PROTEINS: Structure, Function, and Bioinformatics 60:187–194 (2005)

Progress in Protein–Protein Docking: Atomic Resolution Predictions in the CAPRI Experiment using RosettaDock With an Improved Treatment of Side-Chain Flexibility

Ora Schueler-Furman,[1†] Chu Wang,[1†] and David Baker[1,2*]
[1]*Department of Biochemistry, University of Washington, Seattle, Washington*
[2]*Howard Hughes Medical Institute, University of Washington, Seattle, Washington*

ABSTRACT **RosettaDock uses real-space Monte Carlo minimization (MCM) on both rigid-body and side-chain degrees of freedom to identify the lowest free energy docked arrangement of 2 protein structures. An improved version of the method that uses gradient-based minimization for off-rotamer side-chain optimization and includes information from unbound structures was used to create predictions for Rounds 4 and 5 of CAPRI. First, large numbers of independent MCM trajectories were carried out and the lowest free energy docked configurations identified. Second, new trajectories were started from these lowest energy structures to thoroughly sample the surrounding conformation space, and the lowest energy configurations were submitted as predictions. For all cases in which there were no significant backbone conformational changes, a small number of very low-energy configurations were identified in the first, global search and subsequently found to be close to the center of the basin of attraction in the free energy landscape in the second, local search. Following the release of the experimental coordinates, it was found that the centers of these free energy minima were remarkably close to the native structures in not only the rigid-body orientation but also the detailed conformations of the side-chains. Out of 8 targets, the lowest energy models had interface root-mean-square deviations (RMSDs) less than 1.1 Å from the correct structures for 6 targets, and interface RMSDs less than 0.4 Å for 3 targets. The predictions were top submissions to CAPRI for Targets 11, 12, 14, 15, and 19. The close correspondence of the lowest free energy structures found in our searches to the experimental structures suggests that our free energy function is a reasonable representation of the physical chemistry, and that the real space search with full side-chain flexibility to some extent solves the protein–protein docking problem in the absence of significant backbone conformational changes. On the other hand, the approach fails when there are significant backbone conformational changes as the steric complementarity of the 2 proteins cannot be modeled without incorporating backbone flexibility, and this is the major goal of our current work. Proteins 2005;60:187–194.**

Key words: rotamer minimization; high-resolution docking; Monte Carlo minimization; free energy landscape; energy function

INTRODUCTION

Protein–protein interactions play an essential role in cellular processes. A protein complex structure allows detailed analysis of an interaction at the atomic level. However, most of the solved structures of proteins are monomers, with only a small fraction (~10%) of multimers[1] that is expected to decrease even more in the era of structural genomics. It is therefore important to develop reliable docking methods that can predict the structure of a complex starting from the free monomers. Various docking approaches have been developed over the years (see Halperin et al.[2] and Vajda and Camacho[3] for reviews). CAPRI[4,5] is a community-wide, blind experiment aimed at an objective evaluation of the performance of different methods in creating models for protein–protein complexes.

RosettaDock uses real-space Monte Carlo minimization (MCM) on both rigid-body and side-chain degrees of freedom to identify the lowest free energy docked arrangement of 2 protein structures.[6,7] The search is based on a free energy function that is dominated by a Lennard–Jones potential, an orientation-dependent hydrogen bond potential,[8] and an implicit solvation model.[9] Originally, side-chain flexibility was based on rotamers from a backbone-dependent library.[10] Recent modifications of the protocol have greatly improved side-chain modeling by enhancing side-chain conformational sampling through gradient-based off-rotamer optimization, and also by including information from the unbound structures.[11] We present here the results of predictions for CAPRI Rounds 4 and 5 that were created with the improved protocol. For most of the targets, we were able to create high-resolution models with interface–residue backbone root-mean-square devia-

The Supplementary Materials referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat.

†These authors contributed equally to this work.

Grant sponsor: National Institutes of Health. Grant sponsor: Damon Runyon Cancer Research Foundation; Grant number: Fellowship DRG-1704-02 (to O. Schueler-Furman).

*Correspondence to: David Baker, Department of Biochemistry, Box 357350, University of Washington, Seattle, WA 98195. E-mail: dabaker@u.washington.edu

Received 10 February 2005; Accepted 16 February 2005

DOI: 10.1002/prot.20556

TABLE I. Summary of Performance in Capri Rounds 4 and 5 (Targets Grouped According to the Number of Backbone Clashes)

Target	Name	Target class[a]	Backbone clashes[b]	Interface backbone–atom RMSD (I_rmsd)[c] (Å)	Interface full-atom RMSD (I_all_rmsd)[c,d] (Å)	Ligand RMSD[c] (Å)	Fraction of native contacts (Fnat)[c]	No. of h/m/a/i[c,e]
Group1								
T15*	immD–ColD	B–B[f]	0	**0.23** (1.06)	1.13 (1.95)	0.63 (2.15)	**0.89** (0.86)	3/3/3/1
T12	Cohesin–Dockerin	U–B	1	**0.27 (0.28)**	1.08 (1.14)	**0.46** (0.49)	0.87 (0.86)	7/0/0/3[g]
T14	MYPT1–PP1	B–H	2	**0.38** (0.50)	1.01 (1.03)	0.93 (0.93)	**0.61 (0.61)**	10/0/0/0
Group2								
T11	Cohesin–Dockerin	U–*H*	5	**1.09 (1.13)**	3.05 (3.22)	**5.81** (6.24)	**0.42** (0.35)	0/2/1/7
T13	Ab–SAG1	B–U	12	0.65 (17.0)	0.94 (20.0)	**2.44** (72.4)	0.47 (0.00)	0/1/0/9
T19	Ab–Prion	B–*H*	16	**0.98 (0.98)**	1.53 (1.53)	**2.53 (2.53)**	**0.74** (0.69)	1/6/0/3
Group3								
T18	Xylanase–TAX1	U–U	35	14.3 (14.3)	15.4 (15.5)	46.7	0.00	0/0/0/10
T10	TBEV envelope protein	Trimer	51	17.4 (23.9)	18.2 (24.3)	33.3 (40.1)	0.01	0/0/0/10

[a]Starting structure: B, bound; U, unbound; H, homolog unbound. NMR-solution structures are indicated in italics.
[b]Backbone clashes across the interface (pairs of atoms closer than a minimal distance; see Methods section and Table II in Supplementary Material) of starting structures superimposed onto bound complex structure.
[c]Measures according to CAPRI assessors[12,13] (http://capri.ebi.ac.uk/). The best value among all submitted models is indicated. (Overall best predictions among all submissions are in bold.) The values for the first submission are indicated in parentheses.
[d]RMSD of all interface residue heavy atoms (Méndez, personal communication).
[e]Accuracy of models[12]: h, high; m, medium; a, acceptable; i, incorrect.
[f]Side-chains were stripped from starting models.
[g]The 3 incorrect models are the symmetric conformation (see text).
*Target 15 was cancelled after we had submitted our predictions.

tion (RMSD) (I_rmsd) values within 1 Å and interface–residue all-atom RMSD (I_all_rmsd) values within 1.6 Å with accurately modeled side-chain conformations (see Table I and Fig. 1).

MATERIALS AND METHODS

Prediction protocol

The basic prediction protocol of RosettaDock has been described previously.[6,7] Briefly, it includes the following steps: (1) prepacking of the partners to remove clashes in the free monomers; (2) global search of rigid-body orientations; and (3) clustering of the low-energy models and selection of the largest cluster as prediction. The global search step starts from a large number of random initial orientations and (1) brings the partners into glancing contact and removes clashes (slide-into-contact step); (2) optimizes their rigid-body orientation at a low-resolution level [where each amino acid side-chain is represented by a centroid pseudoatom that is positioned according to an average position determined from a set of known structures from the Protein Data Bank (PDB)[19]; see Table III in the Supplementary Material]; and (3) builds initial side-chain conformations using a Monte Carlo search through a backbone-dependent rotamer library, and then uses ~50 cycles of MCM to optimize the side-chain and rigid-body degrees of freedom using a free energy function dominated by short-range Lennard–Jones and hydrogen-bonding interactions, and an implicit solvation model. Each MCM step consists of (1) a small, random, rigid-body perturbation; (2) updating the side-chain conformations by either (a) cycling through the rotamers at positions in which the energy increases, or (b) by full combinatorial repacking of rotamers for all interface side-chains; and (3) gradient-based minimization of the rigid-body orientation.

We have modified this protocol to include an additional local refinement of models created by the global run (500–1000 small perturbations) to explore the local energy landscape. In case of an energy funnel, the lowest energy conformation is selected as a model for submission. Otherwise, low-energy outliers are subjected to further refinement. Variations of the protocol that skip the slide-into-contact or the low-resolution optimization steps are also evaluated at this stage.

Significant shortening of the global search

The improved side-chain modeling, as well as the additional local refinement, allowed us to drastically reduce the number of random orientations to be evaluated in the initial global run (10^4 random orientations were sufficient to detect global minimum energy conformations, a reduction of 1–2 orders of magnitude compared to the original protocol). Another contribution to reducing computer time comes from modification of the energy filters[6] that are now target-specific at each step, resulting in maximal enrichment of low-energy models in the global run.

Side-chain modeling

Improved side-chain modeling in RosettaDock is described in detail in Wang et al.[11] Briefly, increased sampling of side-chain conformations is achieved through an additional step that consists of cycling through rotamers including off-rotamer, gradient-based minimization

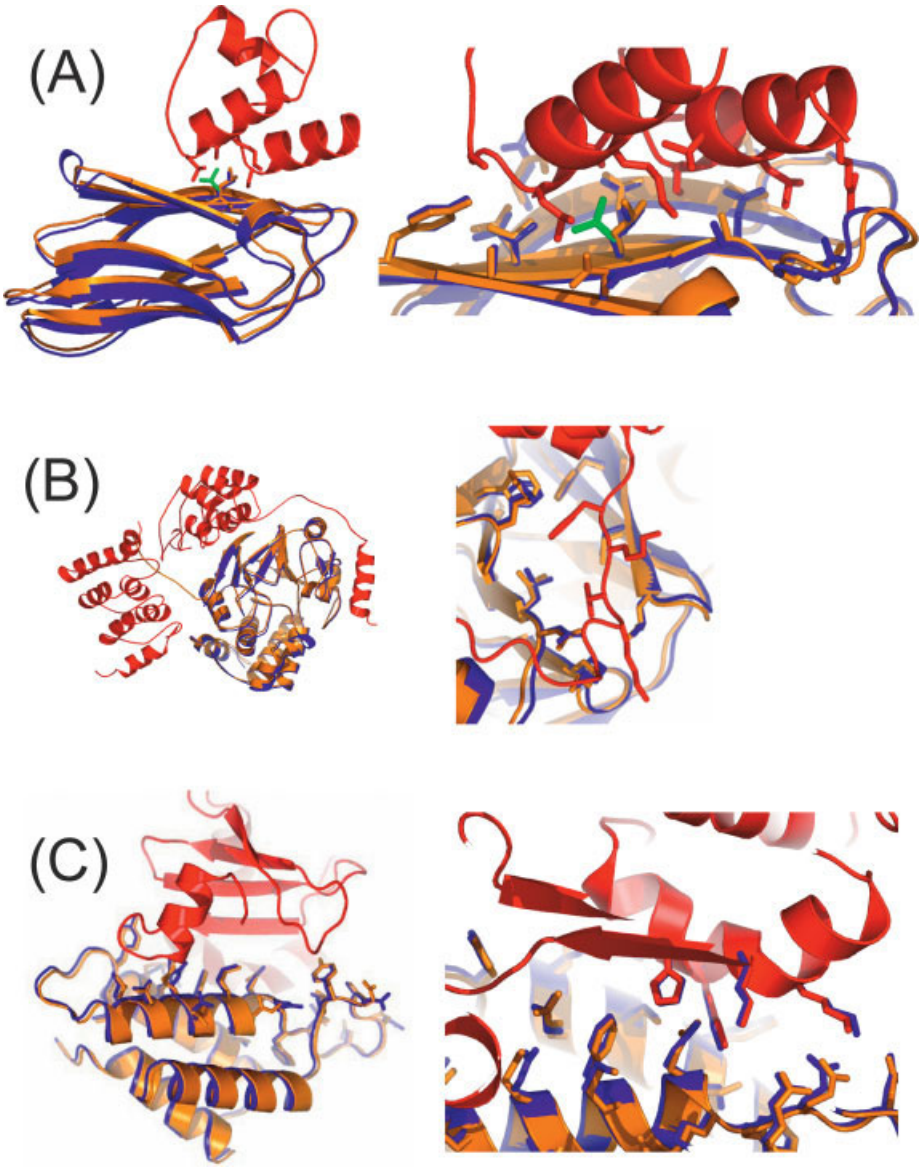

Fig. 1. High-resolution CAPRI predictions of both rigid-body orientation and side-chain conformation. The crystal structure is shown in red and orange, and the orientation of the model is shown in blue. For each target, the whole protein complex is shown in the left panel, while the right panel shows details of the interface. (**A**) Target 12: Dockerin–Cohesin.[14] The side-chain conformation of Leu78 of cohesin is shown in its unbound conformation (green), its predicted conformation in the model (blue), and its experimentally determined conformation (orange). The detailed interface shows in addition the side-chain conformations of residues that are important for the interaction (e.g., positions Asn37, Asp39, Tyr74, and Glu86 for the cohesin domain,[15] and positions Ser45, Thr46, and Arg53 for the dockerin domain[16]). (**B**) Target 14: Myosin phosphatase targeting subunit (MYPT1)– protein S/T phosphatase 1.[17] The detailed interface shows the recognition peptide of MYPT1 (residues 35–38, KVKF) in the binding groove of the phosphatase (residues Leu243, Phe257, and Phe293). (**C**) Target 15: Colicin D–Immunity protein D.[18] The detailed interface shows the side-chains of the catalytic residue His611 and additional positively charged residues that are thought to bind to the RNA, as well as their matching negatively charged residues in the immunity protein.

(RTMIN: Rotamer Trial with MINimization in torsion space[11]). RTMIN is added to the original protocol after each full combinatorial repacking of interface side-chains in the MCM (see above). Also, side-chain conformations of the free monomers are added to the rotamers from the backbone dependent library (and are assigned minimal internal energy).

Selection of decoys

Most prediction runs were performed without including any a priori biological information, and the energy of a model was the primary criterion for the selection of the submissions. In some cases, biological information constraints, including FAB filters were used as described

before[6,7] (see description of individual targets below). The interface size and compactness of the models was monitored by measuring the number of contacts (atom–atom contacts across the interface were counted based on a 5 Å cutoff between 2 atoms from different partners).

Symmetric docking

For the prediction of the structure of homomultimers, we implemented a new protocol that performs the search for the optimal conformation within the space of symmetric conformations. The general docking protocol keeps the first partner fixed and optimizes the relative orientation of the second partner. In contrast, the symmetric docking protocol rotates the first partner, thereby changing the orientation of the ordinates that define the coordinate system relative to the partner. The homomultimer is created from symmetry operations based on the ordinates, ensuring full sampling of possible symmetric conformations.

Target classification

As our method does not incorporate backbone flexibility, it is expected to fail when the unbound structures differ significantly enough from the bound structures that they no longer exhibit steric complementarity at the backbone level. To assess the amount of backbone changes upon binding, the structures provided were superimposed onto the final complex structure, and the backbone clashes between partners were counted. (i.e., atom pairs across the interface that were nearer than a minimal approach distance derived from a representative data set of protein structures.[20] Backbone atoms, as well as centroid pseudoatoms, are included. See Tables II and III in the Supplementary Material). As expected, success and failure of the RosettaDock method correlated closely with the number of clashes: High-resolution and medium-resolution predictions were created for all targets with less than 5 and 20 clashes, respectively.

RESULTS AND DISCUSSION

The predictions are discussed for each target in detail, and summarized in Table I.

Target 10: Tick-Borne Encephalitis Virus (TBEV) Fusion Protein Homotrimer

Target 10 involved the prediction of the structure of the homotrimer (PDB code: 1urz[21]), starting from the monomer conformation observed in the dimer (PDB code: 1svb[22]). It motivated us to develop a symmetric docking protocol (see Methods section). Our best submitted model is 17 Å interface RMSD away from the correct solution (Table I). The prediction failed because of conformational changes that occur in the monomers upon the creation of the homotrimer[21] (51 backbone clashes between the starting monomer conformations when positioned as in the complex; Table I). In addition, sampling only symmetric conformations throughout the search for the minimum energy conformation might be too restrictive.

Targets 11 and 12: Cohesin–Dockerin Complex of the Cellulosome

Models of the homolog dockerin structure of Target 11 were created with the Robetta server.[23] Since both the template and the homolog dockerin are known to bind cohesin (cohesin-2 domain of *Clostridium thermocellum*), we first docked the template structure (1daq[24]) to cohesin (1anu[25]). The resulting model of the complex served as a scaffold, upon which the homology model of dockerin was superimposed and then locally minimized to reduce clashes and optimize the rigid-body orientation. Our first submitted model is of medium quality [1.1 Å interface RMSD; 42% native contacts; see Table I and Fig. 2(A)]. The interface is much larger than the actual interface in the experimental structure; hence, over 50% of the contacts in the model are wrong. Interface residues of the dockerin and cohesin that are important for the interaction (e.g., Ser45 and Arg53[16]) are correctly positioned at the interface; however, the side-chain packing is less accurate (3.2 Å interface full-atom RMSD), not surprisingly, considering the difference in the relative orientation of the 2 helices in dockerin that is not modeled by our current protocol. This result is only achieved if the homology modeling is performed in the last step (i.e., after optimization of the rigid-body orientation, as predictions of the rigid-body orientation starting with the homology model of dockerin failed). No constraints were used for this run.

When starting from the bound dockerin structure in Target 12, we unambiguously identified a conformation with significantly lower energy than the background distribution (Figure 3, left top panel). Our first submission for this target was the lowest energy structure from a refinement run of this orientation [Fig. 1(A)]: a high-quality prediction with 0.28 Å interface RMSD, 87% native contacts, and merely 6% wrongly predicted contacts (see Table I). No biological constraints were used for this run. Interestingly however, runs that did include constraints, located in addition to the minimum energy conformation selected above another conformation that turned out to be a symmetric solution, and positioned Ser11 and Thr12 at the interface instead of the Ser45–Thr46 pair. Indeed, the dockerin sequence and structure contains a repeat, and it is known that only mutation of both Ser–Thr pairs abolishes specificity[16] (suggesting that if only 1 site is eliminated, the second site could still bind in the very much same orientation[14]). All our models to this target consisted of variations of these 2 orientations (7 in the orientation observed in the crystal structure, and 3 in the alternative orientation), as they were clearly lower in energy than all other sampled configurations.

A detailed view of the interface shows that side-chain conformations are modeled very accurately [see Fig. 1(A)]. In addition, the importance of side-chain flexibility is demonstrated in the example of the cohesin side-chain Leu78, which has to be moved away from its unbound conformation in order to prevent a clash with Leu22 in the dockerin partner.

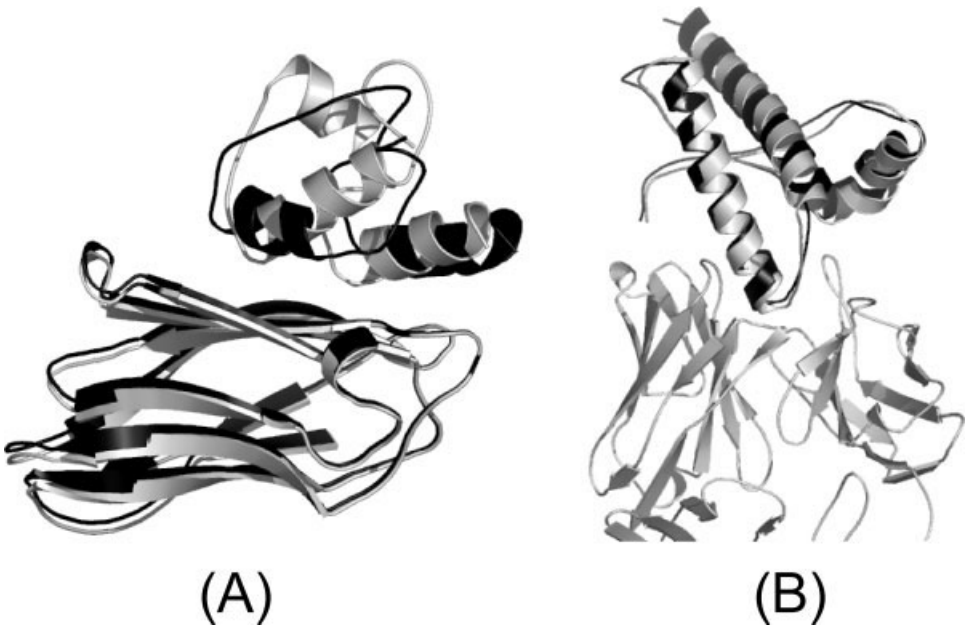

Fig. 2. Accurate CAPRI predictions for homology models based on NMR templates. The crystal structure is shown in gray, and the orientation of the model is shown in black. (**A**) Target 11: Dockerin–Cohesin. (**B**) Target 19: Ovine prion–FAB complex.[23]

Target 13: Antibody–SAG1

The global docking search was performed using a filter for interactions with the CDR loops in the antibody.[7] Based on the fact that the tip of the C-terminal domain of SAG1 is anchored into the parasite membrane, cluster 1, whose interface directly involves this region, was excluded from our consideration. A further local perturbation search did not exhibit strong energy funnels for any clusters. Our model 5, selected from low-energy conformations obtained by locally refining the center of the second largest cluster, correctly predicts 47% of the native interface contacts and the ligand superimposes onto the experimental structure[27] with a backbone RMSD of 2.44 Å, which ranks overall the best among the predicted models by this measure (see Table I). However, the ranking of this model in our final submission was falsely biased by the biochemical data suggesting that the peptide from residue 138 to 154 constitutes a potential epitope on SAG1,[28] which turns out to be irrelevant to this specific antibody.

Target 14: pp1δ–MYPT1

The bound conformation of MYPT[17] (the starting structure) contains an exposed but defined N-terminal region (residues 1–39), and suggested that the complex would be intertwined. Initial global runs produced only models with small interfaces, probably due to our slide-into-contact step (see Gray et al.[6] and Methods section), which is not good at creating initial conformations for intertwined proteins. Since the N-terminal peptide (1–39) binds to the phosphatase,[29] we decided to dock only this part of MYPT. The starting rigid-body orientation was created by approximately orienting the peptide onto the binding groove as described in Egloff et al.,[30] so that peptide residues 35–38 were located in proximity to residues Leu243, Phe257, and Phe293. Those residues were also used as constraints in the runs that followed. pp1δ was modeled based on 2 different template structures of pp1α (PDB codes: 1fjm[31] and 1it6[32]). Initial models of the complex were created based on the template 1fjm. These models served as starting points for local refinement of models based on the template 1it6.

A global search that optimized the rigid-body orientation of the N-terminal peptide to the phosphatase detected a deep energy minimum conformation (Fig. 3). We added the rest of MYPT back, and further refined the complete complex by local refinement searches with and without the slide-into-contact step. The resulting minimum energy conformation based on 1fjm was submitted as first model, which contains the absolute highest fraction of native contacts among all submissions (61%), together with a low faction of non-native contacts (9%) (see Table I). Model 8 is the minimum energy conformation based on 1it6, which resulted in the absolute lowest value of interface RMSD [0.38 Å; Fig. 1(B)].

Target 15: ImmD–ColD

For this target we were given the bound conformations of the partners,[18] without the side-chains. The best prediction (model 3) is of very high accuracy: The interface RMSD is 0.23 Å, and 89% correct contacts were identified [see Table I and Fig. 1(C)]. The catalytic residue of the tRNAse colicin, His611, forms a hydrogen bond with Glu56 of the immunity protein. The energy is similar, and only slightly higher than the energy of the first submitted model (interface RMSD 1.1 Å and 86% of native contacts identified). Interestingly however, the latter contains sig-

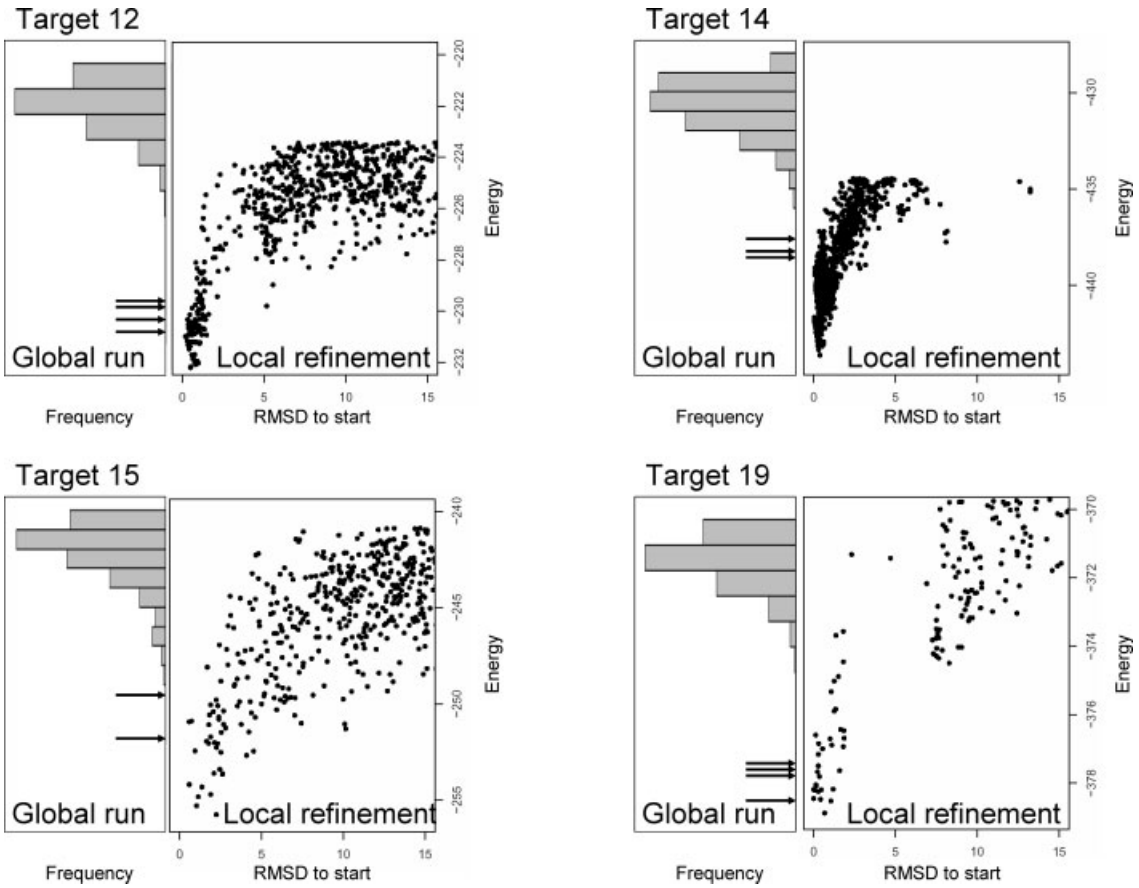

Fig. 3. Selection of models by energy. Examples are shown for high-resolution predictions (Targets 12, 14, 15 and 19). **Left panel:** Energy histogram of models created by initial global search. Arrows indicate low-energy, near native models (within 2 Å RMSD of final submission). A large energy gap between the global minimum energy conformation and the background indicates that a correct solution has been found. **Right panel:** Subsequent local energy refinement defines funnels in the free energy landscape around the free energy minimum, as shown by energy versus RMSD plots.

nificantly more atom–atom contacts at the interface. This is an example of an interface that contains a significant amount of water molecules, and therefore a relatively small number of direct atom–atom contacts. Thus, the contact number is not always a good selection criterion, especially not for polar interfaces. This target was canceled after we had submitted our models.

Target 18: TAXI–inhibitor

Global docking searches were performed, and low-energy cluster centers were subjected to further local perturbation refinements. Unfortunately, none of our submitted models captured the experimentally determined binding mode. From our postassessment of the predictions, we observed that, although it was claimed to be provided in the complexed conformation, the starting TAXI structure showed different backbone conformations from that in the released experimental complex structure,[33] especially in the region of Asn290–Gly294, which plays a critical role in stabilizing the native xylanase-TAXI interface. Indeed, 35 backbone clashes were counted when the starting structures were superimposed onto the final complex structure (see Table I), resulting in inevitable atomic overlaps in any docking model with a native-like binding interface.

Target 19: Prion–Antibody

In this target, the starting structure of the prion was an NMR-solution structure.[34] The global optimization run was performed in 2 different ways: (1) 500 models were generated starting from each of the 20 NMR models in the PDB entry (1dwz) and combined; (2) the minimized, averaged NMR model (1dwy) was used as the starting structure of the prion protein. Both approaches unambiguously located a conformation with significantly lower energy than the rest of the population. The run performed with the averaged minimized structure was more efficient and detected the global minimum more often (Fig. 3, Target 19, left panel). This is, however, not expected to be generally the case, as the averaging is likely to introduce inaccuracies, but as we completely repack the side-chains, some of these inaccuracies may be alleviated. The final model lies within 1 Å interface RMSD of the crystal structure and

predicts 74% of the contacts correctly [Table I and Fig. 2(B)].

SUMMARY

A summary of our predictions is given in Table I, which classifies the targets into groups based on the backbone conformational changes (as measured by the number of backbone clashes when the starting structures are superimposed onto the complex). Group1 includes targets with practically no clashes, indicating that changes in the backbone conformations are not necessary to achieve a good prediction (Targets 12, 14, and 15). Indeed, those targets were predicted with high accuracy. Group2 includes targets with intermediate backbone clashes, and indeed those targets are predicted at medium resolution (Targets 11, 13, and 19). Finally, a significant number of clashes in Group3 (Targets 10 and 18) made it impossible to access near-native conformations with our docking protocol that is based on a clash-sensitive energy function. The significant accuracy of the models in Groups1 and 2 is highlighted by the fact that for 5 of the targets, the all-atom interface RMSD is within 1.6 Å.

The lowest energy model (our first submission) is in general of comparable quality to the overall best submission (except for Target 13). In most of the successful predictions, the prediction is separated from the background by a significant energy gap, and located within the center the free energy basin (Fig. 3). These characteristics indicate successful predictions, and not surprisingly, were not observed in Group3 Targets 10 and 18. In addition to the energy function, biological information can in some cases contribute to successful predictions. For example, the prediction of Target 14 was simplified by focussing on the N-terminal MYPT peptide. Also the prediction of Target 15 could have benefited from considering the environment of the catalytic residue His611, which would have selected the correct model as first submission among 2 models with very similar energy values (see above).

CONCLUSIONS

Successful structure prediction is dependent both on the accuracy of the free energy function and the thoroughness of sampling of conformational space. The two are not independent: a critical feature of our approach in contrast to the traditional grid-based methods is that since the side-chains are completely flexible, no softening of the interatomic interactions is necessary. The energy landscapes in Figure 3 suggest that there is not a "scoring problem" in protein–protein docking provided that (1) the interaction potential is not softened and (2) structures sufficiently close to the native structure are sampled (in CAPRI, this was evident in our many correct predictions for multiple targets). Indeed, it is evident from these landscapes that the free energy funnels are fairly narrow, which is not surprising given the importance of close complementary packing and hydrogen bonding, both very short-range interactions, to the free energy of protein–protein interactions. Given the narrow apertures of the free energy funnels, it seems likely that no "scoring function" can reliably and consistently recognize as close to correct structures greater than 5 Å from the native structure—instead it is necessary to sample close enough to the native structure for the close complementary packing of the side-chains to be realized.

The Rosetta MCM and side-chain packing algorithms, and free energy function were first developed in the context of *ab initio* structure prediction, but despite progress, predictions of the quality of those shown in Figure 1 have not been achievable. The nice feature of protein–protein docking is that the conformational space is very much smaller, even with full side-chain flexibility, and with currently available computing power, the sampling method and free energy function appear to some extent to solve the fixed backbone docking problem. This validation of the free energy function and the sampling methodology is encouraging as we attempt to tackle the much more challenging flexible backbone docking and ab initio structure prediction problems.

ACKNOWLEDGMENTS

Our thanks to the many scientists who have participated in the development of the suite of computational tools used in the Baker laboratory for computations on the structure of proteins. In particular, Jeffrey G. Gray laid the groundwork for RosettaDock. John Karanicolas provided useful scripts, and Dylan Chivian helped with the creation of homolog models. Keith Laidig of Formix, Inc. built and maintained reliable, state-of-the-art computing resources.

REFERENCES

1. 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.
2. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 2002;47:409–443.
3. Vajda S, Camacho CJ. Protein–protein docking: is the glass half-full or half-empty? Trends Biotechnol 2004;22:110–116.
4. Janin J. Assessing predictions of protein–protein interaction: the CAPRI experiment. Protein Sci 2005;14:278–283.
5. 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.
6. 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.
7. 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.
8. Kortemme T, Morozov AV, Baker D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein–protein complexes. J Mol Biol 2003;326:1239–1259.
9. Lazaridis T, Karplus M. Effective energy function for proteins in solution. Proteins 1999;35:133–152.
10. Dunbrack RL Jr, Karplus M. Backbone-dependent rotamer library for proteins: application to side-chain prediction. J Mol Biol 1993;230:543–574.
11. Wang C, Schueler-Furman O, Baker D. Improved side-chain modeling for protein–protein docking. Protein Sci 2005;14:1328–1339.
12. Mendez 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.

13. Wodak SJ, Mendez R. Prediction of protein–protein interactions: the CAPRI experiment, its evaluation and implications. Curr Opin Struct Biol 2004;14:242–249.
14. Carvalho AL, Dias FM, Prates JA, Nagy T, Gilbert HJ, Davies GJ, Ferreira LM, Romao MJ, Fontes CM. Cellulosome assembly revealed by the crystal structure of the cohesin–dockerin complex. Proc Natl Acad Sci USA 2003;100:13809–13814.
15. Miras I, Schaeffer F, Beguin P, Alzari PM. Mapping by site-directed mutagenesis of the region responsible for cohesin–dockerin interaction on the surface of the seventh cohesin domain of *Clostridium thermocellum* CipA. Biochemistry 2002;41:2115–2119.
16. Mechaly A, Yaron S, Lamed R, Fierobe HP, Belaich A, Belaich JP, Shoham Y, Bayer EA. Cohesin–dockerin recognition in cellulosome assembly: experiment versus hypothesis. Proteins 2000;39:170–177.
17. Terrak M, Kerff F, Langsetmo K, Tao T, Dominguez R. Structural basis of protein phosphatase 1 regulation. Nature 2004;429:780–784.
18. Graille M, Mora L, Buckingham RH, Van Tilbeurgh H, De Zamaroczy M. Structural inhibition of the colicin D tRNase by the tRNA-mimicking immunity protein. EMBO J 2004;23:1474–1482.
19. 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.
20. Simons KT, Ruczinski I, Kooperberg C, Fox BA, Bystroff C, Baker D. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. Proteins 1999;34:82–95.
21. Bressanelli S, Stiasny K, Allison SL, Stura EA, Duquerroy S, Lescar J, Heinz FX, Rey FA. Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. EMBO J 2004;23:728–738.
22. Rey FA, Heinz FX, Mandl C, Kunz C, Harrison SC. The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. Nature 1995;375:291–298.
23. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 2004;32:W526–W531.
24. Lytle BL, Volkman BF, Westler WM, Heckman MP, Wu JH. Solution structure of a type I dockerin domain, a novel prokaryotic, extracellular calcium-binding domain. J Mol Biol 2001;307:745–753.
25. Shimon LJ, Bayer EA, Morag E, Lamed R, Yaron S, Shoham Y, Frolow F. A cohesin domain from *Clostridium thermocellum*: the crystal structure provides new insights into cellulosome assembly. Structure 1997;5:381–390.
26. Eghiaian F, Grosclaude J, Lesceu S, Debey P, Doublet B, Treguer E, Rezaei H, Knossow M. Insight into the PrPC→PrPSc conversion from the structures of antibody-bound ovine prion scrapie-susceptibility variants. Proc Natl Acad Sci USA 2004;101:10254–10259.
27. Graille M, Stura E, Bossus M, Muller BH, Letourneur O, Battail-Poirot N, Sibai G, Gauthier M, Rolland D, M.H. LD, Ducancel F. Structure of the immunodominant epitope displayed by the surface antigen 1 (SAG1) of Toxoplasma gondii complexed to a monoclonal antibody. 2005. Submitted for publication.
28. Velge-Roussel F, Dimier-Poisson I, Buzoni-Gatel D, Bout D. Anti-SAG1 peptide antibodies inhibit the penetration of *Toxoplasma gondii* tachyzoites into enterocyte cell lines. Parasitology 2001;123:225–233.
29. Toth A, Kiss E, Herberg FW, Gergely P, Hartshorne DJ, Erdodi F. Study of the subunit interactions in myosin phosphatase by surface plasmon resonance. Eur J Biochem 2000;267:1687–1697.
30. Egloff MP, Johnson DF, Moorhead G, Cohen PT, Cohen P, Barford D. Structural basis for the recognition of regulatory subunits by the catalytic subunit of protein phosphatase 1. EMBO J 1997;16:1876–1887.
31. Goldberg J, Huang HB, Kwon YG, Greengard P, Nairn AC, Kuriyan J. Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1. Nature 1995;376:745–753.
32. Kita A, Matsunaga S, Takai A, Kataiwa H, Wakimoto T, Fusetani N, Isobe M, Miki K. Crystal structure of the complex between calyculin A and the catalytic subunit of protein phosphatase 1. Structure (Camb) 2002;10:715–724.
33. Sansen S, De Ranter CJ, Gebruers K, Brijs K, Courtin CM, Delcour JA, Rabijns A. Structural basis for inhibition of *Aspergillus niger* xylanase by triticum aestivum xylanase inhibitor-I. J Biol Chem 2004;279:36022–36028.
34. Zahn R, Liu A, Luhrs T, Riek R, von Schroetter C, Lopez Garcia F, Billeter M, Calzolai L, Wider G, Wuthrich K. NMR solution structure of the human prion protein. Proc Natl Acad Sci USA 2000;97:145–150.