

Protein-Protein Docking in ROSETTA

ROSETTA Workshop
Membrane Proteins
May 2014

Alyssa Lokits, Steven Combs, & Alberto Cisneros

Outline

- ▶ Overview
 - ▶ Literature and citations
 - ▶ General Workflow & Protocol
 - ▶ RosettaDock Global Search (low-res docking)
 - ▶ Rosetta docking refinement procedure (high-res docking)
- ▶ The docking XML file
 - ▶ Protocol generation
 - ▶ Movers and behavior
- ▶ Options files
 - ▶ Define input, output, and additional considerations
- ▶ Span files
 - ▶ Define membrane regions
- ▶ Application
- ▶ Analysis of results

Primary Citations

doi:10.1016/S0022-2836(03)00670-3

J. Mol. Biol. (2003) 331, 281–299

JMB

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®



Protein–Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations

Jeffrey J. Gray, Stewart Moughon, Chu Wang, Ora Schueler-Furman
Brian Kuhlman, Carol A. Rohl and David Baker*

OPEN ACCESS Freely available online

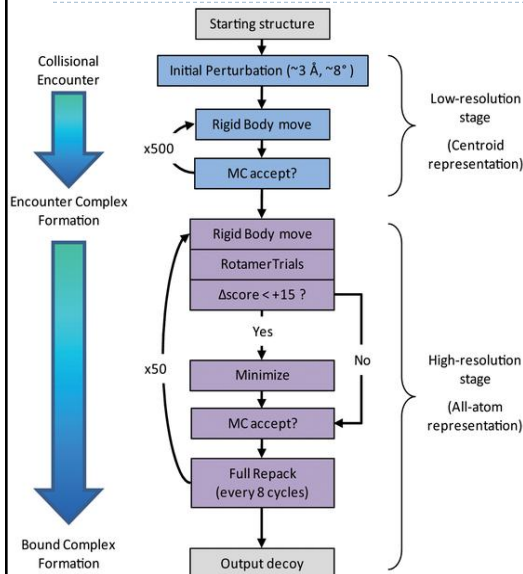
PLoS one

Benchmarking and Analysis of Protein Docking Performance in Rosetta v3.2

Sidhartha Chaudhury¹, Monica Berrondo², Brian D. Weitzner², Pravin Muthu², Hannah Bergman²,
Jeffrey J. Gray^{2*}

¹ Program in Molecular Biophysics, Johns Hopkins University, Baltimore, Maryland, United States of America, ² Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland, United States of America

Overview - Protein Docking



- ▶ Mimics natural process of protein interaction
 1. Centroid-based coarse grain stage
 2. Fine conformational search with side chain flexibility
- ▶ Additional Inputs:
 1. Incorporate experimental data to decrease conformational search space
 2. Cluster, re-dock, continue refinement

General Docking Protocol

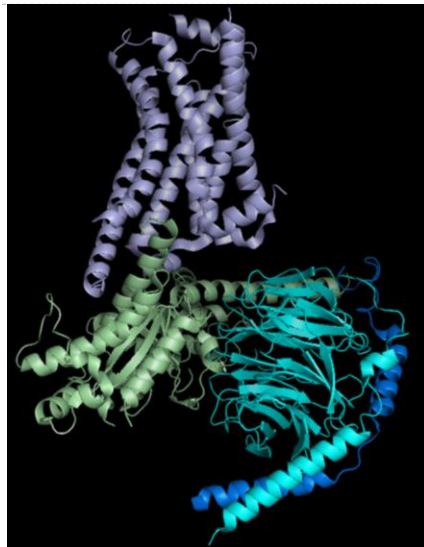
1. Prepare input files
 - ▶ Template PDB
 - ▶ RosettaScripts XML File
 - ▶ Options File
 - ▶ Span File
2. Perform docking calculations
3. Analyze results
 - ▶ Plot Score vs. RMSD
 - ▶ Examine interface character



Example System:

β -2 Adrenergic Receptor (GPCR) and $G\alpha_s\beta\gamma$

- ▶ PDBID 3SN6
 - ▶ 3.0 Å res.
- ▶ Class A GPCR
 - ▶ Agonists: adrenaline (epinephrine)
 - ▶ Vital for flight or fight response
- ▶ $G\alpha_s\beta\gamma$
 - ▶ Alpha subunit activation results in increased cAMP production
 - ▶ Activates PKA and PP2B
 - ▶ β 2AR can also interact with $G\alpha_i$



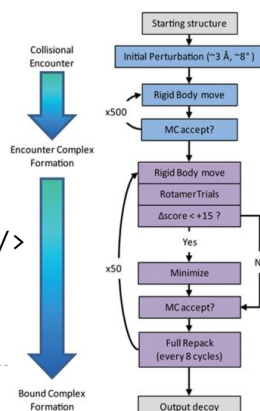
Docking XML file

Please open the docking_full.xml file

```
gedit ~/rosetta_workshop/tutorials/protein_protein_docking/
membrane/input_files/docking_full.xml
```

Where should you start?

```
<PROTOCOLS>
  <Add mover=dock_low/>
  <Add mover=srsc />
  <Add mover=dock_high />
  <Add mover=minimize_interface />
</PROTOCOLS>
</dock_design>
```



Docking movers

- `<Docking name="string" score_low=(string) score_high=(string) fullatom=(bool) local_refine=(bool) jumps=(Integer vector) optimize_fold_tree=(bool) conserve_foldtree=(bool) design=(bool) ignore_default_docking_task=(bool) task_operations=(comma-separated list)/>`
 - ▶ `score_low`: scorefxn for centroid-level docking
 - ▶ `score_high`: scorefxn for full atom docking
 - ▶ `fullatom`: if true, do full atom docking
 - ▶ `local_refine`: if true, skip centroid
 - ▶ `jumps`: where should we carry out rigid body motions?
 - ▶ `optimize_fold_tree`: should DockingProtocol make the fold tree for this pose?
 - ▶ `conserve_foldtree`: should DockingProtocol reset the fold tree to the input one after it is done?
 - ▶ `design`: interface design for all chains downstream of the rb_jump
 - ▶ `ignore_default_docking_task`: allows you to ignore the default docking task and only use the ones defined in your task_operations section
 - ▶ `task_operations`= define tasks outside of default -can use to prevent residues from repacking (prfrp) etc.

Docking movers

```
> <Docking name=dock_low score_low=score_docking_low
score_high=score12 fullatom=0 local_refine=0
optimize_fold_tree=1 conserve_foldtree=0
ignore_default_docking_task=0 design=0
task_operations=ifc1 jumps=1/>
```

```
> <Docking name=dock_high score_low=score_docking_low
score_high=score12 fullatom=1 local_refine=1
optimize_fold_tree=1 conserve_foldtree=0 design=0
task_operations=ifc1 jumps=1/>
```

****fullatom:** true for dock_high, not for dock_low

****local_refine:** true for dock_high, not for dock_low

```
> <SaveAndRetrieveSidechains name=srsc allsc=0 />
  ▪ Speeds the move from centroid to full atom mode
```



Minimization movers

These movers are part of the minimization protocol put together in the minimize_interface ParsedProtocol

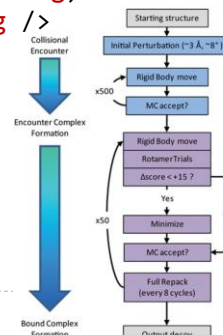
```
> <PackRotamersMover name="string" scorefxn=(string)
task_operations=(comma-separated list)/>
```

```
> <TaskAwareMinMover name="string" tolerance=(integer)
task_operations=(comma-separated list) type=(string)
chi=(bool) bb=(bool) jump=(int) scorefxn=string />
```

Type: determines which minimization algorithm

Chi: minimize sidechains?

Bb: minimize backbone?



Minimization movers

- ▶ `<PackRotamersMover name=soft_rp scorefxn=soft_rep task_operations=rtiv,rtr,ifcl />`
- ▶ `<TaskAwareMinMover name=soft_min tolerance=0.001 task_operations=ifcl,rtiv type=dfpmin_armijo_nonmonotone chi=1 bb=1 jump=1 scorefxn=soft_rep />`
- ▶ The soft-rep score function specifies the use of an alternate weight set that dampens part of the Lennard-Jones potential. The idea is to allow small clashes that can be repaired by backbone relaxation. This option is often used during fixed backbone operations since the backbone cannot relax to accommodate small steric clashes



Minimization after docking

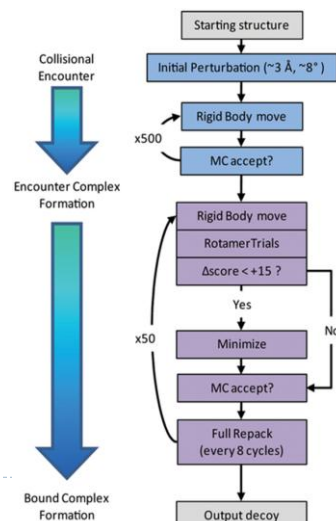
`<ParsedProtocol name=minimize_interface mode=sequence >`

```

<Add mover=soft_rp />
<Add mover=soft_min_sc />
<Add mover=soft_rp />
<Add mover=soft_min />
<Add mover=soft_rp />
<Add mover=min_sc />
<Add mover=soft_rp />
<Add mover=min />
<Add mover=rp />
<Add mover=min_sc />
<Add mover=rp />
<Add mover=min />

```

`</ParsedProtocol>`



TaskOperations

- **<TASKOPERATIONS>**
 - `<InitializeFromCommandline name=ifcl/>`
 - `<RestrictToRepacking name=rtr />`
 - `<RestrictToInterfaceVector name=rtiv`
 - `chain1_num=1 chain2_num=2`
 - `CB_dist_cutoff=10.0`
 - `nearby_atom_cutoff=5.5`
 - `vector_angle_cutoff=75`
 - `vector_dist_cutoff=9.0 />`
- This mover restricts the task to residues defined as the interface.
 - Calculates residues at an interface between to chains or jump based on CB_dist_cutoff.
 - Confirms interface residues by iterating through all SC atoms and checking that their distance is less than the nearby_atom_cutoff.
 - Secondary classification of interface residues defined by vector_angle_cutoff and vector_distance_cutoff

Options file

Docking options

<code>-docking</code>	<code>-partners R_A</code>	-dock chains R as a single rigid body to chain A
	<code>-dock_pert 3 8</code>	-perform perturbations during docking of 3 angstroms and 8 degrees
	<code>-dock_mcm_trans_magnitude .1</code>	-specify the translational magnitude during the refinement steps
	<code>-dock_mcm_rot_magnitude 5.0</code>	-specify the rotational magnitude during the refinement steps
<code>-s model_r_a_renumbered.pdb</code>		-specify the input pdb
<code>-run:max_retry_job 50</code>		-try again this many times a mover fails
<code>-use_input_sc</code>		-add the side chains from the input structure to the rotamer library
<code>-ex1</code>		-add extra rotamers for chi-2 to the rotamer library
<code>-ex2</code>		-add extra rotamers for chi-1 to the rotamer library
<code>-corrections:score:use_bicubic_interpolation true</code>		-use bicubic interpolation of phi/psi bins
<code>-out</code>		
	<code>-file</code>	
	<code>-scorefile docking.fasc</code>	-name of model score file
	<code>-silent docking.silent</code>	-name of the model silent file (need for scoring_
	<code>-silent_print_all_score_headers true</code>	- print score headers for each model in silent

Command Line

`-parser:protocol docking.xml` -specify the xml protocol

`-restore_pre_talaris_2013_behavior` -allows for use of soft_rep, not always needed

Other docking options

-docking:spin - spin one partner around an axis in between the two docking partners

-docking:randomize1 - 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.



SPAN files

- Span files are required for membrane proteins only
 - Include information about residues interacting with the lipid membrane
 - If no experimental data exists, can identify membrane spanning regions using programs such as OCTOPUS
 - GPCRs have 7 TMD
 - [model_r_a_aligned.spanfile](#) for tutorial
-



Perform docking

- Run the RosettaScripts application:

```
~/rosetta_workshop/Rosetta/main/source/bin/rosetta_scripts.default.linuxgccrelease @docking.options -database  
~/rosetta_workshop/Rosetta/main/database/ -parser:protocol  
docking_full.xml -out:suffix _full -nstruct 50 >&  
docking_full.log &
```

- Compare results for full docking protocol and docking using just the refinement protocol (dock_high)
 - Remove < in front of dock_low mover in protocols section to turn it off
 - Add -out:suffix _refine_only to the commandline.
 - Outputs a pdb and score file with _refine_only appended to name
 - Run docking again with modified xml file and command line

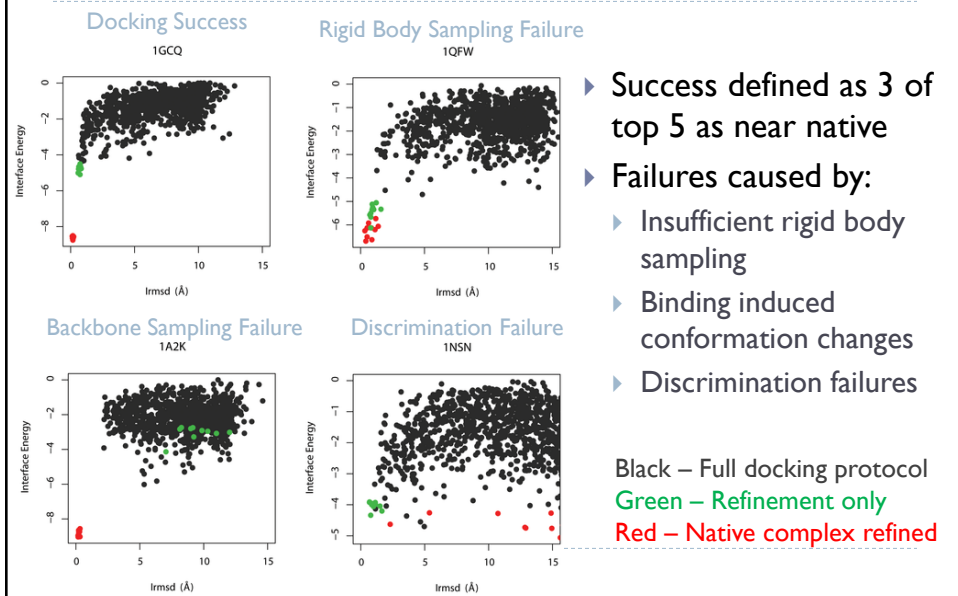
➤

Perform analysis

- Score vs. RMSD plots to identify a docking funnel
 - In experimental studies, the 'native structure' is a good scoring structure, one that fulfills experimental restraints, or a cluster center
- It may be necessary to cluster and dock again, using the top cluster centers as starting templates. This will help to drive the models into an energy well. The 'correct' model will continue down into an energy well while incorrect models will stall.
- Look at the structures
 - Unsatisfied polar residues
 - Binding density
 - Average Degree of Residue Burial

➤ Fleishman, S. J.; Baker, D., Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell* **2012**, *149* (2), 262-73
Fleishman, S. J.; Whitehead, T. A.; Strauch, E. M.; Corn, J. E.; ... ; Baker, D., Community-wide assessment of protein-interface modeling suggests improvements to design methodology. *Journal of molecular biology* **2011**, *414* (2), 289-302.

Score vs. RMSD plots – identifying a docking funnel



Reference List

▶ Docking

- ▶ Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D., Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *Journal of molecular biology* **2003**, 331 (1), 281-99.
- ▶ Daily, M. D.; Masica, D.; Sivasubramanian, A.; Somarouthu, S.; Gray, J. J., CAPRI rounds 3-5 reveal promising successes and future challenges for RosettaDock. *Proteins* **2005**, 60 (2), 181-6.
- ▶ Chaudhury, S.; Sircar, A.; Sivasubramanian, A.; Berrondo, M.; Gray, J. J., Incorporating biochemical information and backbone flexibility in RosettaDock for CAPRI rounds 6-12. *Proteins* **2007**, 69 (4), 793-800.
- ▶ Sircar, A.; Chaudhury, S.; Kilambi, K. P.; Berrondo, M.; Gray, J. J., A generalized approach to sampling backbone conformations with RosettaDock for CAPRI rounds 13-19. *Proteins* **2010**, 78 (15), 3115-23.
- ▶ Sivasubramanian, A.; Sircar, A.; Chaudhury, S.; Gray, J. J., Toward high-resolution homology modeling of antibody Fv regions and application to antibody-antigen docking. *Proteins* **2009**, 74 (2), 497-514.
- ▶ Sircar, A.; Gray, J. J., SnugDock: paratope structural optimization during antibody-antigen docking compensates for errors in antibody homology models. *PLoS computational biology* **2010**, 6 (1), e1000644

Reference list continued

▶ Design

- ▶ Fleishman, S. J.; Whitehead, T. A.; Ekiert, D. C.; Dreyfus, C.; Corn, J. E.; Strauch, E. M.; Wilson, I. A.; Baker, D., Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **2011**, 332 (6031), 816-21.
- ▶ Fleishman, S. J.; Corn, J. E.; Strauch, E. M.; Whitehead, T. A.; Karanicolas, J.; Baker, D., Hotspot-centric de novo design of protein binders. *Journal of molecular biology* **2011**, 413 (5), 1047-62.
- ▶ Der, B. S.; Machius, M.; Miley, M. J.; Mills, J. L.; Szyperski, T.; Kuhlman, B., Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. *Journal of the American Chemical Society* **2012**, 134 (1), 375-85.
- ▶ Mandell, D. J.; Kortemme, T., Computer-aided design of functional protein interactions. *Nature chemical biology* **2009**, 5 (11), 797-807.
- ▶ Joachimiak, L. A.; Kortemme, T.; Stoddard, B. L.; Baker, D., Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. *Journal of molecular biology* **2006**, 361 (1), 195-208.

▶ Analysis

- ▶ Fleishman, S. J.; Baker, D., Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell* **2012**, 149 (2), 262-73.
- ▶ Fleishman, S. J.; Khare, S. D.; Koga, N.; Baker, D., Restricted sidechain plasticity in the structures of native proteins and complexes. *Protein science : a publication of the Protein Society* **2011**, 20 (4), 753-7.
- ▶ Fleishman, S. J.; Whitehead, T. A.; Strauch, E. M.; Corn, J. E.; ... ; Baker, D., Community-wide assessment of protein-interface modeling suggests improvements to design methodology. *Journal of molecular biology* **2011**, 414 (2), 289-302.