

# Protein-Protein Docking Predictions with RosettaDock

Jeffrey J. Gray<sup>\*†</sup>, Michael Daily<sup>†</sup>, Lian Guo<sup>\*</sup>, <sup>\*</sup>Chemical & Biomolecular Engineering and <sup>†</sup>Molecular Biophysics, Johns Hopkins University  
David Baker, Stuart Moughon, Chu Wang, Ora Scheuler-Furman, Biochemistry, University of Washington & Howard Hughes Medical Institute



## Abstract

Protein-protein docking algorithms provide a means to elucidate structural details for presently unknown complexes. Here, we present and evaluate a new method to predict protein-protein complexes from the coordinates of the unbound monomer components. The method employs a low-resolution rigid-body Monte Carlo search followed by simultaneous optimization of backbone displacement and side-chain conformations using Monte Carlo minimization. Up to 10<sup>5</sup> independent simulations are carried out, and the resulting "decoys" are ranked using an energy function dominated by van der Waals interactions, an implicit solvation model, and an orientation-dependent hydrogen bonding potential. Top-ranking decoys are clustered to select the final predictions. Small-perturbation studies reveal the formation of binding funnels in 42 of 54 cases using coordinates derived from the bound complexes and in 32 of 54 cases using independently determined coordinates of one or both monomers. Experimental binding affinities correlate with the calculated score function and explain the predictive success or failure of many targets. Global searches using one or both unbound components predict at least 25% of the native residue-residue contacts in 28 of the 32 cases where binding funnels exist. In the Critical Assessment of PRedicted Interactions (CAPRI), the algorithm created several successful predictions, including one of the two best structures of the laminin-nidogen complex (T08) and a successful prediction starting from a homology model (T11). The results suggest that the method may soon be useful for generating models of biologically important complexes from the structures of the isolated components, but they also highlight the challenges that must be met to achieve consistent and accurate prediction of protein-protein interactions. Finally, recent explorations into flexible-backbone docking will be introduced.

## Background and Motivation

**Computational protein-protein docking** is the task of predicting the structure of a biomolecular complex from its unbound monomer components.

**Cellular Function Depends on Protein-Protein Interactions**  
(signaling, regulation, assembly, aggregation enzymes/inhibitors, antibodies/antigens...)

Faulty interactions result in diseases



**Protein docking tests our fundamental knowledge of biomolecular physics**  
(Conformational space and free energy functions)

**Protein docking studies may teach us how to design complex devices capable of assembling themselves from nanoscopic (macromolecular) components**

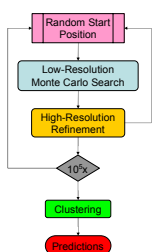


**Computational protein docking could help elucidate biological molecular interactions on a genomic scale**

## Simulation Methods

### Algorithm Features:

- Multi-scale representation:
  - Low-resolution mode enables rapid searching over rigid-body conformation space.
  - High-resolution mode enables accurate decoy discrimination.
- Large number of decoys generated (10<sup>5</sup>) using many compute processors
- Decoy clustering (captures entropic contribution)



### All-Atom Mode: (high-resolution)

- Simultaneous optimization of side-chain conformations and rigid-body displacement
- Side chain rotamer searches



### Monte-Carlo Minimization



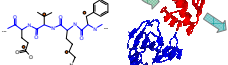
- Scoring function: (in linear combination)

Score	Form / Source	Discriminatory z-value
Repulsive van der Waals	Modified Lennard-Jones 6-12	73.0
Attractive van der Waals	Lennard-Jones 6-12	45.0
Surface area solvation	Surface area (see Tsai 2003)	28.5
Gaussian solvent-exclusion	Lazaridis & Karplus, 1999	27.2
Rotamer probability	Dunbrack & Cohen, 1997	19.6
Hydrogen bonding	Empirical, Kortemme <i>et al.</i> 2003	14.9 & 6.8 (BB/BB)
Residue pair probability	Empirical, Kuhlman & Baker 2000	6.9
Electrostatics	Coulomb model with simple charges	0.4-15.1 (LR rep)

Scoring weights determined using general linear model regressions on small perturbation decoy sets. Different weights are employed for packing, minimization, and discrimination.

### Residue Mode: (low-resolution)

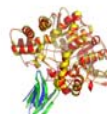
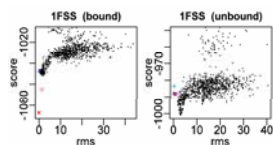
Protein representation: backbone atoms + average centroids



Score	Representation	Physical Force
Contacts	$r_{\text{contact-centroid}} < 6 \text{ \AA}$	Attractive van der Waals
Bumps	$(r - R_b)^2$	Repulsive van der Waals
Residue environment	$-\ln(P_{\text{env}})$	Solvation
Residue pair	$-\ln(P_p)$	Hydrogen bonding electrostatics, solvation
Alignment	-1 for interface residues in Antibody CDR	(bioinformatic)
Constraints	varies	(biochemical)

## Benchmark Results

### Scoring function captures free energy funnels:



A true free energy function would be minimized at the native conformation (rms = 0). We have tuned our scoring function to achieve this for as many targets as possible.

### Selected Global-Search Predictions:

These pictures show both the experimental structures (Red and Blue) and a model (Yellow and Green) selected from ten predictions output by the algorithm.

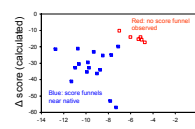


### Benchmark Performance

Number of successful dockings, starting from either bound or unbound protein backbones and searching either near the native structure or globally.

	Bound Perturbations	Unbound Perturbations	Global Searches
Enzyme/Inhibitor	21/22	18/22	17/22
Antibody/Antigen	10/16	9/16	8/16
Other	5/10	5/10	3/10
Difficult	6/6	0/6	0/6
TOTAL	42/54	32/54	28/54

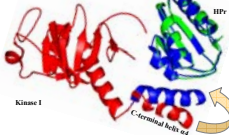
Benchmark set assembled by R. Chen *et al.*, see *Proteins* 2003



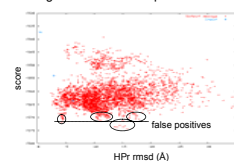
Score change upon binding (using bound backbones) correlates with experimental binding free energy. Low-affinity targets are more difficult to predict.

## Flexible Backbone Explorations

Recently, we have explored techniques to capture backbone conformational change during docking. Here we test on CAPRI T01, HPr + HPr kinase. The N-terminal helix of the kinase swings in upon docking:

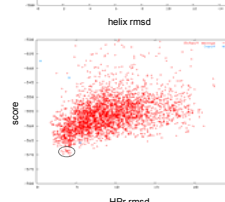
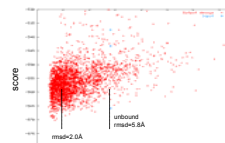
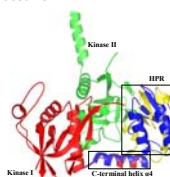


Without backbone flexibility, there is no energy funnel for binding the unbound components:



Torsion angle movement in residues 290-292 would allow the correct conformation to be observed.

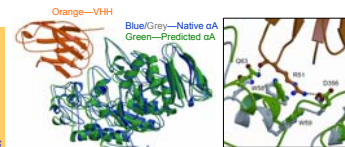
Incorporating torsion angle perturbations and explicit minimizations in the low-resolution search phase, we recover a binding funnel and a structure with a correctly placed helix.



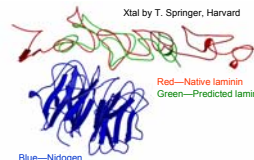
## CAPRI: A Blind Assessment

The **Critical Assessment of PRedicted Interactions** is an international blind prediction challenge (similar to CASP). When a new protein-protein complex has been experimentally solved, theorists across the globe attempt to predict the coordinates before the experimental structures are made public (see Jarin *et al.* *Proteins* 2003). Below are some of our notable predictions.

**CAPRI T06: alpha-Amylase + VHH**  
Model #1:  
• 48/65 contacts  
• 1.33 Å, 3° rotation  
• 1.5 Å rmsd  
  
• Bound VHH coords. were given  
• Captures atomic-scale interactions



Xtal by C. Cambillau, CNRS



**CAPRI T08: Laminin + Nidogen**  
Model #2:  
• 53% contacts  
• 1.5 Å rmsd  
• 0.66 Å interface rmsd  
  
• D800, N802, V804 were constrained to the interface based on experimental knowledge  
• One of the two best models submitted

Prediction using homology model of dockerin Prediction using bound coordinates of dockerin

**CAPRI T11/12: Cohesin + Dockerin**  
Model #6 (T11):  
• 42% contacts  
• 6.1 Å rmsd  
• 1.9 Å interface rmsd  
  
• Dockerin coordinates modeled by homology via the Robetta server  
• RosettaDock produced the best model by correct contacts



Xtal by Romao, Carvalho, Fontes *et al.*, Lisbon

## Summary

- A new docking method has been created which mimics the physical process of protein-protein recognition and binding, including a two-stage search and new energy functions.
- Benchmark study demonstrates efficacy of the energy function and the search strategy.
- Results from blind prediction challenges further demonstrate the ability of the algorithm on a diverse set of binding partners.
- Algorithm is amenable to extensions such as the incorporation of torsion angle moves.

Further Details: Gray *et al.*, *J. Mol. Biol.* 331(1), 281-299 (2003).

## Future Directions

- Benchmark and CAPRI target failures indicate proteins for which either the scoring function is missing important components or the conformational search algorithm is not sampling correct positions or oversampling unphysical positions.
- Backbone motions need to be incorporated more generally, particularly in ways which will not disrupt the bulk structure.
- Techniques to identify flexible and rigid portions of the proteins are needed to keep the conformational search tractable.

### ACKNOWLEDGMENTS:

- NIH K01-HG02316 for support of J.J.G.
- Carol Rohl (UCSC), Brian Kuhlman (UNC), Tanja Kortemme (UCSF) and Kira Misura (UW) contributed components of the algorithm.