WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Incorporating biochemical information and backbone flexibility in RosettaDock for CAPRI rounds 6–12

Sidhartha Chaudhury,[1] Aroop Sircar,[2] Arvind Sivasubramanian,[2] Monica Berrondo,[2] and Jeffrey J. Gray[1,2]*

[1] Program in Molecular and Computational Biophysics, Johns Hopkins University, Baltimore, Maryland 21218

[2] Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland 21218

ABSTRACT

In CAPRI rounds 6–12, RosettaDock successfully predicted 2 of 5 unbound–unbound targets to medium accuracy. Improvement over the previous method was achieved with computational mutagenesis to select decoys that match the energetics of experimentally determined hot spots. In the case of Target 21, Orc1/Sir1, this resulted in a successful docking prediction where RosettaDock alone or with simple site constraints failed. Experimental information also helped limit the interacting region of TolB/Pal, producing a successful prediction of Target 26. In addition, we docked multiple loop conformations for Target 20, and we developed a novel flexible docking algorithm to simultaneously optimize backbone conformation and rigid-body orientation to generate a wide diversity of conformations for Target 24. Continued challenges included docking of homology targets that differ substantially from their template (sequence identity <50%) and accounting for large conformational changes upon binding. Despite a larger number of unbound–unbound and homology model binding targets, Rounds 6–12 reinforced that RosettaDock is a powerful algorithm for predicting bound complex structures, especially when combined with experimental data.

Proteins 2007; 69:793–800.

Key words: *protein–protein docking; RosettaDock; Rosetta; CAPRI; protein structure prediction; protein flexibility; induced fit binding.*

INTRODUCTION

Protein–protein interactions are fundamental to cellular life, and the structure of protein complexes can provide tremendous insight into the mechanisms of protein function. Given the enormous number of undetermined complex structures and the difficulties in structure determination of protein complexes through X-ray crystallography or nuclear magnetic resonance, there is a clear need for accurate computational methods to predict protein complex structures. In such methods, homology models or crystal structures of the unbound protein partners are "docked" with each other to predict the complexed structure. The challenges in docking include both sufficient sampling of putative complexed conformations and discrimination of the correct complexed conformation (near-native).[1] The Critical Assessment of PRotein Interactions[2] (CAPRI) is a blind, community-wide test of protein docking methods. CAPRI targets are newly solved crystal structures that are not released to the public until after the participants have submitted their predictions to ensure that all predictions are blind. We have used RosettaDock, a Monte Carlo-based docking program, in CAPRI since its inception and performed well in previous rounds.[3,4]

CAPRI rounds 6–12 proved significantly more challenging than the previous rounds largely because of an increase in the number of homology targets and unbound–unbound targets, which reflect predictive situations more relevant to the broader biology community. In rounds 1–5,[3,4] RosettaDock primarily relied upon structural information for docking. Protein backbones were held fixed, and rigid-body displacement and side-chain conformations were optimized simultaneously to generate docking predictions.[5] We applied this approach with moderate success in past CAPRI rounds, but as the difficulty of the targets increased, its limitations have become more apparent. In rounds 6–12, we address these limitations with continued development in RosettaDock in two critical areas: the inclusion of experimental biochemical data and the modeling of backbone conformation changes. We outline our development in these two areas and summarize our results for the other CAPRI targets.

The utility of incorporating experimental information, such as mutational analysis, in structure prediction is well-established in CAPRI (e.g., Refs. 6–8).

The authors state no conflict of interest.
Grant sponsor: National Institute of Health (NIH); Grant numbers: R01 GM078221; T32 GM008403.
Sidhartha Chaudhury and Aroop Sircar contributed equally to this work.
*Correspondence to: Jeffrey J. Gray, Department of Chemical and Biomolecular Engineering, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218. E-mail: jgray@jhu.edu
Received 31 May 2007; Revised 26 June 2007; Accepted 28 June 2007
Published online 25 September 2007 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.21731

Table I
Summary of Targets

Target	Complex	Type/SeqID	N_{res}	Best model	F(nat)	L_RMSD (Å)	I_RMSD (Å)	Quality	Notes
20	HemK/eRF1	U-H30	636	6	0.04	38.59	7.15	–	Standard local + multiple loops
21	Orc1/Sir1	U-U	308	3	0.46	7.62	1.92	★★	Standard local + computational mutagenesis
24	Arf1/ARHGAP10	U-H40	272	10	0.15	34.02	11.94	–	Standard local + flexible docking
25	Arf1/ARHGAP10	U-B	288	7	0.00	26.94	10.66	–	Standard local
26	TolB/Pal	U-U	505	6	0.47	2.70	1.24	★★	Standard local + site constraints
27-1	E2-25K/Ubc9	U-U	359	7	0.00	48.55	9.28	–	Standard global + local
27-2	E2-25K/Ubc9	U-U	359	7	0.00	31.85	13.15	–	Standard global + local
28	NEDD4L Dimer	H53-H53	756	2	0.01	65.56	14.53	–	Symmetry-constrained global + local

Type indicates the origin of the starting monomer coordinates: bound (B), unbound (U), or homologous (H) structures. Percent sequence identity is provided for homology targets (SeqID). Best model is the rank of the submitted model of highest quality, detailed in the remaining columns, F(nat) is the fraction of native residue–residue contacts predicted correctly, L_RMSD is the RMSD of the backbone atoms of the smaller protein after optimal superposition of the backbone atoms of the larger protein, I_RMSD is the RMSD of the backbone atoms in the interface after optimal superposition, and quality ratings follow the criteria of Méndez *et al.*[11]

In standard RosettaDock, such information is typically incorporated as site constraints, which specify certain residues which should appear in the interface, and distance constraints, which specify residues that should be within a certain distance of each other. Both of these constraints serve to decrease the docking conformational space and enhance the discrimination of near-native decoys by precluding false positives. We sought to test whether mutational information could yield additional discrimination ability through explicit energetic calculations. We recently used such a computational mutagenesis technique to predict the binding of a therapeutic antibody, using a homology model, to a large flexible receptor antigen.[9] For Target 21, we exploited this technique to take advantage of mutational analysis data available in the literature, ultimately producing a moderately accurate prediction.

As CAPRI continues, the need to accommodate backbone conformational variability in docking is becoming increasingly apparent. This is primarily due to two trends. First, rounds 6–12 included a smaller number of the more rigid enzyme-substrate and antibody–antigen targets and a larger number of targets that show significant conformational changes upon binding such as signaling proteins. Second, homology targets, which often show significant backbone differences from the homology template to the bound structure, comprise an increasing portion of CAPRI targets. The shift in targets is partially due to the greater number from structural genomics initiatives, such as the Structural Genomics Consortium (SGC).[10] These trends are clearly illustrated in Target 24, where there was uncertainty in the structure and position of two large regions in the homology model of one of the partners. We developed a simultaneous conformational sampling and docking algorithm to optimize the conformation of these regions while docking. Although we did not achieve a successful prediction, it served as an interesting case study for modeling large conformational variability in docking.

TARGETS AND PREDICTIONS

In CAPRI rounds 6–12, we used RosettaDock to successfully predict two targets to medium accuracy according to the criteria used by Méndez *et al.*[11] We obtained predictions using global docking, local docking, and docking constraints, when applicable, with standard RosettaDock as described previously.[4,5,12] Docking was complemented with computational mutagenesis in Target 21, and backbone flexibility was modeled in Target 20 and 24. All homology models were generated using the Robetta structure prediction server.[13] Table I shows a summary of the results for all targets. The three metrics used to assess the quality of predictions include root mean square deviation (RMSD) of the smaller partner's (ligand) backbone atoms (N, C_α, C, O) after superposition of the larger partner, RMSD of the interface backbone atoms after optimal superposition, and the fraction of native residue–residue contacts, compared with the crystal structure of the complex.[11] Individual methods and results for each target are detailed below.

Computational mutagenesis

Target 21: Orc1-Sir1

The Bromo Associated Domain (BAH) of Orc1 is sufficient for interaction with Silencer Recognition Defective (SRD) region of Sir1, and biochemical experiments had identified both binding-loss mutations and nonbinding-loss mutations.[14] We locally docked the unbound structures of the isolated BAH and SRD domains (Protein Data Bank[15] codes 1M4Z[16] and 1Z1A[17]) with an additional constraint that penalized the presence of either truncated region in the interface. We then subjected the top-scoring decoys to computational mutagenesis using RosettaInterface[18] and selected those decoys whose

DOI 10.1002/prot

mutagenesis patterns most closely matched experimental results for submission.

In experimental mutagenesis, the effect of a mutation on binding of a protein can be determined qualitatively, as a loss or gain of binding, or quantitatively, as a difference in binding free energy ($\Delta\Delta G$). In computational mutagenesis using RosettaInterface,[19] the $\Delta\Delta G$ of binding is determined for a structural model by calculating the difference in interaction energy before and after a mutation is made to one of the partners:

$$\Delta\Delta G = (\Delta G^*_{AB} - \Delta G^*_A - \Delta G^*_B) - (\Delta G_{AB} - \Delta G_A - \Delta G_B)$$

where ΔG_{AB} is the free energy of the wild type complex, ΔG_A and ΔG_B are the free energies of the solvated wild-type monomers A and B, and the asterisks indicate the corresponding quantities for the mutant species. After optimizing the side-chain conformations using a rotamer library, the free energy is estimated using a calibrated combination of the Rosetta scoring function components dominated by van der Waals, solvation, and hydrogen bond energies.[18] The calculation is used to categorize mutations as binding-loss or nonbinding-loss, depending on whether the $\Delta\Delta G$ estimate is greater or less than 1 kcal/mol, respectively. RosettaInterface will not capture energetics that arise from backbone conformational changes or dynamics.[18]

For each of the top-scoring decoys, we simulated the experimental mutations of Bose *et al.*[14] and selected the decoy whose pattern of binding-loss and nonbinding-loss mutations most closely matched the experiment. Model 3, shown in Figure 1(a), replicated all nonbinding-loss mutations and 3 of 8 binding-loss mutations. This structure was our best prediction and was categorized as medium quality with 7.62 Å ligand RMSD, 1.92 Å interface RMSD, and 46% of the native residue–residue contacts from the crystal structure. The structure ranked 2nd among all submitted predictions by interface RMSD. The importance of the mutagenesis technique is underscored by the fact that Wang *et al.*, using the standard RosettaDock approach without RosettaInterface, were unable to achieve an acceptable prediction.[21]

To validate the accuracy of the computational mutagenesis method in identifying binding-loss mutations from the complex structure, we retrospectively applied this technique to the released crystal structure for this complex (1ZHI[17]). Table II shows the computational mutagenesis results on both our best submitted prediction and the crystal structure [mutation sites shown in Fig. 1(b)]. There is a qualitative match between the mutagenesis results of the best predicted structure and the crystal structure, indicating that decoy discrimination with this method was optimal. For the loss-of-binding mutations in the crystal structure, 6 of 8 showed a positive $\Delta\Delta G$, but only 3 passed the threshold to be classified as a binding-loss mutation. For the nonbinding-loss mutations in the crystal structure, 6 of 8 showed no change in $\Delta\Delta G$, and all 8 were under the threshold for a binding-loss mutation. Although RosettaInterface predicted the correct direction of $\Delta\Delta G$ in 12 of 16 mutations, inaccuracies remain in the calculation of $\Delta\Delta G$ for categorically identifying binding-loss mutations.

Flexible backbone docking

Target 20: HemK-eRF1

Mutational analysis of eRF1 revealed that the 9-residue GGQ loop in eRF1 interacts with HemK, and this loop undergoes a significant conformational change upon binding.[22] Therefore, we modeled this flexibility by generating a diverse set of loop conformations with a loop-building protocol[23] to represent dominant conformations in a pre-existing equilibrium ensemble. We created a homology model of eRF1 using the previously determined crystal structure of eRF2 (1GQE[24]) as a template, which has 40% sequence identity. We then generated a large set of loop conformations in the eRF1 homology model in the absence of HemK, selected the five lowest energy loops from this set, and grafted each of these separately to the eRF1 homology model. We locally docked each eRF1 model independently to the unbound structure of HemK (1NV8[25]). The top-scoring structures from the two largest clusters for each of the five eRF1 models were selected for submission. The released structure of the complex (2B3T[26]) showed a large change in the GGQ loop conformation suggesting that capturing loop movement was essential, but five loop models insufficiently sampled the conformational space, with the closest loop at 5.0 Å RMSD after aligning over residues 226-242 of eRF1 which includes the GGQ loop. Although we did not achieve any acceptable predictions for this target, this method served as a prototype for loop-ensemble docking methods currently being developed in our lab and underscored the need for simultaneous conformational refinement while docking.

Target 24: Arf1-GTP—ARHGAP10

We used the Robetta server to generate five homology models of the Arf1-binding domain of ARHGAP10 (ArfBD) using the template 1BTN,[27] which has 30% sequence identity, to dock with the unbound structure of Arf1-GTP (1O3Y[28]). Visual inspection of the homology models of ArfBD revealed two regions of large variability between the models: a 15-residue loop and a 33-residue C-terminal region. Therefore, we developed and implemented a novel flexible docking method within RosettaDock to account for the uncertainties in the homology model for this target. Our goal was to produce a large diversity of backbone conformations in the context of the protein–protein complex.

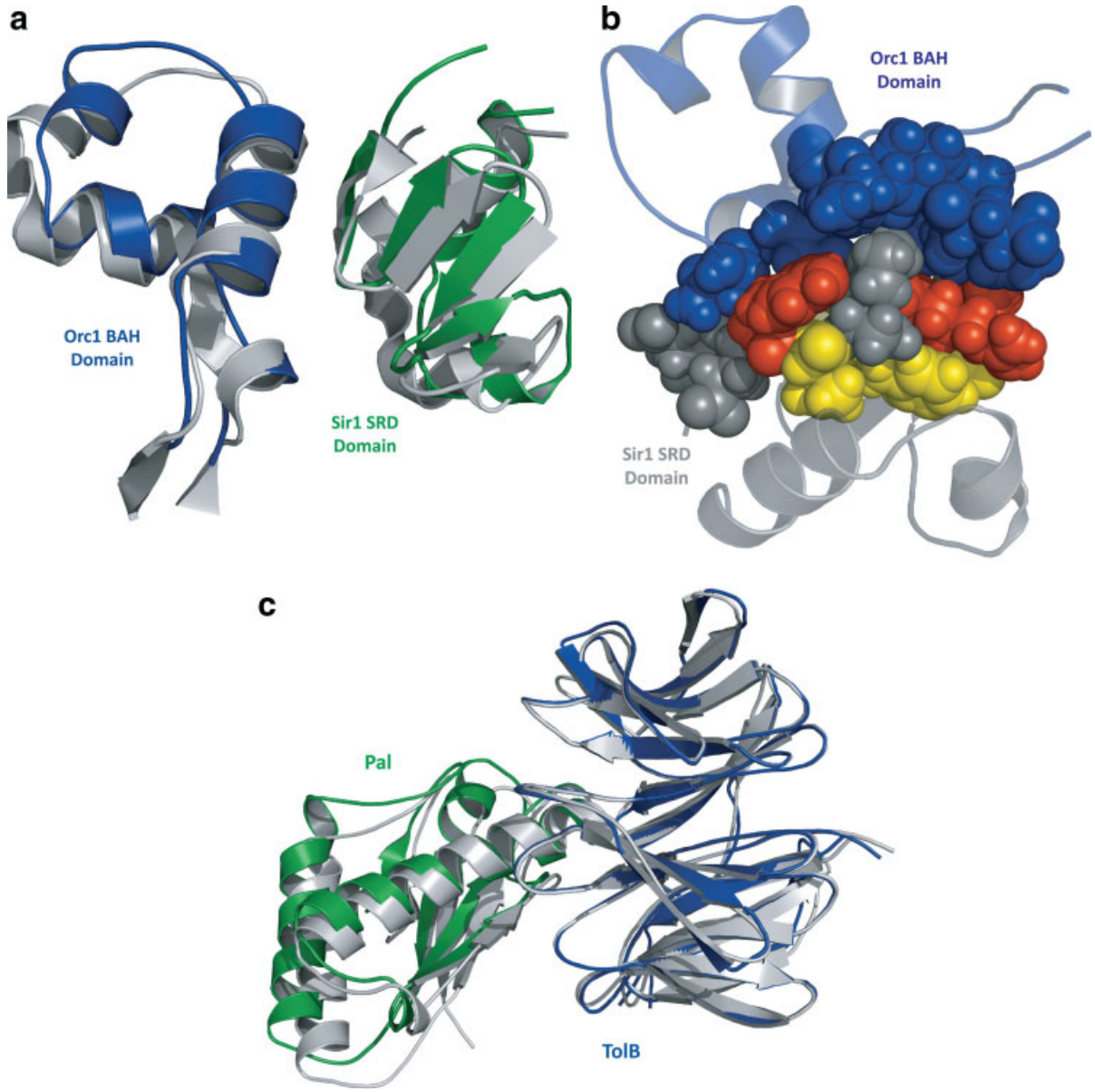

Figure 1

*Successful RosettaDock predictions. (**a**) Medium accuracy prediction (Model 3) of T21 using BAH Domain of Orc1 (blue) and SRD domain of Sir1 (green), superposed on the native structure (gray, 1ZH1[17]) by aligning interface atoms (all atoms within 8 Å of partner). (**b**) Computational mutagenesis results for Model 3 of T21. Blue: Orc1 BAH domain; gray: Sir1 SRD domain. Interface residues shown as spheres with successfully predicted binding-loss mutations in red and binding-loss mutations that were not predicted in yellow. (**c**) Medium accuracy prediction of T26 using Pal (green) and the β-propeller region of TolB (blue) superposed on the native structure (gray, 2HQS[20]) by aligning to TolB in the native structure.*

We added flexible docking to RosettaDock by integrating a loop modeling method within Rosetta[23] to the standard RosettaDock algorithm using a "fold tree" representation of the protein complex[29] which allows for simultaneous treatment of conformational sampling and docking. A fold tree provides a framework for sampling local backbone conformations while maintaining a fixed overall global structure by limiting the propagation of backbone torsion angle perturbations through chain breaks at specified "cut points," which are subsequently repaired. The fold tree consists of "edges," defined as structurally continuous backbone segments, and "jumps" which define the spatial relationship of the edges. Flexible regions of the protein are designated as conformationally "variable regions" and must be adjacent to at least one cut point in the fold tree to allow conformational sampling (Fig. 2, bottom). In the low-resolution stage of RosettaDock, backbone conformations of the variable

 DOI 10.1002/prot

Table II
Computational Prediction of the Change in Binding Free Energy (kcal/mol) for the Best Model and Crystal Structure of Orc1–Sir1 (T21)

Mutation	$\Delta\Delta G_{Model}$	Match	$\Delta\Delta G_{Native}$	Match
Nonbinding-loss mutations				
E487A	0.52	√	0.00	√
E488A	0.00	√	0.28	√
E506A	0.00	√	0.89	√
E507A	0.00	√	0.00	√
K513A	0.00	√	0.00	√
D514A	0.00	√	0.00	√
K522A	0.00	√	0.00	√
D523A	0.00	√	0.00	√
Binding-loss mutations				
Y489S	1.52	√	2.10	√
V490D	0.00	X	0.00	X
R493G	1.04	√	3.41	√
F494S	0.00	X	0.00	X
D503N	0.48	X	0.56	X
A505T	0.62	X	0.97	X
A505S	−0.33	X	0.01	X
A505D	2.80	√	2.34	√

Calculated $\Delta\Delta G$ upon mutation that is greater than 1 kcal/mol in RosettaInterface is indicative of binding loss. The check/crosses indicated qualitative agreement/disagreement between experimental and computational mutagenesis. Model indicates the best prediction, Model 3, and native indicates the crystal structure, 1ZH1.[17]

regions were sampled during docking using a Monte Carlo algorithm that iteratively applied 3-residue fragment-based torsion angle moves[30] and closed the resulting chain breaks using the cyclic coordinate descent (CCD) algorithm.[31] In the high-resolution refinement phase of RosettaDock, the backbone torsion angles of these variable regions were optimized during docking using continuous gradient-based energy minimization following each side-chain repacking step.

A crystal structure of Arf1 bound to a related partner, Sec7,[32] revealed that residues 73-81 in the Switch 2 region of Arf1 were at the interface; mutational analysis in a separate study[33] implicated residues 1052-1096 in the C-terminal region of ArfBD. Therefore, we manually oriented the C-terminal region of ArfBD toward the Switch 2 region of Arf1-GTP, and we locally docked each ArfBD homology model to Arf1 using standard RosettaDock. We then clustered the results and selected the best-scoring structures from the two largest clusters to serve as starting structures for flexible backbone docking. Since the homology models exhibited variation in areas thought to be important for binding, residues 973-988 and 1044-1064 in ArfBD were designated as "variable regions" in the fold tree with Arf1 represented by a single

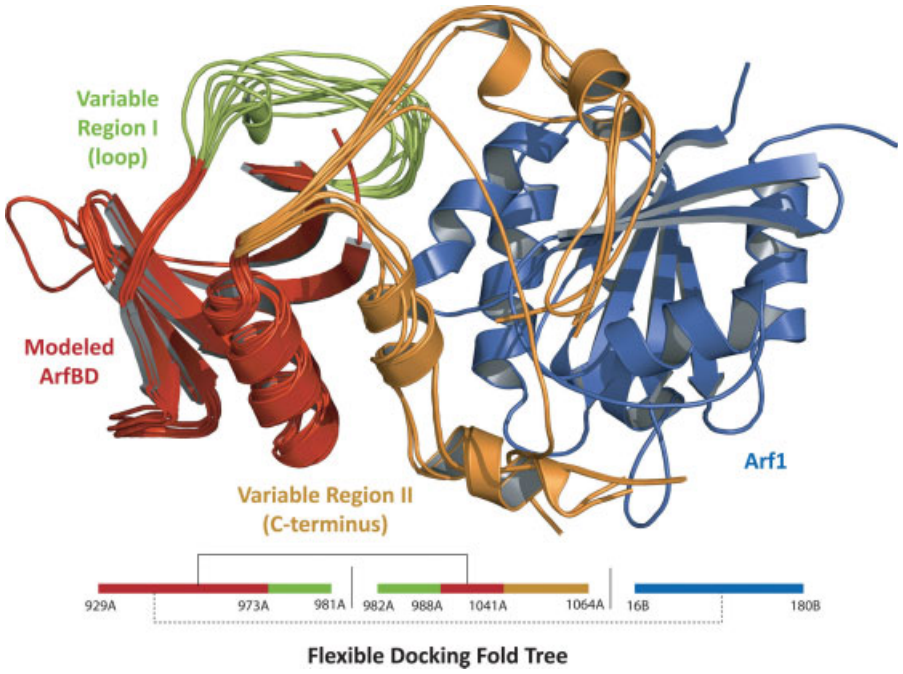

Figure 2
(Top) Conformations of T24 generated by the flexible backbone docking protocol. Red: ArfBD core; blue: Arf1; green: loop region of ARHGAP10 treated as flexible during docking; orange: conformers of the C-terminus of ARHGAP10 built from fragments during docking. (Bottom) Fold tree prescribing the propagation of conformational changes. The colors in the fold tree match their respective regions in the complex structure. The loop jump is shown with a solid line to indicate that the loop stems are held fixed relative to each other. The docking jump is shown as a dotted line to indicate that the two partners are free to move relative to each other. Residue numbering corresponds to the Mus musculus Arf binding-domain of ARHGAP10 (Chain A) and Homo sapiens Arf1 (Chain B).

edge and ArfBD represented by two edges (Fig. 2). For our final predictions, we selected top-scoring structures from clusters with a diverse set of ArfBD C-terminal conformations. We were unable to achieve an acceptable prediction or recover a near-native ArfBD backbone conformation. However, we did succeed in generating a diverse set of conformations of the variable regions from the starting ArfBD model (Fig. 2, top).

Although this method generated complexes with good surface complementarity and hydrogen bonding, visual inspection of the structures revealed that the variable regions of ArfBD tended to unfold and refold around Arf1 leading to unusual extended backbone conformations of ArfBD. Since RosettaDock creates encounter complexes and then optimize their total energies, when given freedom to modify the conformation of large regions of backbone in a complex, it can significantly improve intermolecular energies at the expense of intramolecular energies. Consequently, the free energy of the monomeric state of the putative bound conformation is far worse than the free energy of the unbound conformation which contradicts existing models of protein binding which suggest that the free energies of bound and unbound conformations cannot differ drastically for binding to be kinetically possible.[34] Large induced changes in the overall backbone structure also conflicts with the general observation that proteins retain much of their unbound structure when bound.[35,36] Therefore, a conformational search algorithm designed to model variability of large regions the protein must remain limited to realistic backbone changes from the unbound state by accounting for the unbound energy of the binding proteins, the intramolecular energies and/or compactness.

Standard RosettaDock

Target 26: TolB-Pal

Biochemical experiments suggested that residues 89-130 of Pal interacted with the C-terminal region of TolB, predicted to be a β-propeller.[37] We locally docked the unbound structures of Pal and the C-terminal region of TolB (1CRZ[38] and 1OAP[39]), manually orienting residues 89-130 of Pal toward the β-propeller of TolB, with an additional constraint penalizing interfaces near the truncated N-terminal region of TolB. Model 6, shown in Figure 1(c), was our best prediction with ligand RMSD of 2.70 Å, an interface RMSD of 1.24 Å, and 47% of the native residue–residue contacts from the released crystal structure (2HQS[20]). As a medium-quality prediction, it ranked 3rd among all submitted predictions by ligand RMSD.

Target 25: Arf1-GTP-ARHGAP10

We globally docked the bound structure of ArfBD (2J59[40], chain M) with the unbound structure of Arf1-GTP (1O3Y[28]) and then clustered the results, submitting top-scoring predictions from the largest clusters that agreed with experimental information used for Target 24. Unfortunately, a later-revealed bug in the Rosetta source tree prevented accurate predictions on this target.

Target 27: Ubc9-Hip2

We globally docked the unbound structures of Ubc-9 and Hip-2 (1A3S[41] and 1YLA[42]), clustered the results, and selected the top-scoring structure from the two largest clusters as starting points for local docking and refinement. The largest cluster agreed with a previous study implicating a large positively charged region on Ubc9 as being important in protein interaction.[41]

A substrate-enzyme interaction has been reported where Hip2 is SUMO-lated by Ubc9 on Lys 14 on the N-terminal α-helix.[43] The substrate-interacting residues on Ubc9 included residues 130-136 and 126 along with the catalytic Cys93.[44] Since no cluster in the global docking reflected these findings, we incorporated this biochemical data as site constraints in local docking to produce additional models, one of which had 7% of native contacts.[45]

The results of the two largest clusters from the global run exemplify recent findings that the RosettaDock scoring function is biased toward interfaces of large surface area, even at the cost of low surface complementarity.[46] The local docking failure using site-constraints may also be attributed to the different importance for electrostatics in true enzyme-substrate interactions relative to enzyme-inhibitor or other protein–protein interfaces.

Target 28: HECT-E3 ligase homodimer

We used Robetta to generate five homology models from two different templates (1ND7[47] and 1ZVD[48]) and executed symmetry-constrained global docking[49] using the top-scoring model for each template, generating 10^5 decoys. The 200 highest-scoring structures were clustered for each template and the five largest clusters were used for local, symmetry-constrained docking and refinement. The literature suggested both significant deviation of the HECT-domain orientation from the template structure and substantial conformational flexibility along that region,[47,50] which was later confirmed by the released structure,[51] but we were unable to integrate our flexible docking method from Target 24 with symmetry-constrained docking in time for the predictions.

DISCUSSION

We used RosettaInterface for Target 21 to successfully identify 38% of binding-loss mutations and 100% of nonbinding-loss mutations in a blindly docked model structure. Of the six residues that were shown to lead to binding loss upon mutation, three can be classified as interface residues, that is, residues in which at least one side-chain atom is at least 4 Å from an atom belonging

 DOI 10.1002/prot

to the other partner in the complex. In our computational mutagenesis of the Orc1–Sir1 complex, all three interface binding-loss mutations were identified, while all three noninterface binding-loss mutations were missed [Fig. 1(b)]. This pattern is seen in previous results[18] where RosettaInterface successfully identified 79% of interface binding-loss mutations and only 0.7% of noninterface binding loss mutations for a set of 19 protein complexes of known structure. Rosetta's energy function is short-ranged and only local side-chain motions are captured. Therefore, noninterface mutations, which may alter $\Delta\Delta G$ through protein stability, changes in backbone conformation and packing, or entropy, are beyond the scope of the algorithm.

Since experimental mutational information is often not available for most CAPRI targets, RosettaInterface could be adapted to use evolutionarily conserved surface residues as a proxy for binding-loss mutation residues, or hot spots. Kortemme and Baker demonstrated that computational hot spot residues have the least tolerance for substitution.[18] Since RosettaInterface's overall false positive rate is low (16%), it may be relatively robust to the risk of erroneously identifying false-positive decoys from bioinformatics-identified hot spot residues.[18]

Our approaches to flexible docking for Targets 20 and 24 reflect two opposing views of protein binding: the pre-existing equilibrium hypothesis and the induced-fit hypothesis. In Target 20, our approach was inspired by the pre-existing equilibrium hypothesis of conformational states: that the bound conformation would be a low-energy structure in the ensemble of unbound conformations. We generated a large ensemble of loop conformations in the unbound eRF1 homology model and selected the five lowest energy structures for docking. In hindsight, five structures were probably insufficient to sample the large number of low energy conformations available to that loop, suggesting that larger ensembles may be necessary. In Target 24, we followed an induced-fit approach, that conformational changes occur simultaneously with binding, by first docking ArfBD to Arf1, and then using our flexible docking algorithm to simultaneously optimize the backbone conformation and rigid-body orientation of ArfBD with respect to Arf1. In this case, the results revealed structures where the intermolecular energy was significantly improved at the expense of intramolecular energy, leading to unrealistic bound backbone conformations.

Our two approaches for flexible docking may be combined to complement each other. In such a method, a diverse pregenerated ensemble of structures would be used with smaller-scale backbone refinement during docking. The backbone conformational search is then restricted to small continuous regions of conformational space centered on discrete low-energy structures in the ensemble that are searched stochastically during docking and refinement. Ideally this would reduce both the necessary size of an ensemble to represent conformational space in a purely ensemble-docking approach and the amount of sampling needed to include diverse conformations while docking. This combined approach is representative of the 3-step mechanism for protein binding consisting of diffusion, free conformer selection, and refolding, proposed by Grunberg *et al.*,[34] and we are currently implementing and testing new flexible backbone ensemble docking approaches.

In CAPRI rounds 6–9 we addressed a variety of challenges in computational protein-protein docking. We used biochemical data to quantitatively discriminate near-native decoys based on interface energetics, modeled multiple loop conformations in docking, and developed a novel method for flexible docking. In these and other respects, CAPRI continues to be an excellent testing ground for pushing development of novel methods in RosettaDock.

ACKNOWLEDGMENTS

We thankfully acknowledge the efforts of the CAPRI organizers in finding and evaluating targets and facilitating communication between protein docking groups. We are also grateful to the crystallographers for offering their complexes as CAPRI targets. We also thank Elizabeth Denning for her contributions to Target 20 and Michael Daily for constructive discussions and guidance. The Rosetta protein structure modeling suite is developed collaboratively by several laboratories in the RosettaCommons (www.rosettacommons.org).

REFERENCES

1. Gray JJ. High-resolution protein-protein docking. Curr Opin Struct Biol 2006;16:183–193.
2. Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJ, Vajda S, Vakser I, Wodak SJ. CAPRI: a critical assessment of predicted interactions. Proteins 2003;52:2–9.
3. Gray JJ, Moughon SE, Kortemme T, Schueler-Furman O, Misura KM, Morozov AV, Baker D. Protein–protein docking predictions for the CAPRI experiment. Proteins 2003;52:118–122.
4. Daily MD, Masica D, Sivasubramanian A, Somarouthu S, Gray JJ. CAPRI rounds 3–5 reveal promising successes and future challenges for RosettaDock. Proteins 2005;60:181–186.
5. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D. Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. J Mol Biol 2003;331:281–299.
6. Zhang C, Liu S, Zhou Y. Docking prediction using biological information, ZDOCK sampling technique, and clustering guided by the DFIRE statistical energy function. Proteins 2005;60:314–318.
7. Ma XH, Li CH, Shen LZ, Gong XQ, Chen WZ, Wang CX. Biologically enhanced sampling geometric docking and backbone flexibility treatment with multiconformational superposition. Proteins 2005;60: 319–323.
8. van Dijk AD, de Vries SJ, Dominguez C, Chen H, Zhou HX, Bonvin AM. Data-driven docking: HADDOCK's adventures in CAPRI. Proteins 2005;60:232–238.
9. Sivasubramanian A, Chao G, Pressler HM, Wittrup KD, Gray JJ. Structural model of the mAb 806-EGFR complex using computa-

tional docking followed by computational and experimental mutagenesis. Structure 2006;14:401–414.
10. Structural Genomics Consortium. http://wwwsgc.utoronto.ca/
11. Méndez R, Leplae R, De Maria L, Wodak SJ. Assessment of blind predictions of protein–protein interactions: current status of docking methods. Proteins 2003;52:51–67.
12. Wang C, Schueler-Furman O, Baker D. Improved side-chain modeling for protein–protein docking. Protein Sci 2005;14:1328–1339.
13. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 2004;32:W526–W531.
14. Bose ME, McConnell KH, Gardner-Aukema KA, Muller U, Weinreich M, Keck JL, Fox CA. The origin recognition complex and Sir4 protein recruit Sir1p to yeast silent chromatin through independent interactions requiring a common Sir1p domain. Mol Cell Biol 2004;24:774–786.
15. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
16. Zhang Z, Hayashi MK, Merkel O, Stillman B, Xu RM. Structure and function of the BAH-containing domain of Orc1p in epigenetic silencing. Embo J 2002;21:4600–4611.
17. Hou Z, Bernstein DA, Fox CA, Keck JL. Structural basis of the Sir1-origin recognition complex interaction in transcriptional silencing. Proc Natl Acad Sci USA 2005;102:8489–8494.
18. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein–protein complexes. Proc Natl Acad Sci USA 2002;99:14116–14121.
19. Kortemme T, Kim DE, Baker D. Computational alanine scanning of protein–protein interfaces. Sci STKE 2004;2004:l2.
20. Bonsor DA, Grishkovskaya I, Dodson EJ, Kleanthous C. Molecular mimicry enables competitive recruitment by a natively disordered protein. J Am Chem Soc 2007;129:4800–4807.
21. Wang C, Schueler-Furman O, Andrea I, London N, Fleishman S, Bradley P, Qian B, Baker D. RosettaDock in CAPRI rounds 6–12. Proteins 2007;69:758–763.
22. Heurgue-Hamard V, Champ S, Engstrom A, Ehrenberg M, Buckingham RH. The hemK gene in *Escherichia coli* encodes the N(5)-glutamine methyltransferase that modifies peptide release factors. Embo J 2002;21:769–778.
23. Rohl CA, Strauss CE, Chivian D, Baker D. Modeling structurally variable regions in homologous proteins with rosetta. Proteins 2004;55: 656–677.
24. Vestergaard B, Van LB, Andersen GR, Nyborg J, Buckingham RH, Kjeldgaard M. Bacterial polypeptide release factor RF2 is structurally distinct from eukaryotic eRF1. Mol Cell 2001;8:1375–1382.
25. Schubert HL, Phillips JD, Hill CP. Structures along the catalytic pathway of PrmC/HemK, an N5-glutamine AdoMet-dependent methyltransferase. Biochemistry 2003;42:5592–5599.
26. Graille M, Heurgue-Hamard V, Champ S, Mora L, Scrima N, Ulryck N, van Tilbeurgh H, Buckingham RH. Molecular basis for bacterial class I release factor methylation by PrmC. Mol Cell 2005;20:917–927.
27. Hyvonen M, Macias MJ, Nilges M, Oschkinat H, Saraste M, Wilmanns M. Structure of the binding site for inositol phosphates in a PH domain. Embo J 1995;14:4676–4685.
28. Shiba T, Kawasaki M, Takatsu H, Nogi T, Matsugaki N, Igarashi N, Suzuki M, Kato R, Nakayama K, Wakatsuki S. Molecular mechanism of membrane recruitment of GGA by ARF in lysosomal protein transport. Nat Struct Biol 2003;10:386–393.
29. Bradley P, Baker D. Improved beta-protein structure prediction by multilevel optimization of nonlocal strand pairings and local backbone conformation. Proteins 2006;65:922–929.
30. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol 1997;268:209–225.
31. Canutescu AA, Dunbrack RL Jr. Cyclic coordinate descent: a robotics algorithm for protein loop closure. Protein Sci 2003;12: 963–972.
32. Goldberg J. Structural basis for activation of ARF GTPase: mechanisms of guanine nucleotide exchange and GTP-myristoyl switching. Cell 1998;95:237–248.
33. Dubois T, Paleotti O, Mironov AA, Fraisier V, Stradal TE, De Matteis MA, Franco M, Chavrier P. Golgi-localized GAP for Cdc42 functions downstream of ARF1 to control Arp2/3 complex and F-actin dynamics. Nat Cell Biol 2005;7:353–364.
34. Grunberg R, Leckner J, Nilges M. Complementarity of structure ensembles in protein–protein binding. Structure 2004;12:2125–2136.
35. Betts MJ, Sternberg MJ. An analysis of conformational changes on protein–protein association: implications for predictive docking. Protein Eng 1999;12(4):271–283.
36. Daily MD, Gray JJ. Local motions in a benchmark of allosteric proteins. Proteins 2007;67:385–399.
37. Ray MC, Germon P, Vianney A, Portalier R, Lazzaroni JC. Identification by genetic suppression of *Escherichia coli* TolB residues important for TolB-Pal interaction. J Bacteriol 2000;182:821–824.
38. Abergel C, Bouveret E, Claverie JM, Brown K, Rigal A, Lazdunski C, Benedetti H. Structure of the *Escherichia coli* TolB protein determined by MAD methods at 1.95 A resolution. Structure 1999;7:1291–1300.
39. Abergel C, Walburger A, Bouveret E, Claverie JM. MAD structure of a periplasmic domain in *E. Coli* Pal protein, to be published.
40. Menetrey J, Perderiset M, Cicolari J, Dubois T, El Khatib N, El Khadali F, Franco M, Chavrier P, Houdusse A. Structural basis of Arhgap21-mediated cross-talk between Arf and Rho signalling pathways, to be published.
41. Tong H, Hateboer G, Perrakis A, Bernards R, Sixma TK. Crystal structure of murine/human Ubc9 provides insight into the variability of the ubiquitin-conjugating system. J Biol Chem 1997;272:21381–21387.
42. Choe J, Avvakumov G, Newman E, Mackenzie F, Kozieradzki I, Bochkarev A, Sundstrom M, Arrowsmith C, Edwards A, Dhe-paganon S, (SGC) SGC. Ubiquitin-conjugating enzyme E2–25 kDa (Huntingtin interacting protein 2, to be published.
43. Pichler A, Knipscheer P, Oberhofer E, van Dijk WJ, Korner R, Olsen JV, Jentsch S, Melchior F, Sixma TK. SUMO modification of the ubiquitin-conjugating enzyme E2–25K. Nat Struct Mol Biol 2005;12:264–269.
44. Lin D, Tatham MH, Yu B, Kim S, Hay RT, Chen Y. Identification of a substrate recognition site on Ubc9. J Biol Chem 2002;277:21740–21748.
45. Walker J, Avvakumov G, Xue S, Newman E, Mackenzie F, Weigelt J, Sundstrom M, Arrowsmith C, Edwards A, Bochkarev A, Dhe-paganon S, SGC SGC. A novel and unexpected complex between SUMO-1-conjugationg enzyme ubc9 and the ubiquitin-conjugating enzyme E2–25kDa, to be published.
46. Liang S, Liu S, Zhang C, Zhou Y. A simple reference state makes a significant improvement in near-native selections from structurally refined docking decoys. Proteins 2007;69:244–253.
47. Verdecia MA, Joazeiro CA, Wells NJ, Ferrer JL, Bowman ME, Hunter T, Noel JP. Conformational flexibility underlies ubiquitin ligation mediated by the WWP1 HECT domain E3 ligase. Mol Cell 2003;11:249–259.
48. Ogunjimi AA, Briant DJ, Pece-Barbara N, Le Roy C, Di Guglielmo GM, Kavsak P, Rasmussen RK, Seet BT, Sicheri F, Wrana JL. Regulation of Smurf2 ubiquitin ligase activity by anchoring the E2 to the HECT domain. Mol Cell 2005;19:297–308.
49. Schueler-Furman O, Wang C, Baker D. Progress in protein–protein docking: atomic resolution predictions in the CAPRI experiment using RosettaDock with an improved treatment of side-chain flexibility. Proteins 2005;60:187–194.
50. Huang L, Kinnucan E, Wang G, Beaudenon S, Howley PM, Huibregtse JM, Pavletich NP. Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. Science 1999;286:1321–1326.
51. Walker J, Avvakumov G, Xue S, Butler-Cole C, Weigelt J, Sundstrom M, Arrowsmith C, Edwards A, Bochkarev A, Dhe-paganon S. NEDD4-like E3 Ubiquitin-protein Ligase, to be published.

 DOI 10.1002/prot