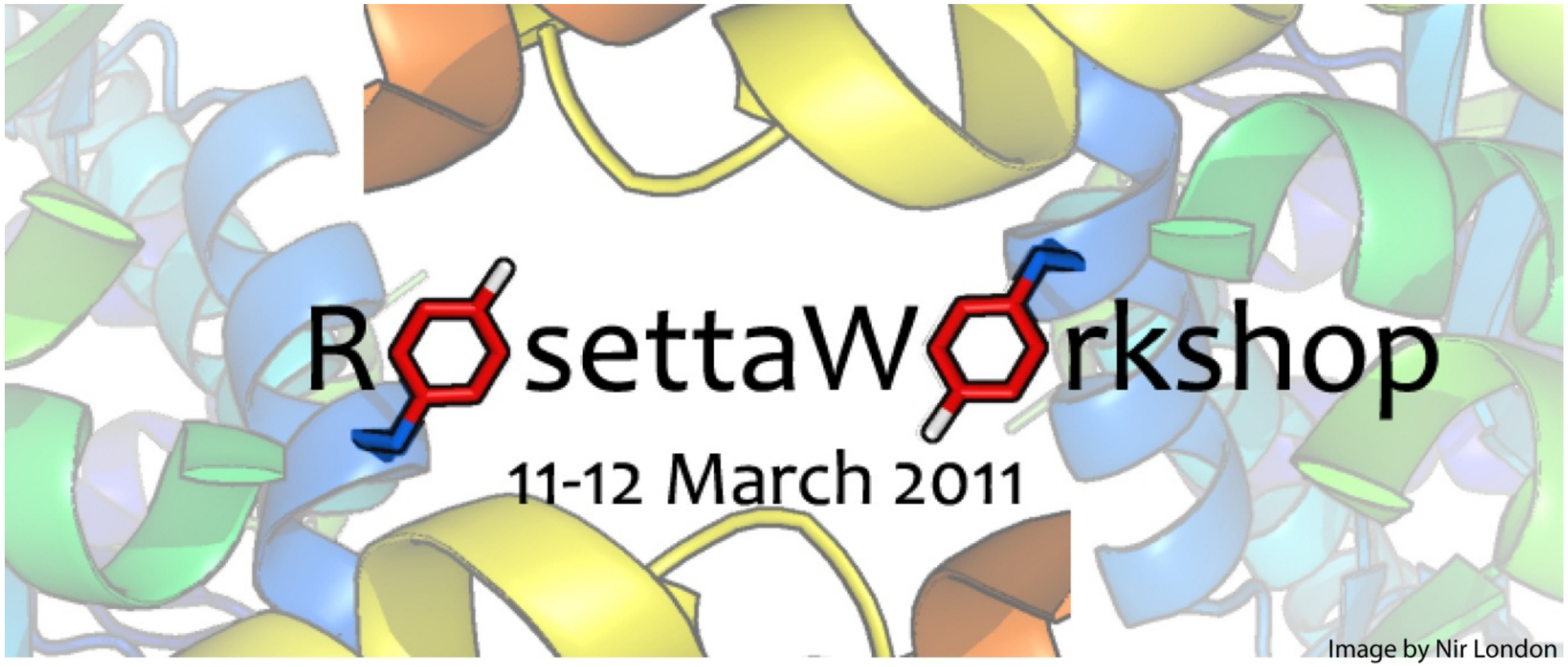
Protein-Protein Docking in Rosetta 3.2



Outline

- Quick Overview
  - Why Protein-Protein Docking
  - Rosetta Monte Carlo Grid Search
- Rosetta Docking Procedure Low Resolution
  - Antibody-Antigen Docking Example (Epitope Search)
  - Data Analysis
- Rosetta Docking Procedure High Resolution
  - Antibody-Antigen Docking Example (Epitope Refinement)
  - Data Analysis
- References and Resources
  - Ab/Ag SnugDock, Rosetta Dock

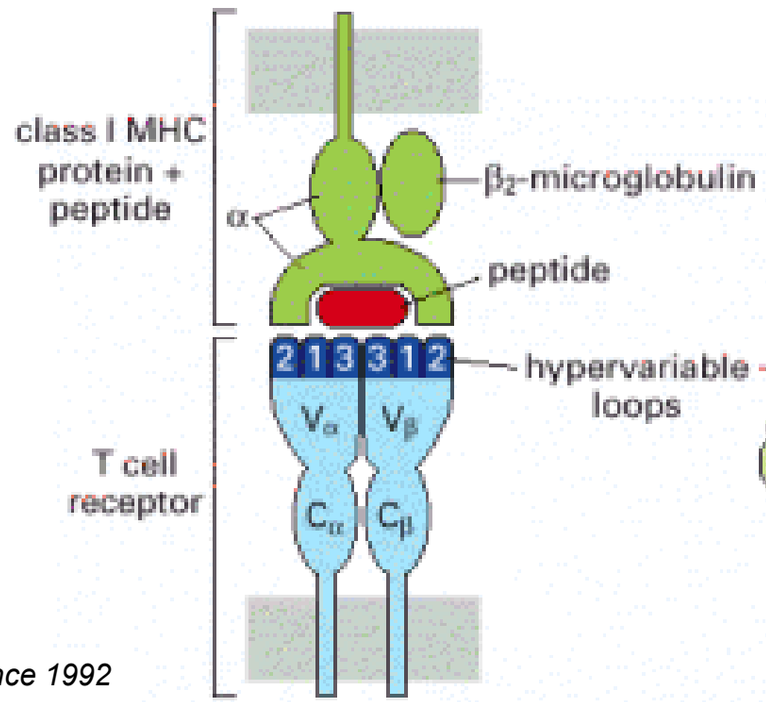
Overview - Protein Docking

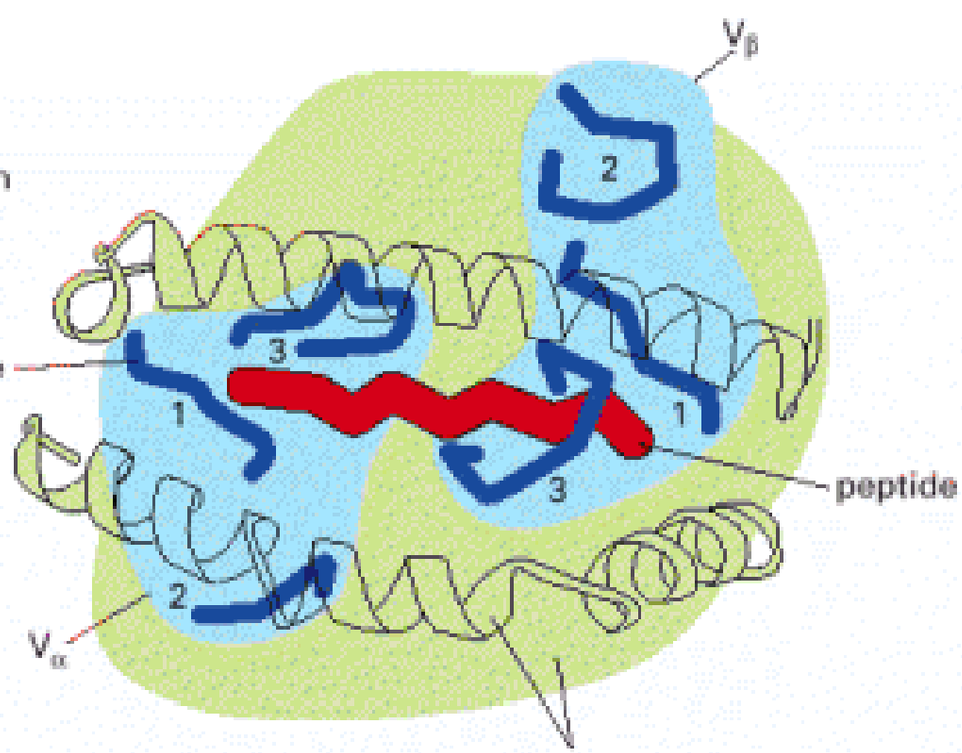
"..post-genomic goal is the characterization of the structures of protein–protein complexes, and computational tools offer an inexpensive means to carry out large-scale studies."

Gray et. al. 2005

Protein - Protein Interactions are integral to :

- Protein localization
- Competitive inhibition
- Allosteric regulation
- Gene regulation (signaling pathways)
- Signal transduction
- **Pathogen clearance (antibody-antigen)**





A.A. Kossiakoff, Science 1992

Overview - Protein Docking

Characterize how proteins interact in two ways:

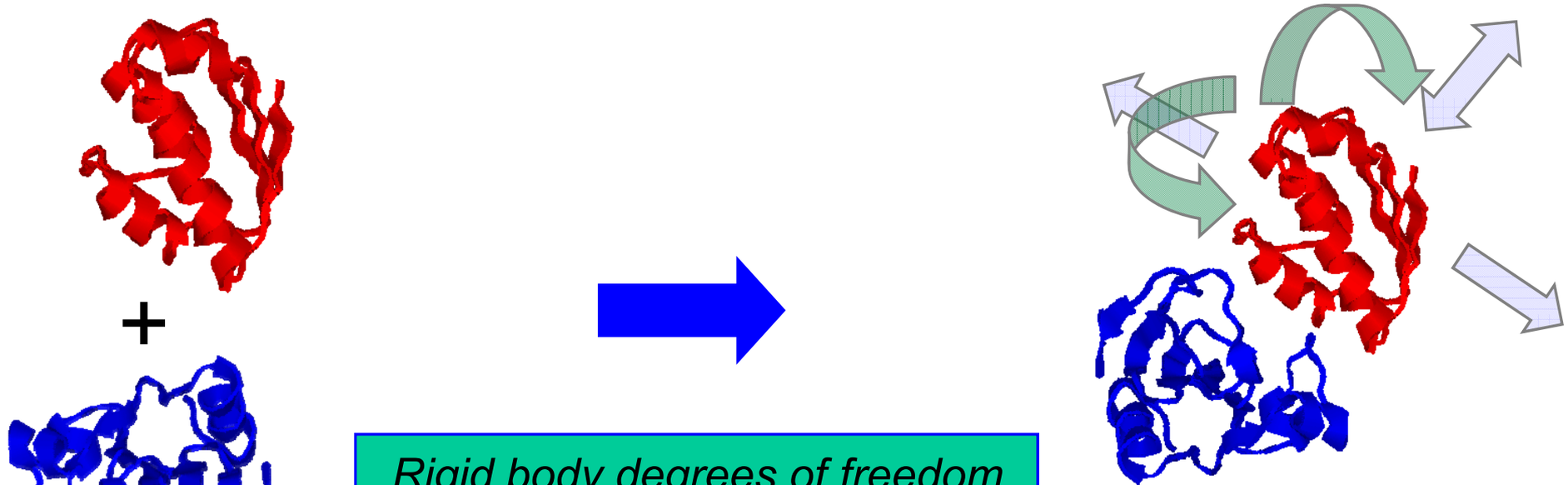
1. Principles from chemistry and physics (molecular mechanics)

2. Knowledge based (information from the PDB, laboratory experiments, and thermodynamic measurements)

Knowledge based gives parameters to limit infinite search space

Overview - Protein Docking

Predict complex from their independent binding partners



Monomers

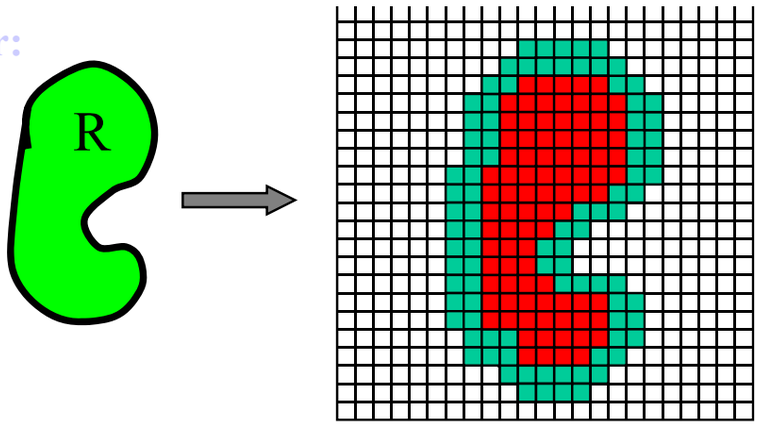
Rigid body degrees of freedom
3 translation
3 rotation

Complex

© Jens Meiler

Overview - Protein Docking

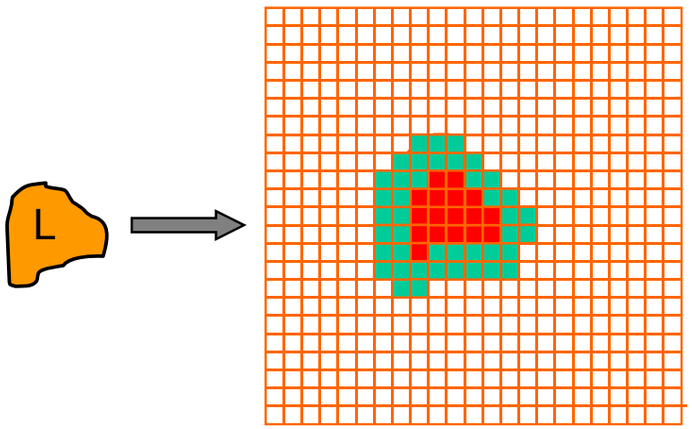
Receptor:



Assign value to each cell:

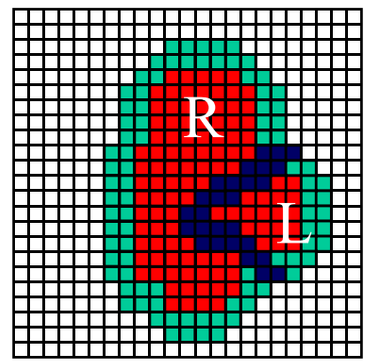
- Exterior: a(i,j) = 0
- Surface: a(i,j) = +1
- Interior: a(i,j) = -15

Ligand:



- Exterior: b(i,j) = 0
- Surface: b(i,j) = +1
- Interior: b(i,j) = -15

pace coupled with side cha



$S(R,T)$

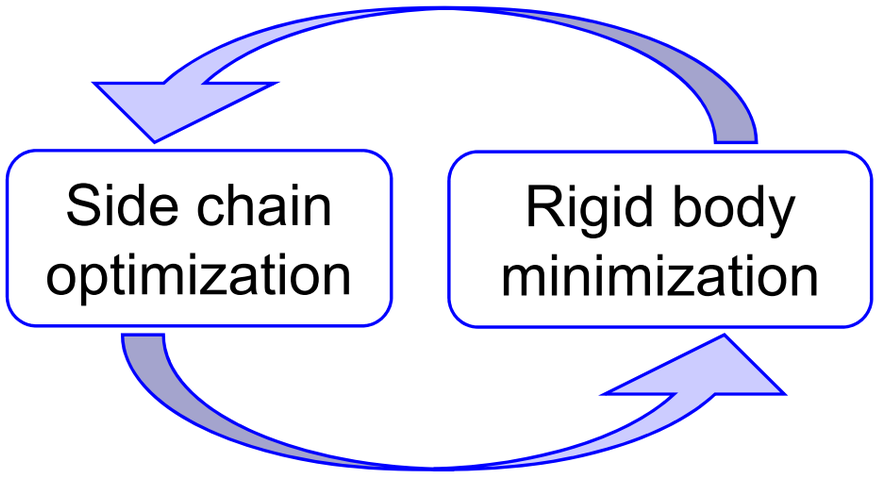
$$S(R,T) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} a(i,j,k)\,b'\left(i+T_x, j+T_y, k+T_k\right)$$

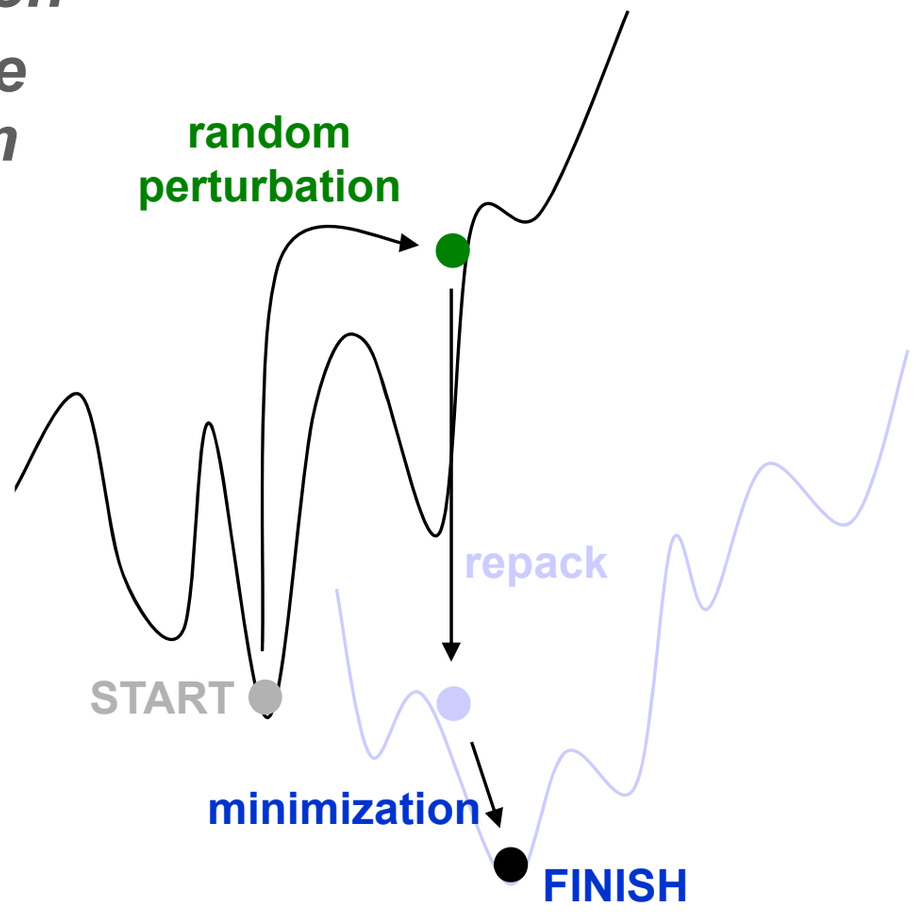
Overview - Protein Docking

Rosetta Dock

Monte Carlo search with minimization

Cycling between rigid body and side chain optimization locates minimum





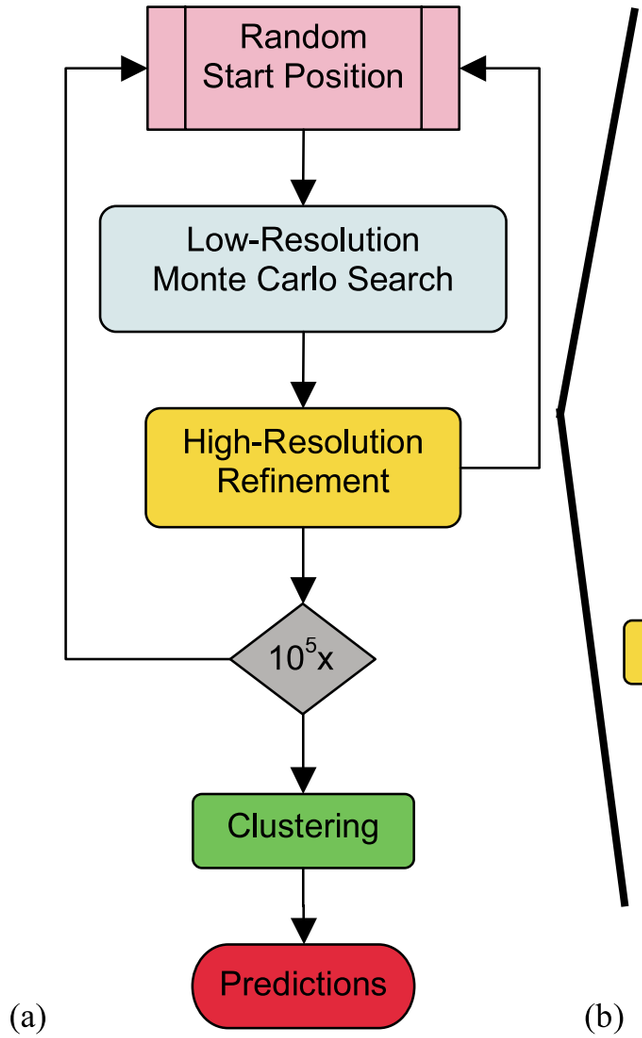
© Jens Meiler

Overview - Protein Docking

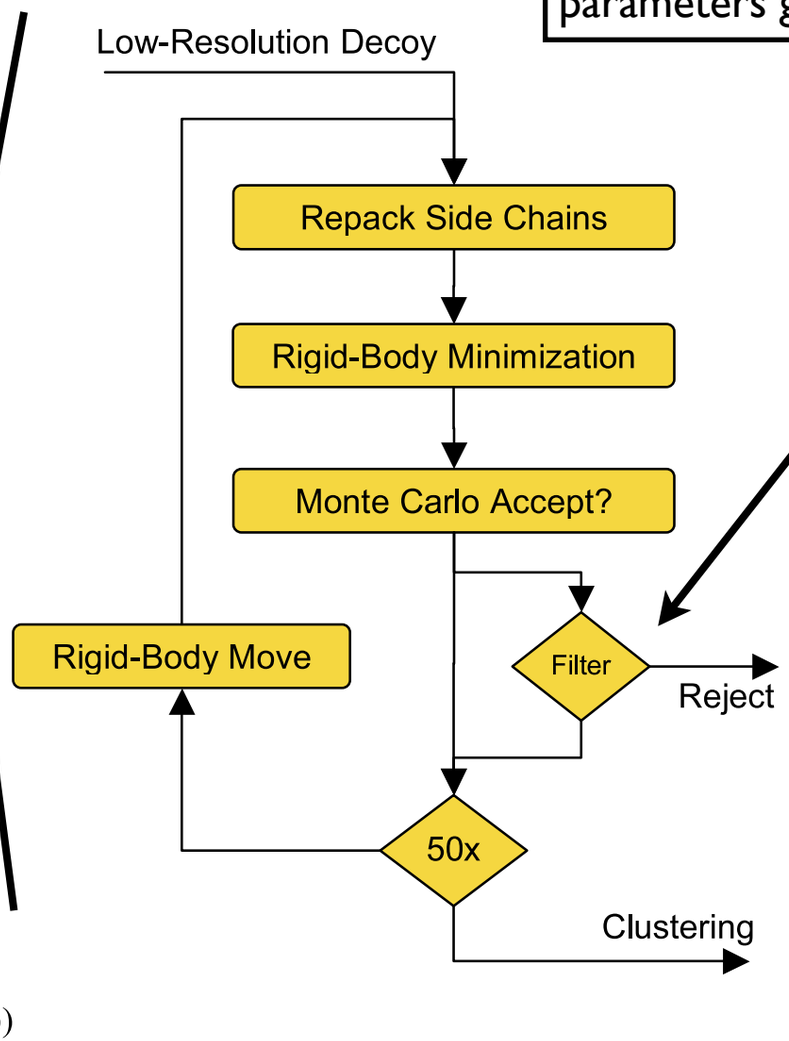
Rosetta Dock

Low Resolution - The global search of the protein target.
High Resolution - Adding side chains, energy minimization.

parameters given from experimental data help filter.



(a)



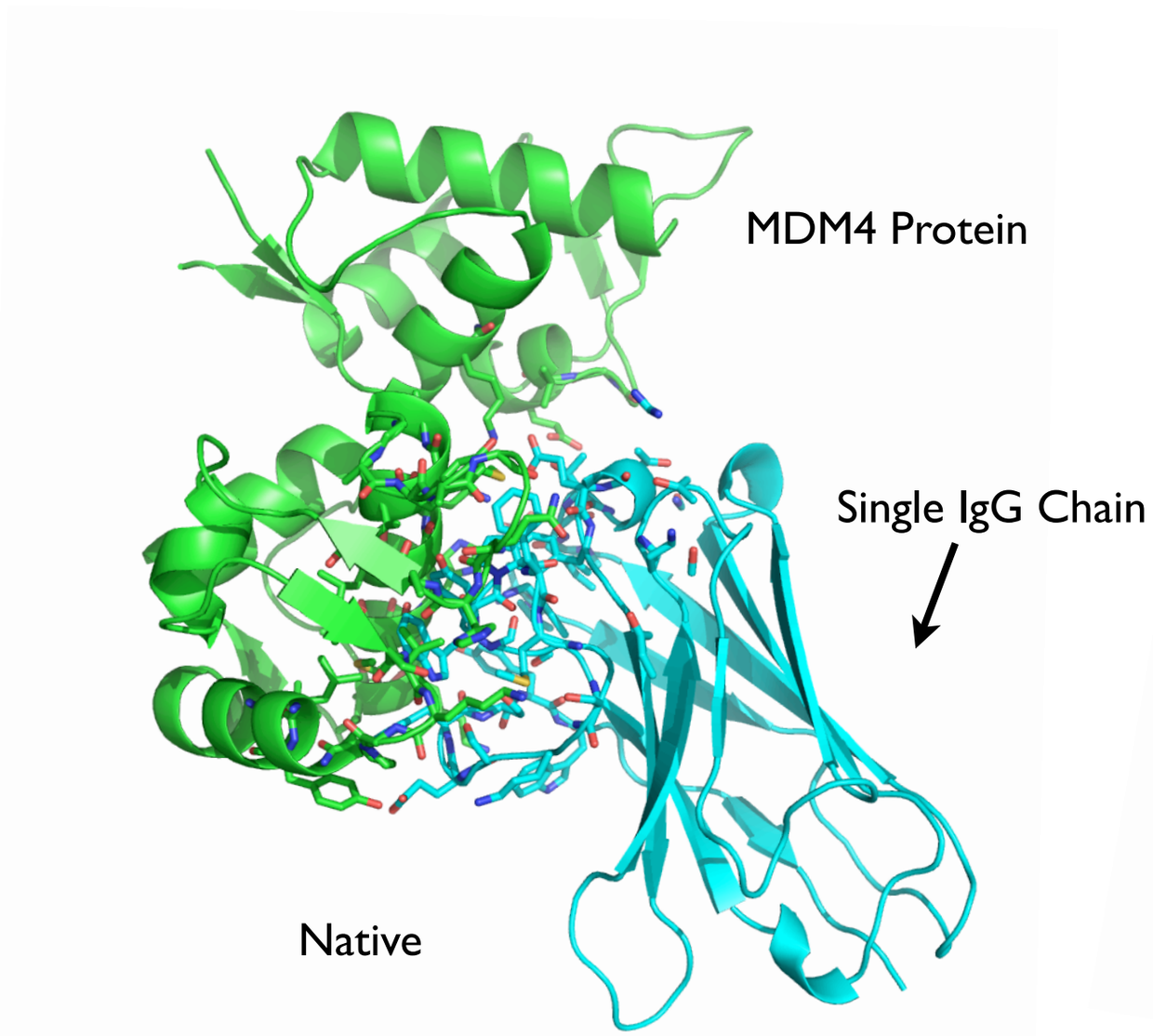
(b)

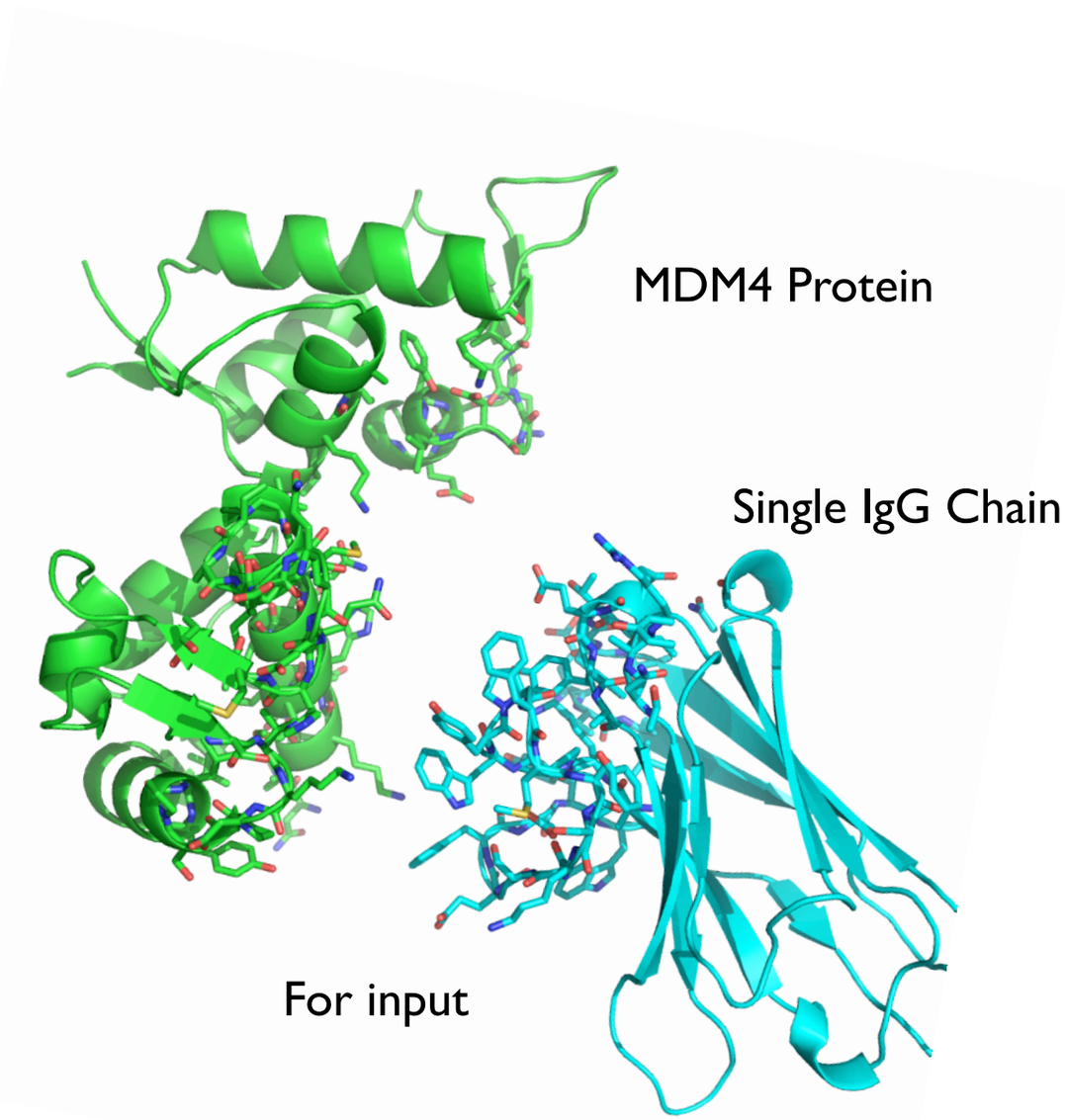
Gray et. al. 2005

Figure 1. Docking protocol. (a) Process flowchart; (b) detail of the refinement stage.

Using Protein Docking

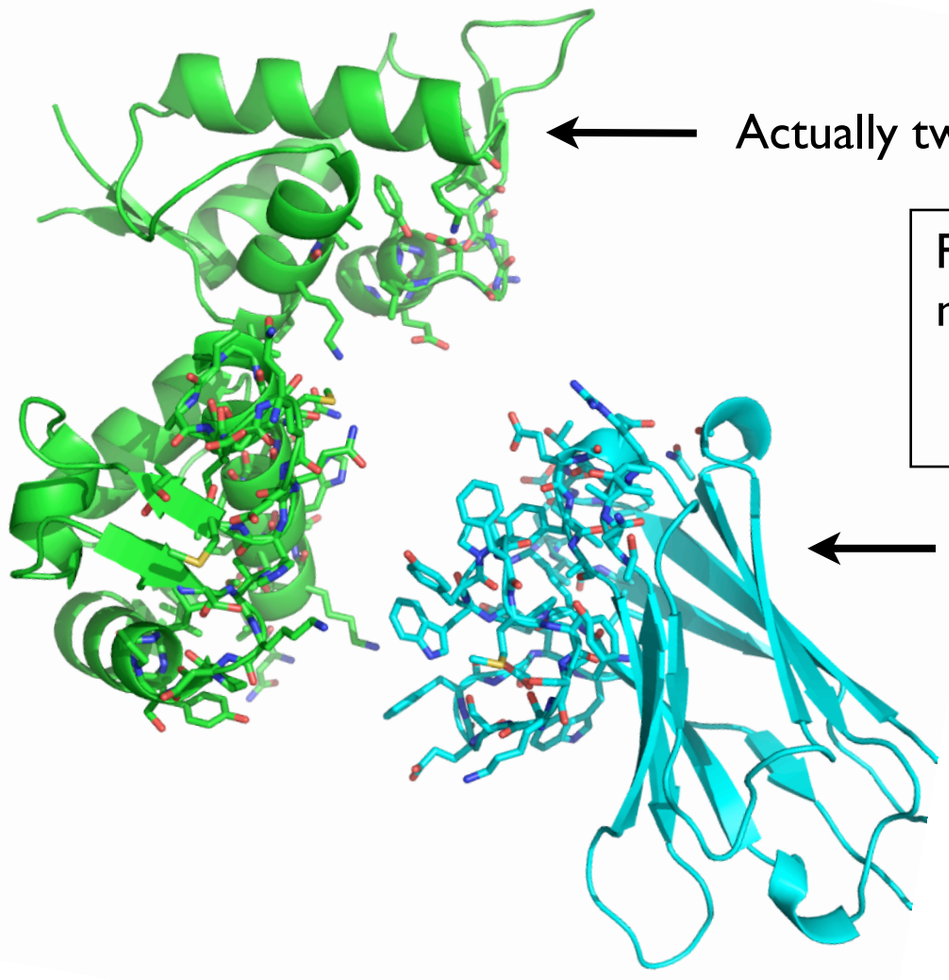
For this example, we will assume we know nothing about the orientation of either binding partners and that we are doing a global dock that includes both the low-resolution search and the high resolution refinement





Using Protein Docking

PDB Prepearation



Actually two chains in the PDB (B, C)

For Rosetta to recognize these as two distinct complexes, the antigen must be viewed as one entity. That is, we will label both chain B and C as chain B. Chain B will then need to come first in the PDB file. It must also be renumbered so it is viewed as one complex

Chain H

Using Protein Docking

Rosetta Scripts - Allows for specification of modular movers
-We will follow the docking protocol through a series of movers specified in Rosetta Scripts

Command Line:

```
$ROSETTA_BIN/rosetta_scripts.$ROSETTA_SUFFIX @$WORKSHOP_ROOT/
tutorials/protein-proteindocking/tutorial_files
flags.txt -database $ROSETTA_DATABASE > log.txt &
```

Using Protein Docking

```
-parser:protocol low_res_docking.xml
-s 2VYR_input_low_reso.pdb
-docking
        -dock_pert 8 5
        -spin 1
        -randomize1
        -docking_centroid_outer_cycles 50
        -docking_centroid_inner_cycles 500
-docking:dock_mcm_trans_magnitude .1
-docking:dock_mcm_rot_magnitude 1
-native 2VYR_input.pdb
#-parser:view
-nstruct 10
-linmem_ig 10
-ex1
-ex2
-ex1aro
-overwrite
-packing:repack_only
-out:pdb
```

Flags File:

- **parser:protocol** - This specifies the name of the file that contains the list of movers we want to do to our protein (looked at in more depth)
- s - specifies the input file we will use (antibody and antigen not in complex)
- **docking:dock pert** - specifies the degree of rotation and translation in the low resolution docking step that we allow for each move. 8 angstrom translation and 5 degree rotation.
- docking:spin - spin one partner around an axis in between the two docking partners
- **docking:randomize** - this flag is used truly if we do not know anything about a given binding site. This allows the input structures to start at a random location on the binding partner and walk along the energy landscape from that position.
- docking:docking_centroid_outer/inner - repeats the number of docking moves in the low resolution, centroid mode.
- docking:dock_mcm_trans_magnitude - how far can the binding partner translate in high resolution mode...this assumes we have already found the binding site in low resolution mode.
- docking:dock_mcm_rot_magnitude - how far can the binding partner rotate in high resolution mode
- nstruct - how many output models do we need
- linmem_ig 10 - linear memory of the interaction graph used in the repacker
- ex1, ex2, ex1aro - specifies the rotamer libraries we will use
- overwrite - overwrites the current pose (just in case)
- packing:repack_only - ensures we don't start designing amino acids at the interface when we call on the packer
- out:pdb - output the file as a pdb

Using Protein Docking

Rosetta Scripts

```
<dock_design>
        <SCOREFXNS>
        </SCOREFXNS>
         <TASKOPERATIONS>
                <InitializeFromCommandline name=ifcl/>
         </TASKOPERATIONS>

        <FILTERS>
        </FILTERS>
        <MOVERS>
                <Docking name=dock_low score_low=score_docking_low score_high=score12 fullatom=0 local_refine=0 optimize_fold_tree=1 conserve_foldtree=1 design=0
task_operations=ifcl/>
                <Docking name=dock_high score_low=score_docking_low score_high=score12 fullatom=1 local_refine=1 optimize_fold_tree=1 conserve_foldtree=1 design=0
task_operations=ifcl/>
                <PackRotamersMover name=pr scorefxn=score12/>
                <MinMover name=min scorefxn=score12 chi=1 bb=1 jump=1 tolerance=0.01/>

        </MOVERS>
        <APPLY_TO_POSE>
        </APPLY_TO_POSE>
        <PROTOCOLS>
          <Add mover_name=dock_low/>
          <Add mover_name=pr/>
          <Add mover_name=min/>
          <Add mover_name=dock_high/>
          <Add mover_name=pr/>
          <Add mover_name=min/>
        </PROTOCOLS>
</dock_design>
```

Using Protein Docking

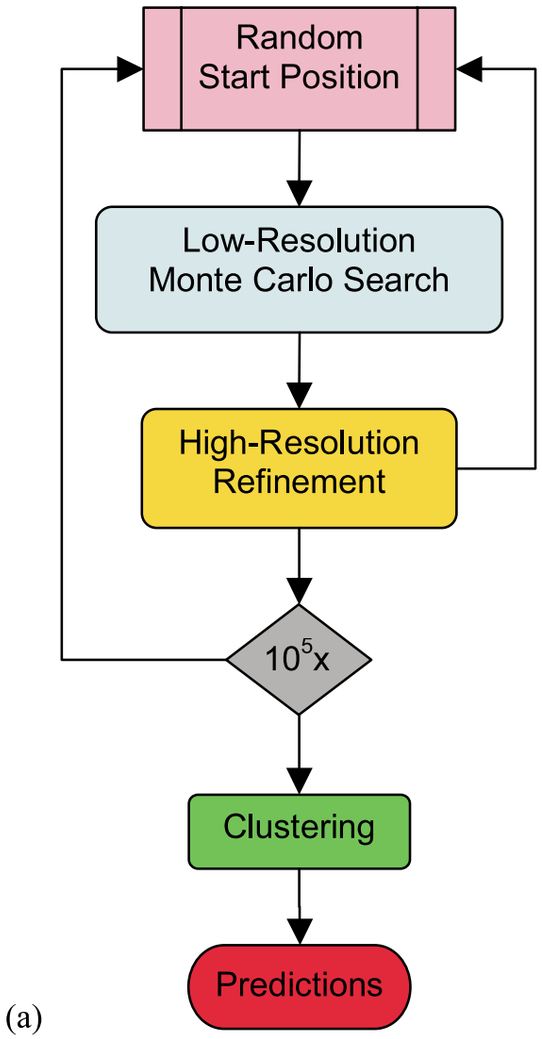
Rosetta Scripts - XML

- Initialize from command line - This tells the scripter that we will use options from the command line that have not yet been hardcoded into the scripter.
- **Movers** - The part of the scripter that does "things" to the pose
  - **Dock Low** - This is the low resolution docking mover. It converts to centroid mode, and begins searching the docking space of both binding partners.
  - **Dock High** - This is the high resolution refinement, after we have found the binding site, slight perturbations in full atom mode minimize the energy.
  - **Repack** - Repacks complex with rotamers form rotamer libraries.
  - **Minimize** - Gradient based energy minimization of entire complex.

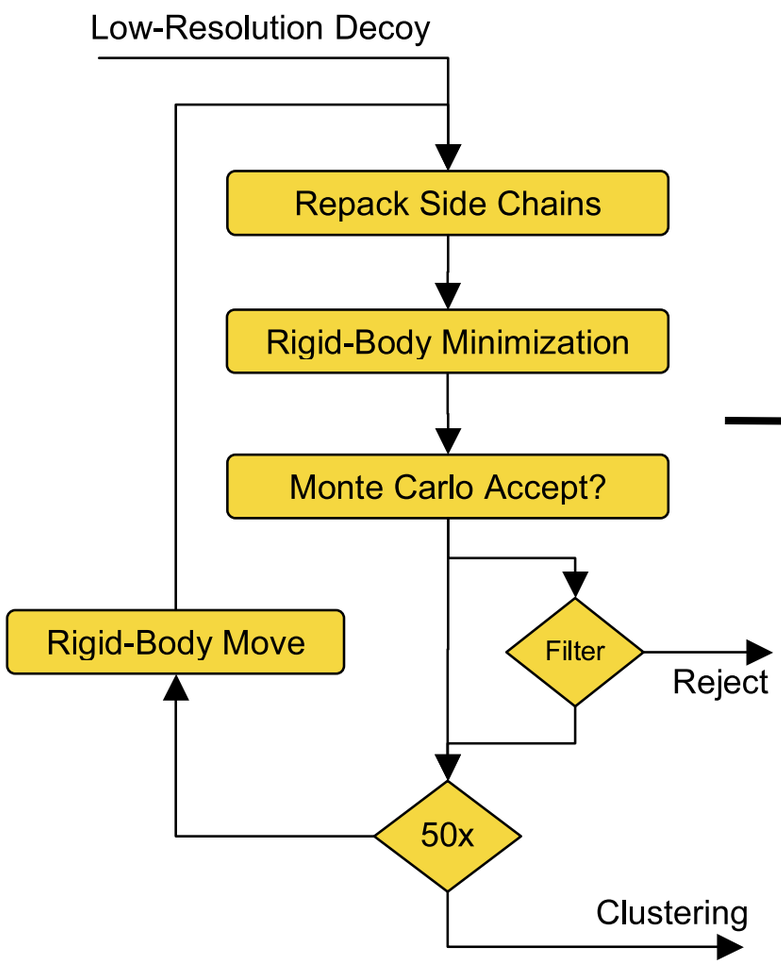
Movers are modular. Putting in a specific order (XML Protocol) mimics the RosettaDock Protocol

Using Protein Docking

Rosetta Scripts - XML



(a)



(b)

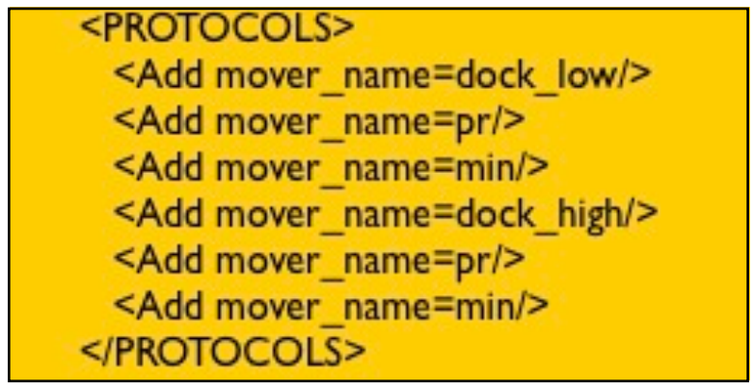


Figure 1. Docking protocol. (a) Process flowchart; (b) detail of the refinement stage.

Analysis of Results

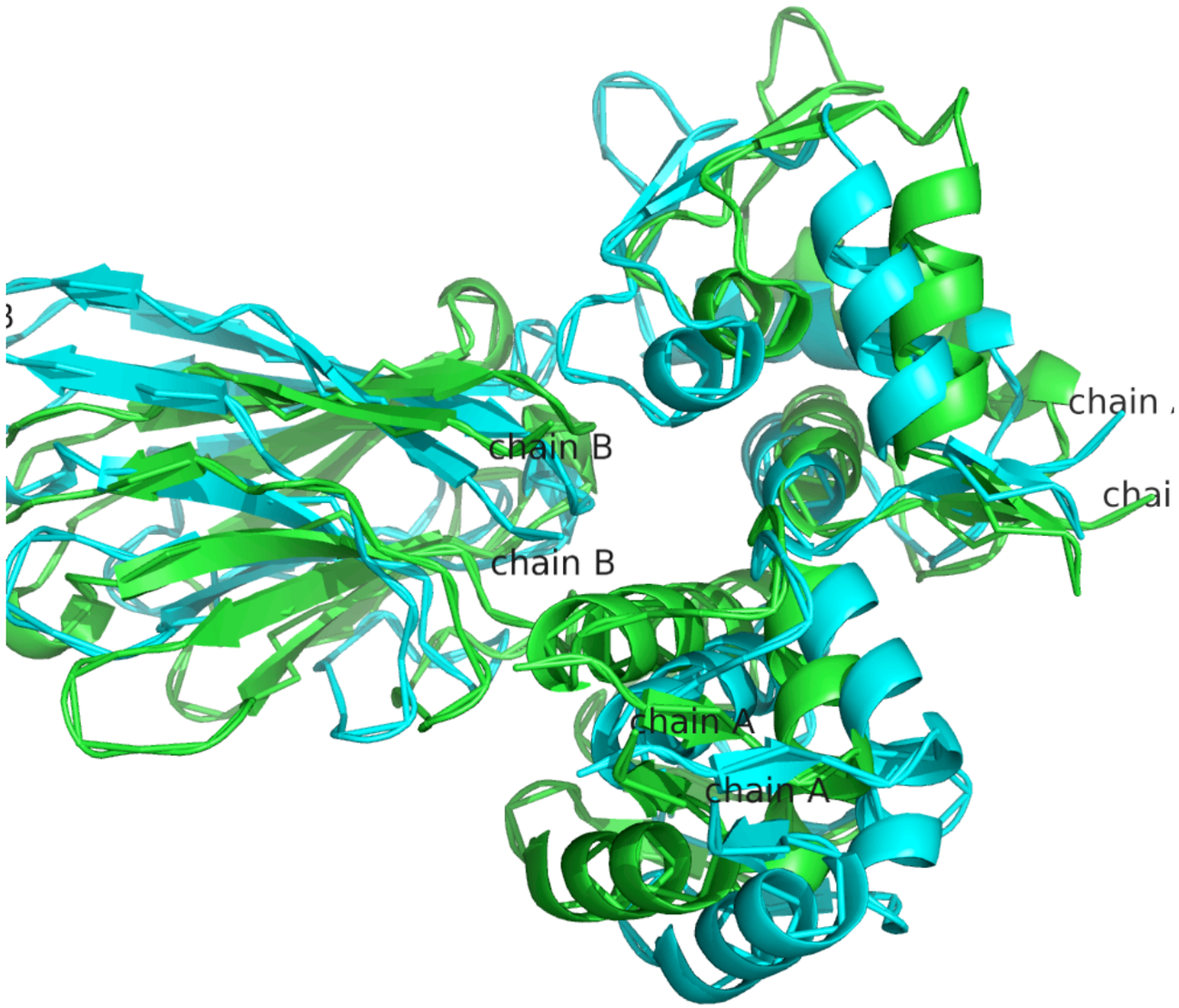
Using score_vs_rmsd.py should create a energy funnel, ie. the lowest energy model should be closest to native structure.

```
$ROSETTA_SCRIPTS/score_vs_rmsd.py --native 2VYR_input.pdb --table=ex.out --term=total
2VYR_input_low_reso_000*.pdb
```

file	score	RMSD
2VYR_input_low_reso_0001.pdb	-719.24	19.5958229731
2VYR_input_low_reso_0002.pdb	-750.993	19.6144996137

Only 2 files gives large RMSD...we need many models for global searches. Usually on the order of 10,000 decoys in order to create an energy funnel

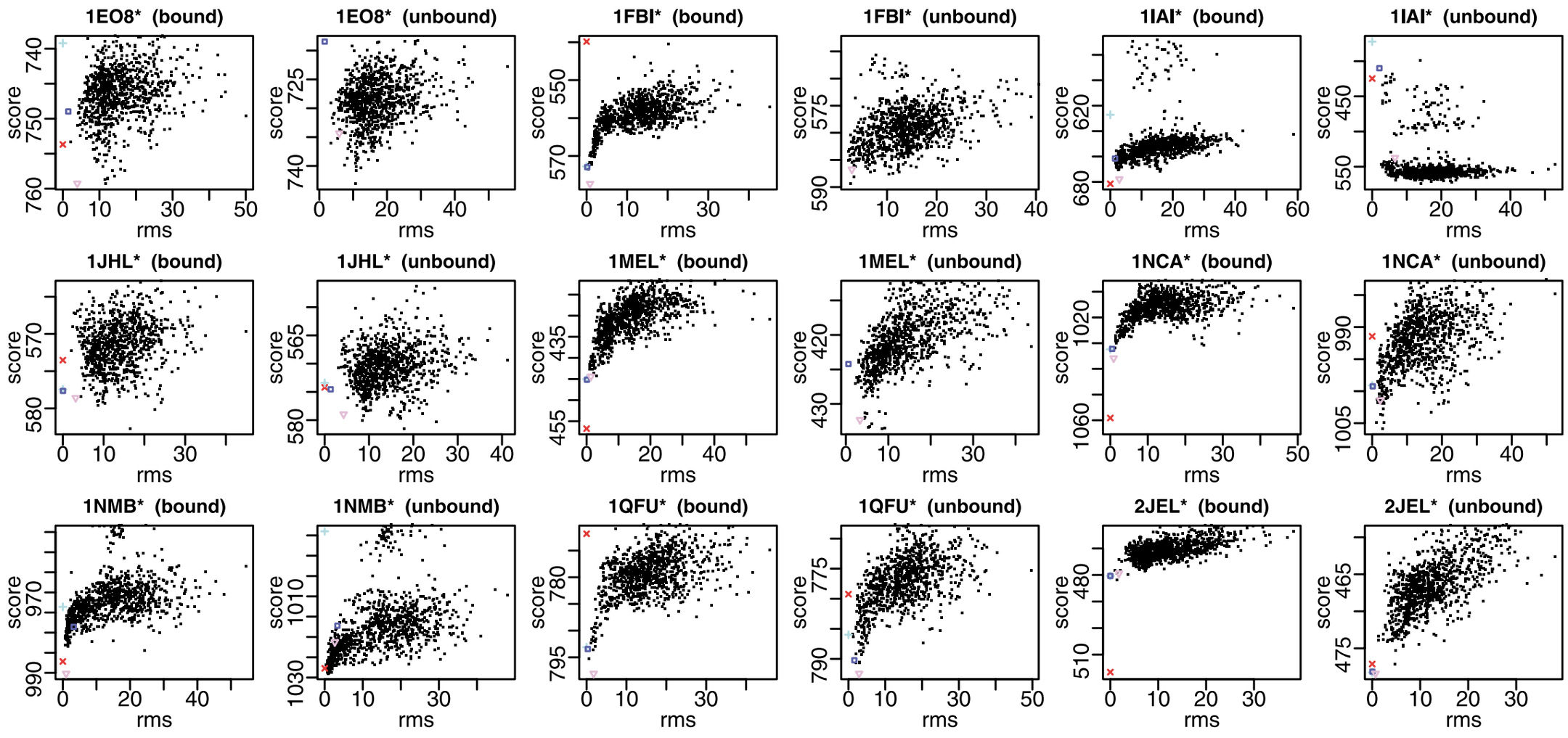
Analysis of Results



Epitope found even in one global search. Green is native structure, blue is modeled.

Analysis of Results

10,000 models provides us with a an energy vs. RMSD funnel.



Gray et. al. 2005

Lowest Score is closest to native structure when an energy funnel exists.

Resources

1. Chaudhury, S. and J.J. Gray, *Conformer selection and induced fit in flexible backbone protein-protein docking using computational and NMR ensembles.* J Mol Biol, 2008. **381**(4): p. 1068-87.
2. Daily, M.D., et al., *CAPRI rounds 3-5 reveal promising successes and future challenges for RosettaDock.* Proteins, 2005. **60**(2): p. 181-6.
3. Gray, J.J., *High-resolution protein-protein docking.* Curr Opin Struct Biol, 2006. **16**(2): p. 183-93.
4. **Gray, J.J., et al., *Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations.* J Mol Biol, 2003. 331(1): p. 281-99.**
5. Lyskov, S. and J.J. Gray, *The RosettaDock server for local protein-protein docking.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W233-8.
6. Sircar, A., et al., *A generalized approach to sampling backbone conformations with RosettaDock for CAPRI rounds 13-19.* Proteins, 2010. **78**(15): p. 3115-23.
7. Sivasubramanian, A., et al., *Toward high-resolution homology modeling of antibody Fv regions and application to antibody-antigen docking.* Proteins, 2009. **74**(2): p. 497-514.

Resources

SnugDock - Comparative modeling of Antibodies + Docking

http://www.ploscompbiol.org/article/info:doi%2F10.1371%2Fjournal.pcbi.1000644

RosettaDock - Automated Docking

http://rosettadock.graylab.jhu.edu/

RosettaAntibody - Automated Antibody Homology Modeling

http://antibody.graylab.jhu.edu/