

## Protein-Protein Docking in Rosetta 3.5



## Why Do Protein-Protein Docking Algorithms Exist?

- Protein – protein interactions are integral to:
  - Protein localization
  - Allosteric regulation
  - Gene regulation
  - Signal transduction
  - Pathogen clearance (antibody-antigen)
- Interface Design

## Outline

- RosettaDock algorithm
  - Coarse, rigid-body docking
  - Full-atom refinement with side-chain flexibility
- Tutorial
  - Template Preparation
  - RosettaDock via RosettaScripts
  - Model Analysis
- References and Resources
  - RosettaDock, SnugDock, FlexPepDock

## Primary RosettaDock Citations

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

J. Mol. Biol. (2003) 331, 281–299

**JMB**Available online at [www.sciencedirect.com](http://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

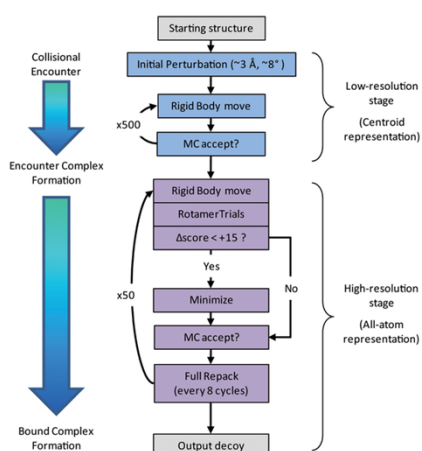


### Benchmarking and Analysis of Protein Docking Performance in Rosetta v3.2

Sidhartha Chaudhury<sup>1</sup>, Monica Berrondo<sup>2</sup>, Brian D. Weitzner<sup>2</sup>, Pravin Muthu<sup>2</sup>, Hannah Bergman<sup>2</sup>,  
Jeffrey J. Gray<sup>2\*</sup>

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

## Overview - Protein Docking



- Mimics natural process of protein interaction
- Coarse rigid-body docking
  - Low-resolution
  - Centroid-based
- Refinement Stage
  - Fine conformational search
  - High-resolution
  - Full atom (side chain flexibility)
- Hints:
  - Incorporate experimental data to restrict conformational search space
  - Use a complementary algorithm for global conformational search
  - Cluster, re-dock, continue refinement

Chaudhury *et al*, PlosONE, 2011

## Docking Score Function

### Centroid docking score function

| Component             | name               | weight |
|-----------------------|--------------------|--------|
| Contact               | interchain_contact | 2.0    |
| Bumps                 | interchain_vdw     | 1.0    |
| Environment           | interchain_env     | 1.0    |
| Pair-wise interaction | interchain_pair    | 1.0    |

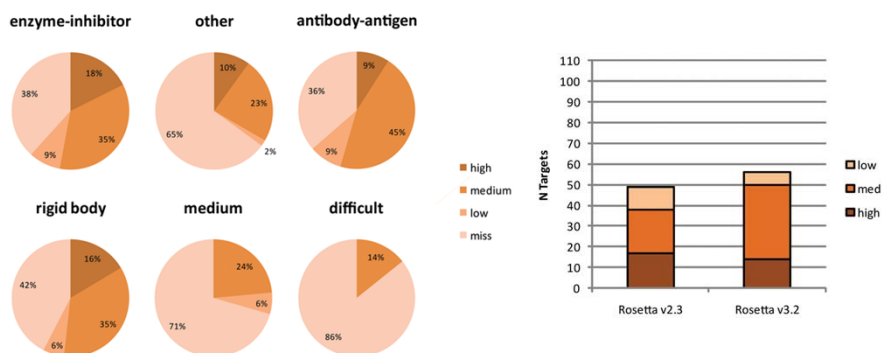
### All-atom docking score function

| Component                  | name                                            | weight |
|----------------------------|-------------------------------------------------|--------|
| Van der Waals (attractive) | fa_atr                                          | 0.338  |
| Van der Waals (repulsive)  | fa_rep                                          | 0.044  |
| Dunbrack side-chain        | fa_dun                                          | 0.036  |
| Solvation                  | fa_sol                                          | 0.242  |
| Pair-wise interaction      | fa_pair                                         | 0.164  |
| Hydrogen Bond              | hbond_lr_bb, hbond_sr_bb, hbond_bb_sc, hbond_sc | 0.245  |
| Electrostatic              | hack_elec                                       | 0.026  |

doi:10.1371/journal.pone.0022477.t001

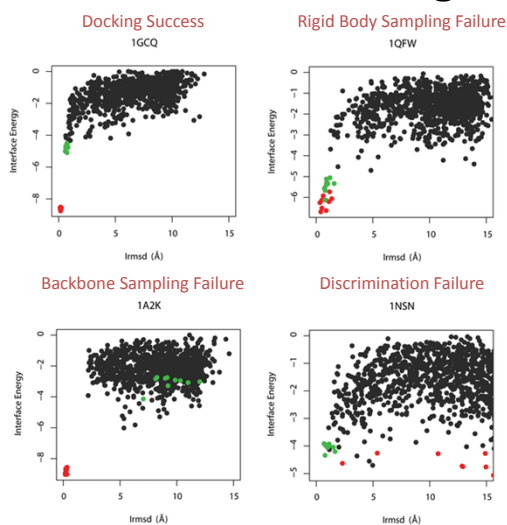
Chaudhury *et al*, PlosONE, 2011

## Protein Docking Benchmark 3.0 shows successes by interaction class and difficulty



Chaudhury *et al*, PlosONE, 2011

## Score vs. RMSD plots – identifying a docking funnel



- Success defined as 3 of top 5 as near native
- Failures caused by:
  - Insufficient rigid body sampling
  - Binding induced conformation changes
  - Discrimination failures

Gray – Full docking protocol  
 Green – Refinement only  
 Red – Native complex refined

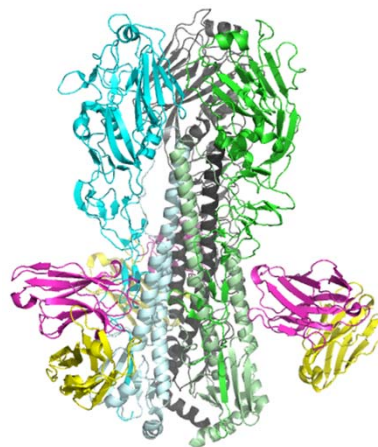
Chaudhury *et al*, PlosONE, 2011

## General Docking Protocol

1. Template Preparation
  - Clean each docking partner and orient for docking
2. RosettaDock via RosettaScripts
  - Prepare RosettaScripts XML protocol
3. Analyze results
  - Plot scores vs. RMSD
  - Shape complementarity
  - Physical and chemical characteristics of interface

## Example System: Influenza Hemagglutinin bound by mAb CR6261

- PDBID 3GBN, 3GBM
  - 2.2 & 2.7 Å res.
- Influenza HA Antigen
  - H1N1 1918 Flu Pandemic
  - H5N1 Bird Flu
  - 2 chains: HA1, HA2
- CR6261
  - Human derived monoclonal antibody
  - Stem-binding antibody
  - Binds many Group 1 HAs
  - 2 chains: Heavy and Light



Ekiert *et al*, Science, 2009

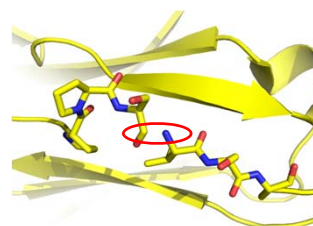
## Template Preparation

1. Prepare the proteins for docking
  - a. Download 3gbn and 3gbm from rcsb.org
  - b. Clean the docking partners
    - i. Remove unwanted atoms and extraneous information
      - Symmetry factors, Crystallization conditions, etc
      - Waters, undesired HETATMs
        - » Docking with a HETATM has been shown to improve results
    - ii. Delete un-resolved regions of extraneous domains
    - iii. Renumber the PDBs from 1 to *N*



## Template Preparation (cont.)

- c. Close small chain breaks
  - i. Identify chain breaks
  - ii. Create LOOPS file
  - iii. Run Rosetta loopmodel application
- d. Repack or relax template structures to generate conformation diversity prior to docking
  - Take best *X* models for each partner
  - Dock  $N_{\text{partner1}} \times M_{\text{partner2}}$  conformations (see matrix at right)
  - Certain combinations have better binding energies (green at right)
- e. Orient docking partners using available knowledge
  - Antibody CDRs toward antigen
  - Experimental data



|           |   | Partner 1 |   |   |   |   |
|-----------|---|-----------|---|---|---|---|
|           |   | 1         | 2 | 3 | 4 | 5 |
| Partner 2 | 1 |           |   |   |   |   |
|           | 2 |           |   |   |   |   |
|           | 3 |           |   |   |   |   |
|           | 4 |           |   |   |   |   |
|           | 5 |           |   |   |   |   |

## Prepare RosettaScripts XML Protocol

```
<ROSETTASCRIPTS>
  <SCOREFXNS>
</SCOREFXNS>
  <TASKOPERATIONS>
    <InitializeFromCommandline name=ifcl/>
    <RestrictToRepacking name=rtr />
    <RestrictToInterfaceVector name=rtiv chain1_num=1,2
      chain2_num=3,4 CB_dist_cutoff=10.0
      nearby_atom_cutoff=5.5 vector_angle_cutoff=75
      vector_dist_cutoff=9.0 />
    <PreventResiduesFromRepacking name=prfrp residues=11,41,345 />
  </TASKOPERATIONS>
<FILTERS>
</FILTERS>
```

## Movers for docking

```
<MOVERS>
  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=ifcl,prfrp 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=ifcl,prfrp jumps=1/>

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

## Movers for minimization

### MINIMIZATION MOVERS

These movers are part of the minimization protocol put together in the minimize\_iface ParsedProtocol

```
<PackRotamersMover name=soft_rp scorefxn=soft_rep
  task_operations=rtiv,rtr,ifcl,prfrp />
<PackRotamersMover name=rp scorefxn=score12
  task_operations=rtiv,rtr,ifcl,prfrp />
<TaskAwareMinMover name=soft_min tolerance=0.001 chi=1 bb=1 jump=1
  task_operations=ifcl,rtiv type=dfpmin_armijo_nonmonotone
  scorefxn=soft_rep />
<TaskAwareMinMover name=soft_min_sc tolerance=0.001 chi=1 bb=0 jump=0
  task_operations=ifcl,rtiv type=dfpmin_armijo_nonmonotone
  scorefxn=soft_rep />
<TaskAwareMinMover name=min_sc tolerance=0.0001 chi=1 bb=0 jump=0
  task_operations=ifcl,rtiv type=dfpmin_armijo_nonmonotone
  scorefxn=score12 />
<TaskAwareMinMover name=min tolerance=0.0001 chi=1 bb=1 jump=1
  task_operations=ifcl,rtiv type=dfpmin_armijo_nonmonotone
  scorefxn=score12 />
```

## XML-file continued

This minimization protocol is adapted from Whitehead, Fleishman, ... Baker. See the XML scripts available in the Supp Mat.

```
<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>
```

## XML-file continued

```

<APPLY_TO_POSE>
</APPLY_TO_POSE>
<PROTOCOLS>
  <Add mover=dock_low/>
  <Add mover=srsc />
  <Add mover=dock_high />
  <Add mover=minimize_interface />
</PROTOCOLS>
</ROSETTASCRIPTS>

```

## Options file

```

-database rosetta-3.5/rosetta_database -specify the database location
-parser:protocol docking_full.xml -specify the xml protocol
-s 3gbm_HA_3gbn_Ab.pdb -specify the input pdb
-run:max_retry_job 10 -try a mover again this many times if mover fails
-jd2:ntrials 50 -if a structure fails to output, how many times to try again
-docking
  -partners AB_HL -dock chains H & L as a single rigid body to chains A & B
  -dock_pert 3 8 -perform perturbations during coarse docking of 3 Å and 8°
  -dock_mcm_trans_magnitude 0.1 -translational magnitude during refinement
  -dock_mcm_rot_magnitude 3 -rotational magnitude during refinement
-nstruct 50 -the number of models to generate
-use_input_sc - add the side chains from the input structure to the rotamer library
-ex1 -add extra rotamers for chi-2 to the rotamer library
-ex2 -add extra rotamers for chi-1 to the rotamer library
-corrections:score:use_bicubic_interpolation true -use bicubic instead of linear interpolation
of the phi/psi bins
-out:file
  -scorefile docking.fasc -write model scores to this file
  -silent docking.fasc -write models to a silent file with this name
  -silent_print_all_score_headers -print score headers for each model in the silent file

```

## 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:randomize2
- docking:docking\_centroid\_outer/inner - repeats the number of docking moves in the coarse, centroid mode.

See manual or options page for other options

## Perform docking

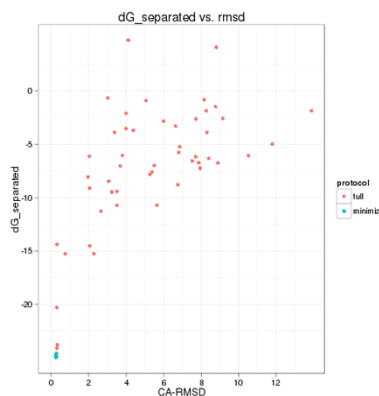
- Run the RosettaScripts application:  

```
rosetta-3.5/rosetta_source/bin/rosetta_scripts.default.linux
gccrelease @docking.options -parser:protocol
docking_full.xml -out:suffix _full -nstruct 50 > &
docking_full.log &
```
- Compare results of the full docking protocol with just the minimization protocol (docking\_minimize.xml)  

```
rosetta-3.5/rosetta_source/bin/rosetta_scripts.default.linux
gccrelease @docking.options -parser:protocol
docking_minimize.xml -out:suffix _minimize -nstruct
10 > & docking_minimize.log &
```

## Identifying the 'Best' Model

- Score vs. RMSD-to-native plots
  - Look for a docking funnel
  - In experimental studies, the 'native structure' is a good scoring model, a cluster center, or a model that fulfills experimental restraints
- Cluster and dock again
  - Use the top cluster centers as templates
  - Increased regional sampling leads to greater precision in scoring
  - 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
  - Shape complementarity
  - Unsatisfied buried polar residues
    - High energy
  - Binding density (ddG/dSASA)
  - 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.

## Model Analysis with RosettaScripts

- Analysis command
 

```
rosetta-3.5/rosetta_source/bin/rosetta_scripts.default.linuxgccrelease
@docking_analysis.options -database rosetta-3.5/rosetta_database/ >& docking_analysis.log &
```
- Options file
 

```
-parser:protocol docking_analysis.xml
-use_input_sc
-ex1
-ex2
-corrections:score:use_bicubic_interpolation true
-in:file:silent docking.silent -read structures from this silent file
-out:file:scorefile docking_analysis.csv -name of scorefile
-native 3gbm_native.pdb -use this model for comparisons
-out:no_nstruct_label -do not add _0001 to model name
```

## Docking Analysis XML Protocol

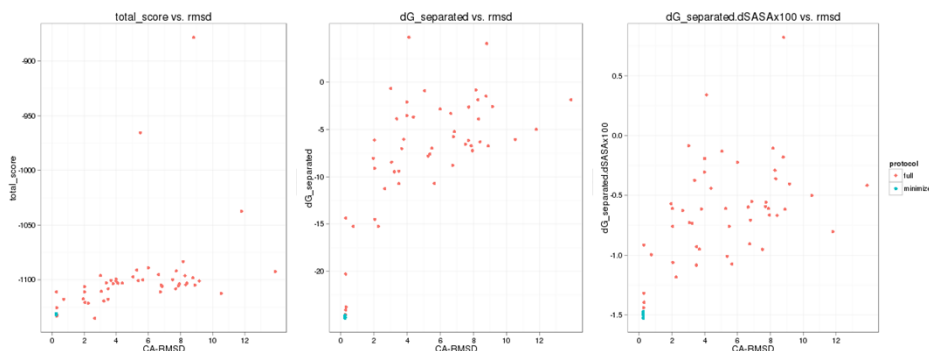
```
<dock_design>
  <SCOREFXNS>
  </SCOREFXNS>
  <TASKOPERATIONS>
    <RestrictToInterfaceVector name=rtiv chain1_num=1,2
      chain2_num=3,4 CB_dist_cutoff=10.0 nearby_atom_cutoff=5.5
      vector_angle_cutoff=75 vector_dist_cutoff=9.0 />
  </TASKOPERATIONS>
  <MOVERS>
    <InterfaceAnalyzerMover name=iface_analyzer
      scorefxn=score12 packstat=0 pack_input=0
      pack_separated=1 fixedchains=A,B tracer=0 />
  </MOVERS>
  <FILTERS>
    <AverageDegree name=avg_degree threshold=0
      distance_threshold=10 task_operations=rtiv />
    <Rmsd name=rmsd superimpose=1 threshold=2500 >
      <span begin_res_num=11 end_res_num=13 />
    </Rmsd>
  </FILTERS>
  <APPLY_TO_POSE>
  </APPLY_TO_POSE>
  <PROTOCOLS>
    <Add mover=iface_analyzer />
    <Add filter=rmsd />
  </PROTOCOLS>
</dock_design>
```

### Interface Analyzer Scores

- Normal Score Terms
  - fa\_atr, fa\_rep, etc.
- Binding related
  - Binding energy
  - dSASA
  - Binding density
- Unsatisfied Hydrogen Bonds
- Number of Interface Residues
- Interface Energy per Interface Residue
- Energy of each partner
  - Also normalized per residue

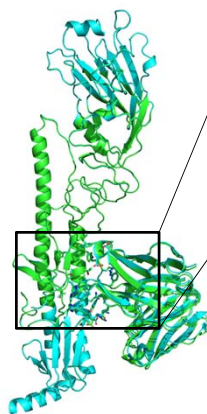
## Plots with sc\_vs\_rmsd.R

```
sc_vs_rmsd.R docking_analysis.csv total_score
sc_vs_rmsd.R docking_analysis.csv dG_separated
sc_vs_rmsd.R docking_analysis.csv dG_separated.dSASAx100
```



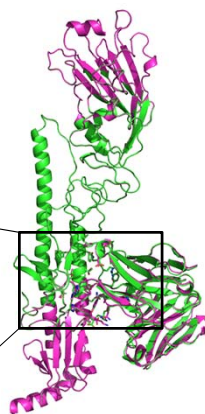
## Docking Results – Overlay with Native

Global Docking

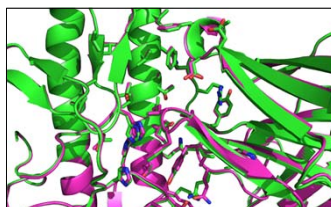
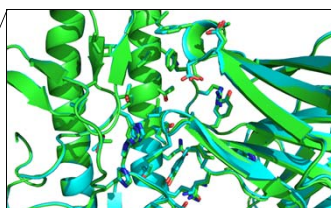


Global Docking  
RMSD: 0.294

Minimization Only



Minimization  
RMSD: 0.257



H5-HA with CR6261 (3gbm)

## Limited 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
- 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.; Karanikolas, 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.; Szyperki, 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.