1	1.2	1.2	0.6	1.0	0.3	0.8	0.6	0.6	1.8	1.2
3UU7	2.7	1.6	3.0	1.8	0.8	2.7	0.7	1.0	1.3	1.4	3.6	2.2	6.7	2.8	2.8	2.8	2.7	0.9	0.6	0.7	1.7	4.3
4IU7	0.7	1.2	7.0	2.0	1.5	0.8	1.3	1.1	1.3	1.5	0.8	1.2	1.0	2.8	2.8	1.3	6.5	1.4	0.2	1.4	0.9	2.3
4IV4	0.8	7.1	7.0	1.9	1.2	6.3	0.9	1.0	1.1	1.5	0.9	1.3	1.0	3.0	6.5	1.3	6.5	1.4	0.3	0.2	0.3	1.8
4IW6	0.8	7.1	7.1	5.7	1.2	0.8	1.2	1.5	1.1	1.5	1.0	1.3	1.2	2.8	6.5	1.4	0.8	1.1	0.3	0.3	1.4	1.8
4IWC	2.7	7.1	1.9	2.7	0.9	2.7	1.0	1.3	1.9	2.0	1.0	6.8	1.6	2.9	3.1	3.3	2.9	1.9	1.2	2.0	2.2	0.3

Average RMSD 1.9 Å

RMSD < Good ... Bad