PROTEINS: Structure, Function, and Bioinformatics 60:207–213 (2005)

ZDOCK and RDOCK Performance in CAPRI Rounds 3, 4, and 5

Kevin Wiehe,[1] Brian Pierce,[1] Julian Mintseris,[1] Wei Wei Tong,[2] Robert Anderson,[1] Rong Chen,[1] and Zhiping Weng[1,2]*

[1]*Bioinformatics Program, Boston University, Boston, Massachusetts*
[2]*Department of Biomedical Engineering, Boston University, Boston, Massachusetts*

ABSTRACT **We present an evaluation of the results of our ZDOCK and RDOCK algorithms in Rounds 3, 4, and 5 of the protein docking challenge CAPRI. ZDOCK is a Fast Fourier Transform (FFT)-based, initial-stage rigid-body docking algorithm, and RDOCK is an energy minimization algorithm for refining and reranking ZDOCK results. Of the 9 targets for which we submitted predictions, we attained at least acceptable accuracy for 7, at least medium accuracy for 6, and high accuracy for 3. These results are evidence that ZDOCK in combination with RDOCK is capable of making accurate predictions on a diverse set of protein complexes. Proteins 2005;60:207–213.** © 2005 Wiley-Liss, Inc.

Key words: ZDOCK; RDOCK; CAPRI; protein docking; scoring function; blind test

INTRODUCTION

CAPRI is an experiment for members of the protein docking community to test their algorithms for accurately predicting the 3-dimensional (3D) structures of protein complexes given only the independently solved structures of their constituents.[1] We participated in the first 2 rounds of CAPRI with the ZDOCK algorithm and achieved some successful predictions.[2,3] ZDOCK is an initial-stage docking algorithm that utilizes a Fast Fourier Transform (FFT) search algorithm and a 3-term energy function consisting of shape complementarity, electrostatics, and desolvation energy.[4] The shape complementarity term is based on pairwise atom distance calculations and was shown to perform better than more commonly used grid-based shape complementarity terms.[5] In addition, we have developed a benchmark of protein complexes in order to extensively test our docking algorithms that is freely available for the protein docking community.[6] The benchmark has recently been updated and now contains over 80 protein complex test cases.[7]

Before Round 3 of CAPRI, we developed the RDOCK algorithm to refine and rerank the top 2000 predictions of ZDOCK with an energy minimization protocol utilizing CHARMM.[8] We demonstrated that RDOCK could substantially improve docking results when used with 3 versions of ZDOCK that employ various combinations of the 3 energy terms in their respective scoring functions.[9] Round 3 of CAPRI marked the first time RDOCK was implemented into our docking approach in a blind test setting.

Through the results from the latest 3 rounds of CAPRI, we have demonstrated that our approach of using ZDOCK initial-stage docking combined with RDOCK refining and reranking is highly successful for making accurate protein docking predictions.

MATERIALS AND METHODS

Our approach to Rounds 3, 4, and 5 of CAPRI involved refining and reranking ZDOCK's top 2000 initial predictions with RDOCK, followed by clustering these results and then visually examining the lowest energy members of each cluster. Biological information from the literature was incorporated in our approach by either preventing undesirable contacts within ZDOCK or selecting for known contacts by using distance filters on predictions after RDOCK. For the last 2 rounds of CAPRI, we also applied a slightly different approach to clustering than in the earlier rounds. Because some test cases in the past were more accurately predicted with different versions of ZDOCK, we sought a method to combine several independent runs of ZDOCK, each with 2000 predictions. Specifically, we ran 2 versions of ZDOCK, one with just the pairwise shape complementarity term and the other with all 3 energy terms in its scoring function. We then refined and reranked each with RDOCK. Finally we clustered these predictions along with predictions from the ZDOCK run with all 3 energy terms. The entire docking approach applied to the latest rounds of CAPRI is outlined in detail as follows.

Blocking

In most targets submitted, biological information was used on some level to discourage contacts between certain residues in the ZDOCK predictions. Blocking is accomplished by assigning all atoms of the residues to be prohibited from the putative binding site a special atom type associated with zero desolvation energy before ZDOCK is executed. In comparison to a contact filter on ZDOCK output, blocking is advantageous because predictions with undesirable contacts are unlikely to appear in the top 2000

Grant sponsor: National Science Foundation; Grant numbers: DBI-0078194, DBI-0133834, and DBI-0116574.

*Correspondence to: Zhiping Weng, Bioinformatics Program, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, MA 02215. E-mail: zhiping@bu.edu

Received 16 January 2005; Accepted 31 January 2005

DOI: 10.1002/prot.20559

predictions. Thus, more acceptable structures are available to be evaluated with RDOCK.

ZDOCK

ZDOCK is a rigid-body protein-docking algorithm that explicitly searches rotational space and uses an FFT algorithm to significantly speed up searching in translational space. For CAPRI we used 2 different versions of ZDOCK that differ only by the terms in their scoring function. We used both ZDOCK2.1,[10] which includes only the pairwise shape complementarity term, and ZDOCK2.3,[11] which includes pairwise shape complementarity, electrostatics, and desolvation energy. The rotational sampling interval was set at 6°, and all default parameters were used.

M-ZDOCK

A prototype of the M-ZDOCK algorithm was used in docking Target 10, the tick-borne encephalitis virus (TBEV) homotrimer. M-ZDOCK is a rigid-body docking program designed to predict the structure of C_n multimers based on the structure of a subunit.[12] Its scoring function is the same as for ZDOCK 2.3: the sum of shape complementarity, desolvation, and electrostatics. It performs a search exclusively in the 4D space of C_n multimers to improve speed and accuracy, using 2D Euler angle sets that evenly sample the symmetric space. As with ZDOCK, an FFT is used to speed up the translational portion of the search.

RDOCK

RDOCK is a 3-stage energy minimization algorithm designed as a refinement and reranking tool for ZDOCK's top predictions. RDOCK is implemented as a protocol in CHARMM involving the following 3 steps: (1) Remove clashes that occur from the soft-shape complementarity parameter in ZDOCK that allows for small conformational change; (2) optimize polar interactions; and (3) optimize charge interactions. For steps 1 and 2, the charges are turned off for ionic residues. This treatment of charges is particularly successful in antibody–antigen docking cases. After refinement, the predictions are rescored using both electrostatics and desolvation terms. The new scores are used to rerank the top ZDOCK predictions. For CAPRI we refined and reranked the top 2000 ZDOCK predictions for each run of ZDOCK.

All of our software, including ZDOCK, M-ZDOCK, and RDOCK, are freely available to academic users at http://zdock.bu.edu/software.php.

Clustering

In all 3 of the latest rounds of CAPRI, we clustered the top predictions after RDOCK to reduce structural redundancy within the 2000 structures of each prediction set. Clustering was accomplished using all-atom root-mean-square deviation (RMSD) between any 2 ligand orientations. We applied a greedy clustering algorithm to cover all orientations with a minimal number of clusters. A *cluster* was defined as a group of structures for which RMSDs from the center structure are smaller than a specified distance cutoff.

In Round 3, we applied a simple approach, clustering the top 2000 predictions after RDOCK by ligand RMSD using a 6-Å cutoff and manually inspecting each cluster that included at least 1 prediction ranked higher than 200. For Rounds 4 and 5, clustering was done independently for the following 3 prediction sets: the top 2000 predictions from ZDOCK2.1 combined with RDOCK, the top 2000 predictions from ZDOCK2.3 combined with RDOCK, and the top 2000 predictions from ZDOCK2.3 alone. Clustering was done at an 8-Å cutoff. The 20 lowest energy representatives from each cluster were kept. These structures were then combined with the top 20 predictions for each of the 3 runs before clustering. The combined set of 120 structures was then manually inspected.

Contact Filtering

For some of the CAPRI targets there was sufficient biological information in the literature to select for residues known to be in the binding site. In these cases, we applied various filters to our prediction sets to minimize the distance between putative interface residues and the surface of the partner protein. This information aided us significantly in the manual inspection portion of our general approach.

Manual Inspection

In order to further cull the prediction set that remained from the previous steps down to the 10 structures per target that would make up our final submissions, manual inspection was applied. Manual inspection consisted of individually displaying the remaining predictions and investigating each for several criteria, including favorable contact filtering scores, charge complementarity, buried hydrophobic residues, and overall agreement with the biological data from the literature. We then manually ranked our final 10 predictions based on the original RDOCK and ZDOCK ranks, as well as the criteria mentioned.

RESULTS

Overall Performance

Our overall performance in the latest 3 rounds of CAPRI is summarized in Table I. We completed 9 targets for submission, but unfortunately we did not submit Target 15 before it was removed from the challenge. Of the 9 targets we submitted, we succeeded in achieving at least 1 prediction of acceptable accuracy for 7, at least 1 prediction of medium accuracy for 6, and at least 1 prediction of high accuracy for 3. We made very accurate predictions in targets without large conformational changes. Our success in the new category of homology docking was dependent on the accuracy of our homology models. Our best prediction of the 3 rounds based on a comparison with the field was Target 13. With no additional biological information other than the complementarity determining regions (CDRs) of the antibody, we were able to successfully predict a structure with a high accuracy rating and attained the

TABLE I. Overall Performance of ZDOCK and RDOCK

Target	Protein complexes	Accuracy[a]	I_RMS (Å)[b]	L_RMS (Å)[c]	Contact %[d]	Ranked[e]
8	Nidogen–laminin	Medium	1.09	10.38	47	1
9	LicT homodimer	Incorrect	9.77	14.68	8	2
10	TBEV trimer	Incorrect	5.05	11.98	11	3
11	Cohesin–dockerin	Acceptable	2.17	9.15	13	1
12	Cohesin–dockerin	High	0.49	1.02	84	1[f]
13	SAG1–FAB	High	0.64	2.57	87	1
14	MYPT1–PP1	High	0.95	3.83	53	8
18	GH11 Xylanase–TAXI	Medium	1.86	5.54	91	1
19	Ovine prion–FAB	Medium	1.26	4.67	57	8

[a]Accuracy is determined by the CAPRI evaluation team based on interface RMSD, ligand RMSD, and percentage of correct contacts predicted.[2]
[b]Interface RMSD.
[c]Ligand RMSD.
[d]Percentage of correct interface residue contact pairs predicted.
[e]Rank of best prediction out of the 10 submissions for that target unless noted otherwise.
[f]For Target 12 our best prediction was ranked 9th, but our first ranked prediction also achieved high accuracy.

highest contact percentage of any group for this target. The evaluation of each target, along with target-specific methods, including the details of the application of biological data, follows.

Target 8 (Nidogen G3–Laminin)

Target 8 was an unbound–bound docking case, where the constituents were comprised of a bound nidogen and an unbound laminin. We blocked the N-terminus on nidogen G3 because it is connected to the G2 domain, effectively prohibiting that part of the nidogen G3 from docking to laminin in vivo. A mutagenesis study of laminin showed that nidogen bound to the middle domain (domain IV) of laminin.[13] Consequently, we cut 10 residues from the N-terminus in domain III of laminin, as well as blocked the last 44 residues of the C-terminus in domain V to moderately constrain our predictions to the middle domain of laminin. It was also known from the mutagenesis study that residues Asp73, Asn75, and Val77 of laminin were crucial for binding; therefore, a contact filter was run after clustering to eliminate predictions that had large distances between these 3 residues and the nidogen surface.

Our best prediction [Fig. 1(C)] was ranked first among our 10 submissions and achieved medium accuracy on the CAPRI evaluation team's scale of incorrect, acceptable-, medium-, and high-accuracy ratings. The ratings of all targets were obtained from the CAPRI website (http://capri.ebi.ac.uk/capri_ranking_predictions.txt). There was a large conformational change[15] of over 20 Å RMSD in the C-terminal domain (domain V) of laminin for which our predictions did not account as is evident by the 10.38 Å ligand RMSD for our best prediction. Nonetheless, we correctly predicted 47% of the interface contacts, and our top prediction had a very low interface RMSD of 1.09 Å.

Target 9 (LicT Homodimer)

Target 9 was an unbound–unbound case, where participants were asked to dock the 2 monomeric subunits of a mutant LicT homodimer by first mutating 2 Asp residues back to wild-type His residues, then docking the subunits. We used the SCWRL 2.0 program[16] to replace the sidechains of the 2 mutant residues. Because the homodimers are principally symmetrical, we also ran a filter for symmetry before running RDOCK. In order to filter for symmetry, we created an atomic contact vector[17] (ACV) for both the receptor and the ligand. An ACV represents the interfacial contacts uniquely in vector form by their physicochemical attributes and allows for a simple way to compare 2 interfaces. If the difference between the receptor ACV and ligand ACV is small, the interface contacts are similar and the homodimer interface is most likely symmetrical.

The wild-type LicT homodimer consists of 2 PRD domains. Significant conformational change occurs in the hinge region between the 2 PRD domains resulting in a large domain shift.[17a] Because of this shift, we failed to make an accurate prediction in our 10 submissions because of our approach's reliance on a mostly rigid-body model for docking.

Target 10 (TBEV Trimer)

For Target 10, we were given 1 unbound monomer of the TBEV envelope protein to assemble into its low-pH-induced trimeric form. Because ZDOCK was not initially developed to dock multimeric complexes, we made a modification to ZDOCK to account for the symmetry constraint. The details of this modified algorithm, M-ZDOCK, are outlined in the Methods section. The version of M-ZDOCK we used for Round 4 of CAPRI was a prototype, and it has been further developed and published.[12] We were told at the beginning of Round 4 by the CAPRI evaluation team that this target had a large conformational change in its C-terminal domain, and we blocked this entire domain accordingly. The first 2 domains do not move significantly and were the only portion of the structure evaluated; however, the native orientation of these 2 domains results in a severe steric clash in the third domain if it is kept in its unbound conformation. Had we removed the third domain completely, this clash would not have occurred. Therefore, blocking the third domain

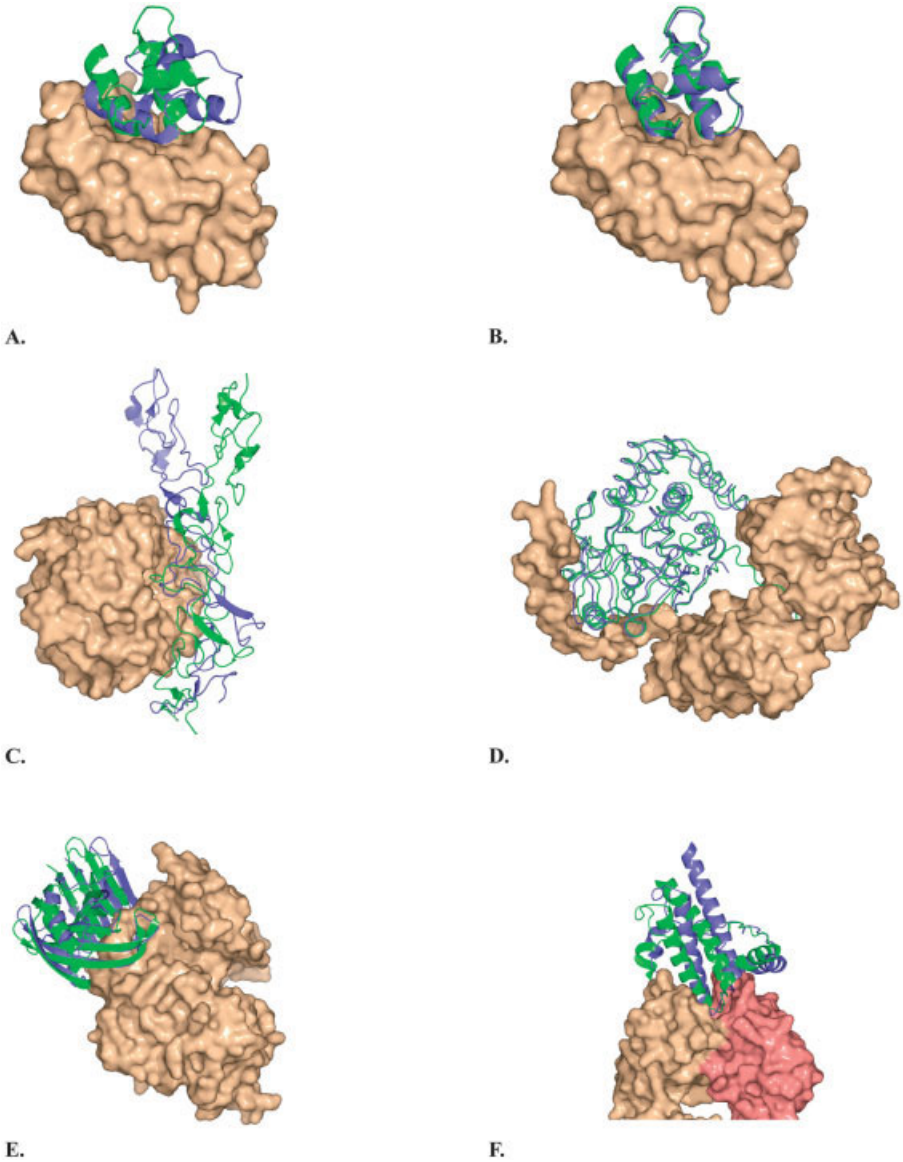

Fig. 1. Comparison of 6 target predictions to their respective crystal structures where the receptors (beige, surface representation) have been superposed and the resulting transformation matrix is applied to the ligands (crystal structure in lime, our prediction in blue, both in cartoon representations). (**A**) Target 11, predicted and native complex of cohesin (receptor) and homology-modeled dockerin (ligand). (**B**) Target 12, cohesin (receptor) and bound dockerin (ligand). (**C**) Target 8, nidogen (receptor) and laminin (ligand). (**D**) Target 14, MYPT1 (receptor) and PP1 (ligand). (**E**) Target 18, TAXI (receptor) and *A. niger* xylanase (ligand). (**F**) Target 19, FAB (receptor, L chain in pink and H chain in beige) and ovine prion protein fragment (ligand). All figures were created using PyMOL.[14]

instead of removing it proved to be a costly error in our docking strategy.

Unfortunately, because of the conformational change[18] in the third domain, we did not predict any accurate structures in our 10 submissions. Although we did not make any successful predictions, this target inspired us to investigate developing a novel algorithm for multimer docking resulting in a new program explicitly for symmetrical multimer assembly, M-ZDOCK. After further development following CAPRI, M-ZDOCK was able to identify a near-native structure ranked in the top 100 predictions for the TBEV trimer case.

Targets 11 and 12 (Cohesin–Dockerin)

Targets 11 and 12 involved predicting the complex of a cohesin-2 domain and a type-2 dockerin. Dockerin and cohesin are both part of a larger multiprotein complex called the cellusome. Cohesin-2 is one of 9 domains that make up the structural protein known as scaffoldin, and dockerin is attached to a catalytic domain.[19]

For Target 11, we were given the structure of an unbound cohesin-2 domain, the structure of a type-7 dockerin, and a sequence alignment of the type-2 and type-7 dockerins with 50% identity. We were instructed to

homology model the type-2 dockerin and then dock the resulting structure with the cohesin-2 domain. Because our group lacks expertise in homology modeling, we chose to use the Swiss Model Homology Server[20] to predict the structure of the type-2 dockerin. The resulting model had a backbone RMSD of 4.58 Å. Target 12 was released after submission of Target 11 and included the bound structure of the type-2 dockerin from the complex.

We incorporated 3 different blocking strategies into separate independent runs of ZDOCK and clustered the 3 prediction sets together. For the first scheme, we blocked both terminal residues of cohesin-2 domain because it is linked at both termini to other scaffoldin domains; thus, these portions of the cohesin-2 domain structure would not be accessible to dockerin in vivo. Also in the first scheme, we blocked the N-terminal residue of the dockerin because it is attached to a large catalytic domain at this end; thus, the region is not accessible to bind cohesin. Two Ser–Thr doublets in dockerin, Ser11–Thr12 and Ser45–Thr46 were implicated in binding in the literature.[21] Additionally, another study noted that Lys18 was important in the cohesin–dockerin interaction.[22] For the second blocking scheme, we used this biological information to construct a putative binding site on dockerin, and we blocked residues 30–37 on the opposite face of the protein in order to discourage ZDOCK predictions from involving these residues in the binding. To further restrict predictions involving the termini, we blocked 8 additional residues from both termini of cohesin and the N-terminus of dockerin in this scheme. The third blocking scheme is inclusive of the second, however; we also blocked residues 20–27 and 91–95 on cohesin, additionally restricting the face with both termini from our prediction results, as well as increasing the area blocked on dockerin to residues 24–37. All 3 blocking schemes followed by clustering were used in predictions of both Targets 11 and 12.

There was biological evidence from a mutagenesis study on the homologous cohesin 7 domain that suggested Asp39, Tyr74, and Glu86 jointly effected cohesin–dockerin binding affinity.[23] We used a contact filter to find predictions with short distances between the corresponding 3 residues in the cohesin-2 domain and the 5 biologically important residues of dockerin. We applied this approach to both Targets 11 and 12.

Our best predictions for Targets 11 and 12 [Fig. 1(A and B)] achieved acceptable and high accuracy, respectively. An important observation from the crystal structure of this complex[24] is that only the Ser–Thr doublet at positions 45–46 participates in the interface. It has now been suggested that dockerin may bind in 2 modes due to its internal symmetry, with either of the Ser–Thr doublets at the binding site. For Target 11, we were hampered by the low quality of the dockerin homology model. Most likely a consequence of the inaccurate positioning of the Ser–Thr doublets in the modeled structure was that many of our models satisfied the false notion of locating both doublets at the binding site. In Target 12, there were very few high-scoring clusters with structures that satisfied both criteria, and much better accuracy was achieved. In fact, the clustering of Target 12 ligand orientations was so tight around the correct binding site on cohesin that we included many similar predictions from it in our submission. The strategy worked, as we had 5 out of 10 predictions achieve high accuracy. Our best prediction was ranked 9th and had a 0.49-Å interface RMSD. We correctly predicted 87% of the interface contacts in this submission. Our first ranked prediction for Target 12 also attained a high accuracy rating.

Target 13 (SAG1–FAB)

In Target 13, we were asked to predict the complex of SAG1, a surface protein antigen of the toxoplasmosis parasite, and an antibody in an unbound–bound case in which the antibody was from the complex and the SAG1 was unbound. We assumed canonical antibody recognition and blocked all non-CDR regions of the antibody. We did not find any additional biological information significant to the binding of this complex in the literature.

Despite not having any biological data on this target, our first ranked prediction achieved a high accuracy rating (structure not shown; Graille and Ducancel, unpublished data). The interface RMSD for that submission was 0.64 Å and we correctly predicted 87% of the interface contacts, which was the highest contact percentage attained for all groups on this target. The difficulty of this target due to the lack of biological data emphasizes our good performance.

Target 14 (MYPT1–PP1)

Target 14 consisted of docking unbound protein serine–threonine phosphatase-1 (PP1) with the myosin phosphatase targeting subunit 1 (MYPT1) taken from the complex structure. We blocked the PP1 N-terminus and C-terminus because the structure file was missing residues at both termini.[25] With the same motivation, we also blocked the C-terminus of the MYPT1, because it was missing a 200-residue stretch of ankyrin repeats at this end. There were extensive biological data available for this interaction in the literature,[26–28] with the most relevant being a surface plasmon resonance study (SPR) between fragments of MYPT1 and PP1.[29] This study showed that residues 1–22, 35–38, and the ankryin repeats (39–263) on MYPT-1 were crucial to the interaction. We utilized a contact filter to identify predictions with short distances between these crucial residues and the surface of the PP1. However, we did not find any predictions with PP1 contacting all 3 regions. Because the SPR study indicated that residues 35–38 and 1–22 were more important than the ankyrin repeats, we manually selected for predictions where the PP1 contacted these 2 regions, of which there were many. The structure of the complex explains why none of our predictions could match perfectly with the biological data. The PP1 does indeed contact all 3 regions; however, the unbound PP1 was missing over 20 residues at its C-terminus. In the crystal structure of the complex,[30] 15 of those residues are included and they form, surprisingly, a straightened tail that fits neatly within the channel created by the ankyrin repeats [Fig. 1(D)]. Our

best prediction received a high accuracy rating and the interface RMSD for this submission was 0.95 Å.

Target 18 (GH11 Xylanase–TAXI)

In Target 18, we were asked to predict the complex of unbound *Aspergillus niger* xylanase, a Family 11 xylanase, and bound *Triticum aestivum* xylanase inhibitor (TAXI). The substrate binding site of *A. niger* xylanase had been postulated to be in the large cleft,[31] and competitive inhibition has been demonstrated in a Family 11 xylanase.[32] We therefore assumed that TAXI would inhibit by binding in the large cleft of *A. niger* xylanase. In keeping with this model, we blocked a scattered set of residues on the opposite face from the cleft on xylanase (residues 50, 102, 136, 141, and 154) to discourage ZDOCK predictions in that entire region. Two glutamate residues conserved in Family 11 xylanases, Glu79 and Glu170, have been implicated in catalysis.[33–36] In addition, mutational analysis on another Family 11 xylanase with 90% identity to *A. niger* xylanase showed that Asp37 was important to catalysis. We used a contact filter to search our prediction set for candidates where Asp37, Glu79, and Glu170 were sufficiently buried by contact with TAXI.

Our best prediction [Fig. 1(E)] received a medium-accuracy rating and was our first ranked submission. We correctly predicted 91% of the residue contacts; however, our predicted xylanase structure was tilted at its distal region, which led to an interface RMSD of 1.86 Å, preventing us from receiving a high-accuracy rating.[37]

Target 19 (Ovine Prion–FAB)

In Target 19, we were asked to predict the complex of the unbound ovine prion protein with a bound antibody. As in Target 11, we were told to homology model the ligand and then dock. We were given the sequence alignment between the ovine prion and the bovine prion, and the NMR structure for the latter. We surmised that an X-ray structure of the human prion protein [Protein Data Bank (PDB) ID: 1I4M] with slightly higher sequence identity than the NMR structure would make a better template for homology modeling. However, it is a dimer with a domain-swapped third helix that is severely dislocated when structurally aligned with the bovine prion. To circumvent this difficulty, we used only the first 192 residues of the human prion X-ray structure, and cut and pasted the remaining residues constituting the third helix and C-terminus from the bovine prion NMR structure. We then used the Swiss Model Homology Server to make the homology model using the ovine prion sequence and the hybrid template structure. Our method was successful, and the modeled structure had a backbone RMSD of 1.25 Å.

We blocked the 8 residues from each terminus of the ovine prion model because its sequence as given was a fragment. We assumed canonical antibody recognition and blocked the non-CDR regions of the bound antibody. Our best prediction [Fig. 1(F)] attained an accuracy rating of medium. The interface residues had a 1.26 Å RMSD, and we correctly predicted 57% contacts.[38]

DISCUSSION

ZDOCK and RDOCK together make a powerful combination for docking, and along with clustering and the application of biological information, including filtering and blocking, we were able to make many successful predictions for CAPRI. Our approach clearly leads to accurate results for many targets but has weaknesses in predicting certain types of docking cases. In the simple unbound–bound target cases where no modeling was necessary, thus excluding Targets 9, 11, and 19, our docking procedure involving ZDOCK and RDOCK worked extremely well. For all such cases, we attained medium or high accuracy in the CAPRI rating scale, and our highest interface RMSD among these cases was just 1.86 Å. In addition, each of such targets had at least 47% of its interface contacts correctly predicted, which demonstrates the strength of our approach toward binding-site prediction. In the unbound–bound targets that did require modeling, we did not achieve consistently successful predictions. Target 9 required a remodeling of 2 side-chains that led to a large conformational change that our docking procedure was not able to overcome. In Targets 11 and 19, the ligand was required to be homology modeled, and our group had very little expertise in the field of protein folding. The results of our predictions in those 2 cases were highly dependent on the accuracy of the homology models. In Target 19, which required a homology modeled ovine prion, we correctly predicted its structure and achieved medium accuracy for the predicted complex. For Target 11, we did not have a sufficiently accurate dockerin model, and our overall performance suffered, as evidenced by comparing our results of Targets 11 and 12. As shown in Figure 1(A and B), with the correct dockerin structure we were able to improve from just an acceptable accuracy rating to a high-accuracy rating with the same docking strategy. In addition, we had a 6-fold increase in our contact percentage and a 4-fold decrease in interface RMSD when bound dockerin was used instead of the homology model.

The most difficult targets are the ones with large conformational change. In CAPRI Rounds 3, 4, and 5, Targets 8, 9, and 10 fit in this category. The conformational change in Target 8 occurred mostly in the nonbinding fifth domain of laminin and did not significantly affect our methods. However, in both Targets 9 and 10, the conformational change severely hindered our prediction accuracy. Target 9 was equally challenging for the entire protein docking community, as indicated by the failure of all but 1 group to make a correct prediction. Although we use a moderately soft clash potential in our shape complementarity term in ZDOCK to allow for some overlap and thus small conformational change, the domain shift in Target 9 was too severe for our approach to make any successful predictions. In Target 10, we might have been able to overcome the conformational change that occurs in the C-terminal domain had we approached the initial preparation of the structure before running ZDOCK differently. We chose to block this domain rather than remove it completely from docking. As a result, it led to severe steric clash, which effectively excluded all correct predictions. Accounting for conformational change remains one of the most important areas in protein docking.

We believe ZDOCK and RDOCK represent a successful first step in the solution to the docking problem. Our performance in the last 3 rounds of CAPRI demonstrates that the 2 algorithms, as part of a comprehensive general docking strategy, can correctly identify at least 1 accurate structure out of 10 submissions for a large majority of blind-test cases. The diversity of protein complexes on which our approach is successful only furthers the idea that combining initial-stage docking techniques such as ZDOCK and RDOCK that can identify potential near-native structures with more sophisticated refinement methods that can account for side-chain rearrangements and backbone conformational changes could lead to great progress in the protein docking field. We look forward to testing this and other approaches in future rounds of CAPRI.

ACKNOWLEDGMENTS

We thank the CAPRI organizers and evaluation team for their extremely helpful efforts.

REFERENCES

1. Janin J, Henrick K, Moult J, Ten Eyck L, Sternberg MJE, Vajda S, Vasker I, Wodak SJ. CAPRI: A Critical Assessment of PRedicted Interactions. Proteins 2003;52:2–9.
2. 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.
3. Chen R, Tong WW, Mintseris J, Li L, Weng ZP. ZDOCK predictions for the CAPRI challenge. Proteins 2003;52:68–73.
4. Chen R, Weng ZP. Docking unbound proteins using shape complementarity, desolvation, and electrostatics. Proteins 2002;47:281–294.
5. Chen R, Weng ZP. A novel shape complementarity scoring function for protein–protein docking. Proteins 2003;51:397–408.
6. Chen R, Mintseris J, Janin J, Weng ZP. A protein–protein docking benchmark. Proteins 2003;52:88–91.
7. Mintseris J, Wiehe K, Pierce B, Anderson RJ, Chen R, Janin J, Weng Z. Protein–Protein Docking Benchmark 2.0: an update. Proteins 2005;60:214–216.
8. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4:187–217.
9. Li L, Chen R, Weng ZP. RDOCK: Refinement of rigid-body protein docking predictions. Proteins 2003;53:693–707.
10. Chen R, Weng Z. A novel shape complementarity scoring function for protein–protein docking. Proteins 2003;51:397–408.
11. Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. Proteins 2003;52:80–87.
12. Pierce B, Tong W, Weng Z. M-ZDOCK: a superior grid-based approach for C_n symmetric multimer docking. Bioinformatics 2005;21:1472–1478.
13. Poschl E, Mayer U, Stetefeld J, Baumgartner R, Holak TA, Huber R, Timpl R. Site-directed mutagenesis and structural interpretation of the nidogen binding site of the laminin gamma1 chain. EMBO J 1996;15:5154–5159.
14. DeLano WL. The PyMOL molecular graphics system. San Carlos, CA: DeLano Scientific; 2002.
15. Takagi J, Yang Y, Liu JH, Wang JH, Springer TA. Complex between nidogen and laminin fragments reveals a paradigmatic beta-propeller interface. Nature 2003;424:969–974.
16. Bower MJ, Cohen FE, Dunbrack RL Jr. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. J Mol Biol 1997;267:1268–1282.
17. Mintseris J, Weng Z. Atomic contact vectors in protein–protein recognition. Proteins 2003;53:629–639.
17a. Graille M, Zhou CZ, Receveur-Brechot V, Collinet B, Declerck N, van Tilbeurgh H. Activation of the LicT transcriptional antiterminator involves a domain swing/lock mechanism provoking massive structural changes. J Biol Chem 2005;280:14780–14789. PMID: 15699035.
18. 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.
19. 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.
20. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 2003;31:3381–3385.
21. 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.
22. Mechaly A, Fierobe HP, Belaich A, Belaich JP, Lamed R, Shoham Y, Bayer EA. Cohesin–dockerin interaction in cellulosome assembly: a single hydroxyl group of a dockerin domain distinguishes between nonrecognition and high affinity recognition. J Biol Chem 2001;276:9883–9888.
23. 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.
24. 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.
25. 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.
26. Tanaka J, Ito M, Feng J, Ichikawa K, Hamaguchi T, Nakamura M, Hartshorne DJ, Nakano T. Interaction of myosin phosphatase target subunit 1 with the catalytic subunit of type 1 protein phosphatase. Biochemistry 1998;37:16697–16703.
27. Hirano K, Phan BC, Hartshorne DJ. Interactions of the subunits of smooth muscle myosin phosphatase. J Biol Chem 1997;272:3683–3688.
28. Zhao S, Lee EY. A protein phosphatase-1-binding motif identified by the panning of a random peptide display library. J Biol Chem 1997;272:28368–28372.
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. Terrak M, Kerff F, Langsetmo K, Tao T, Dominguez R. Structural basis of protein phosphatase 1 regulation. Nature 2004;429:780–784.
31. Krengel U, Dijkstra BW. Three-dimensional structure of endo-1,4-beta-xylanase I from *Aspergillus niger*: molecular basis for its low pH optimum. J Mol Biol 1996;263:70–78.
32. Furniss CS, Belshaw NJ, Alcocer MJ, Williamson G, Elliott GO, Gebruers K, Haigh NP, Fish NM, Kroon PA. A family 11 xylanase from *Penicillium funiculosum* is strongly inhibited by three wheat xylanase inhibitors. Biochim Biophys Acta 2002;1598:24–29.
33. Bray MR, Clarke AJ. Identification of a glutamate residue at the active site of xylanase A from *Schizophyllum* commune. Eur J Biochem 1994;219:821–827.
34. Ko EP, Akatsuka H, Moriyama H, Shinmyo A, Hata Y, Katsube Y, Urabe I, Okada H. Site-directed mutagenesis at aspartate and glutamate residues of xylanase from *Bacillus pumilus*. Biochem J 1992;288:117–121.
35. Wakarchuk WW, Campbell RL, Sung WL, Davoodi J, Yaguchi M. Mutational and crystallographic analyses of the active site residues of the *Bacillus circulans* xylanase. Protein Sci 1994;3:467–475.
36. Miao S, Ziser L, Aebersold R, Withers SG. Identification of glutamic acid 78 as the active site nucleophile in *Bacillus subtilis* xylanase using electrospray tandem mass spectrometry. Biochemistry 1994;33:7027–7032.
37. 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.
38. 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.