Why having a method for docking protein and small molecules?

Virtual screening

-there are lots of compounds, find out which ones bind.

Lead optimization

- **provides plausible configurations for developing a hypothesis for candidate improvement**
- **is ideal if the binding strength can be predicted**

How to build a docking algorithm (starting with protein-protein interfaces)?

What might be the main components to consider?

- experimental prior
- scoring function
- conformational sampling (interface)
- conformational sampling (flexibility)

experimental prior

- **coevolution info**
- **mutational data**
- **homologous structures**
- **other experimental data (such as interface mapping by NMR or protease resistance)**

scoring function

- **molecular mechanics based**
- **knowledge based**
- **heuristics - shape complementarity, etc.**

critical considerations are:
-how to derive the knowledge based forcefield given limited data
-how solvents are handled.

To have good coverage on all degrees of freedom, usually the scoring terms are simplified to speed up the calculations. But the trade off is accuracy.

conformational sampling (interface)

- **stochastic search (such as Monte Carlo)**
- **grid search (systematically search a predefined set of conformations)**
- **shape matching (convex-concave vector surface mapping)**
- **human vision (Foldit)**
- **brute force (Molecular dynamics)**

conformational sampling (flexibility)

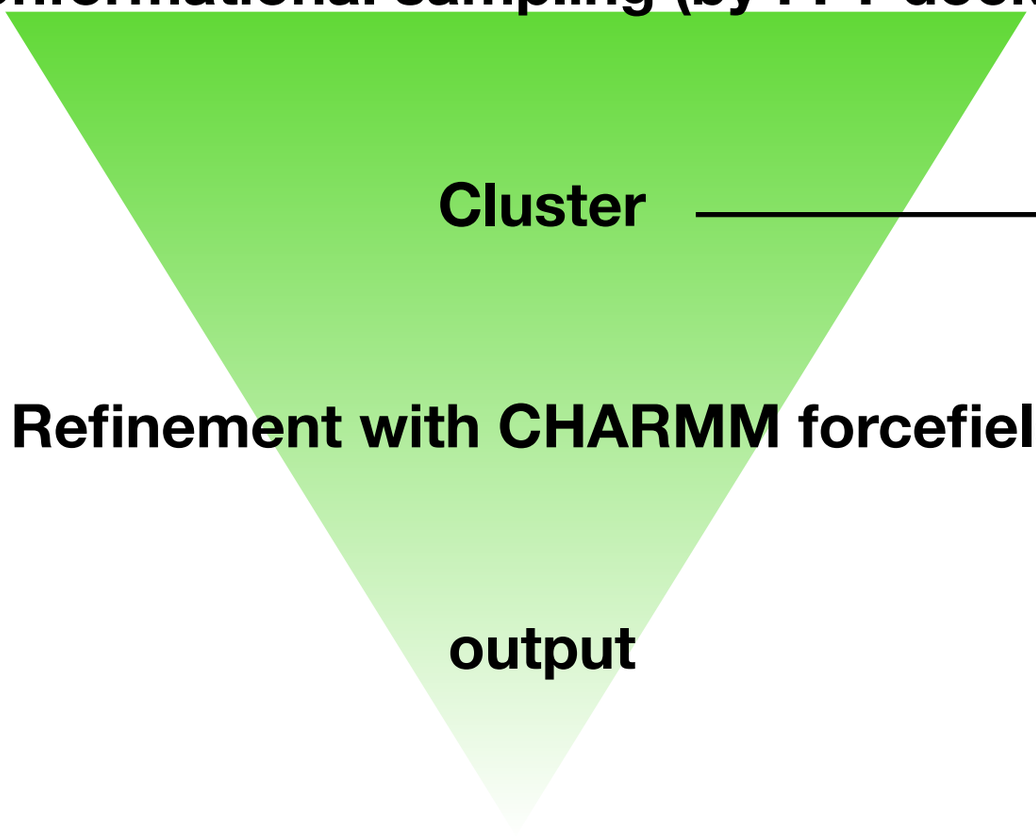
- **Rigid body docking (bound/unbound cases)**
- **chemical flexibility (torsional angles in the backbones and the sidechains)**

In an unbound docking case, structures of the two interacting components are determined separately and they may not be in the states that are perfectly complementary as the bound state. So the calculations will have to estimate a certain degree of clashes. This is often done by softening the Van der Waals potential or use a reduced representation of the atomic details to approximate the solution.

If handling protein-small molecule docking, an additional module that handles the conformations of the small molecule is needed.

Using ClusPro as an example of a docking protocol
One of the best performing algorithms but use unconventional assumptions to find solutions

Conformational sampling (by FFT docking)



ClusPro assumes that pockets of high convergence should contain the low energy state. Therefore it finds the center of the most populated clusters

Partition function for all states: $Q = \sum_j \exp(-E_j/RT)$

For a cluster k: $Q_k = \sum_k \exp(-E_k/RT)$

Probability $P_k = Q_k/Q$

assume the energies are the same in the same cluster $E_j=E$, then

$P_k = Q_k/Q = \exp(-E/RT)*N_k/Q$, N_k being the number of structures in cluster k

$$P_k \propto N_k$$

somehow this assumption allows ClusPro perform well in the CAPRI docking competition