Searching for Protein–Protein Interaction Sites and Docking by the Methods of Molecular Dynamics, Grid Scoring, and the Pairwise Interaction Potential of Amino Acid Residues

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · AUGUST 2005

Impact Factor: 2.63 · DOI: 10.1002/prot.20572 · Source: PubMed

5 AUTHORS, INCLUDING:

Genki Terashi
Kitasato University
24 PUBLICATIONS **203** CITATIONS
SEE PROFILE

Hideaki Umeyama
Kitasato University
165 PUBLICATIONS **2,533** CITATIONS
SEE PROFILE

Searching for Protein–Protein Interaction Sites and Docking by the Methods of Molecular Dynamics, Grid Scoring, and the Pairwise Interaction Potential of Amino Acid Residues

Genki Terashi,[†] Mayuko Takeda-Shitaka,[†] Daisuke Takaya, Katsuichiro Komatsu, and Hideaki Umeyama*
School of Pharmaceutical Sciences, Kitasato University, Tokyo, Japan

ABSTRACT **In CAPRI Rounds 1 and 2, we assumed that because there are many ionic charges that weaken electrostatic interaction forces in living cells, the hydrophobic interaction force might be important entropically. As a result of Rounds 1 and 2, the predictions for binding sites and geometric centers were acceptable, but those of the binding axes were poor, because only the largest benzene cluster was used for generating the initial docking structures. These were generated by fitting of benzene clusters formed on the surface of receptor and ligand. In CAPRI Rounds 3–5, the grid-scoring sum on the protein–protein interaction surface and the pairwise potential of the amino acid residues, which were indicated as coming easily into the protein–protein interaction regions, were used as the calculation methods, along with the smaller benzene clusters that participated in benzene cluster fitting. Good predicted models were obtained for Targets 11 and 12. When the modeled receptor proteins were superimposed on the experimental structures, the smallest ligand root-mean-square deviation (RMSD) values corresponding to the RMSD between the model and experimental structures were 6.2 Å and 7.3 Å, respectively. Proteins 2005;60:289–295.**

Key words: protein–protein docking; protein-binding site; hydrophobic interaction; benzene cluster; induced fit problem; interaction site prediction; CAPRI

INTRODUCTION

Prediction of the protein–protein docking structure, in addition to the prediction of protein–protein interaction sites, has been researched in biochemical, biophysical, genetic, and computational fields. In the CAPRI Rounds 1 and 2, predicted structure estimations of protein–protein interactions were carried out. Information on protein–protein docking structure and protein–protein interaction sites is very important for elucidating protein function and researching metabolic pathways in biological systems.

In this article, we present our results for CAPRI Rounds 3–5. In CAPRI Rounds 1 and 2, we especially remarked that the largest benzene cluster associated with a receptor or ligand protein is very significant in consideration of protein transfer in living cells, which are filled with strong electrolyte molecules.[1] Since additional benzene clusters besides the largest one might also be important, all benzene clusters containing more than 10 benzene molecules are included in the calculations. Consideration of these additional benzene clusters gives better predictions.

Other enthalpylike energies such as electrostatic, charge transfer, and polarization interactions should also be considered in protein–protein docking simulations. We account for these interactions by introducing pairwise potentials of 20 × 20 amino acid combinations[2] and the grid-scoring sum on the protein–protein interaction surface created using the Protein Quaternary Structure file server[3] (PQS) on the putative protein–protein complex in the prediction of the protein–protein interaction. With these modifications, we achieved "medium" and "acceptable" estimations of our predictions for Targets 11 and 12, respectively, from the assessors of CAPRI Rounds 3–5.

MATERIALS AND METHODS

Figure 1 shows the methods used in CAPRI Rounds 3–5, detailed descriptions of which are shown here. There are 4 steps:

1. *Benzene cluster (BC) fitting:* The protein–protein interaction site (PPI site) is searched using "benzene cluster fitting," which has previously been reported in detail.[1] This method is dependent upon the BC being easily formed around the hydrophobic surface of a globular protein in an aqueous solution. Molecular dynamics (MD) calculations[4] were executed for a system consisting of a protein, such as a receptor or ligand, benzene molecules initially placed near hydrophobic amino acid residues, and water molecules, in order to form clusters of benzene molecules. The BC must contain at least 10 benzene molecules; rough fittings between the BCs formed on the surface of each receptor and ligand were executed to dock

[†]These authors contributed equally to this work.

Grant sponsor: Grant-in-Aid for Scientific Research on Priority Areas (C) "Genome Information Science" from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

*Correspondence to: Hideaki Umeyama, Department of Biomolecular Design, School of Pharmaceutical Sciences, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan. E-mail: umeyamah@pharm.kitasato-u.ac.jp

Received 14 January 2005; Accepted 3 March 2005

DOI: 10.1002/prot.20572

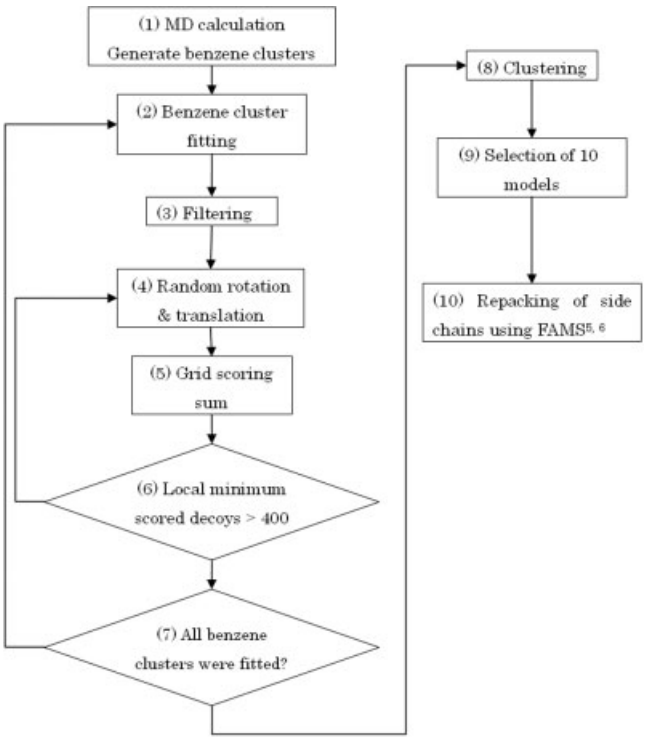

Fig. 1. Flow chart of the protein–protein docking method. (**1**) MD calculation, (**2**) BC fitting, (**3**) filtering of initial structures, (**4–7**) local minimum search for protein–protein interaction energy (grid-scoring sum), (**8**) clustering for the interacting places on the protein surfaces, (**9**) 10 models selected from the total scores of the possible protein–protein interaction places, and (**10**) repacking of the side-chain with a little movement of the main-chain, using the FAMS Complex homology modeling program. The grid-scoring sum step was used on Targets 14–19. In Targets 8–13, only BC fitting and amino acid residues pairwise potentials were used.

the ligand with the receptor. Three principal axes of each BC were determined using principal component analysis. Twenty-four initial structures along 3 principal axes were generated by a ligand BC fitting to a receptor BC. In addition, after this BC fitting, a filter was used to remove unsuitable complexes (decoys) with too many clashes. In the case of antigen–antibody complexes, some BCs were specially selected based on biological information.

2. *Docking based upon grid-scoring sum:* In the small areas around the interface regions identified by BCs, a grid-scoring method was used to search for energetically stable configurations, considering the 3-dimensional (3D) shape complementarity between receptor and ligand. We made a surface pattern database using the PQS server. The surface pattern represents the amino acid residues pattern (e.g., [Ala,Gly,Arg]) that exists within a constant distance from the grid position to the van der Waals surface of the protein. Grids were generated on surfaces of protein structures of the PQS server. In this step, we counted the sum of the number of grids with characteristic surface patterns in the protein–protein contact area (Con_i), and on the whole protein surface (C_i).

In the docking step, we generated grids on each protein surface, and calculated the grid score (Sco_i) in each grid [Eq. (1)]. Remarking grids existing on the each interaction site of various docking structures were determined, then these grid scores were summed:

$$Sco_i = -\log\left(\frac{\left(\frac{Con_i}{\sum_{k=0}^{n} Con_k}\right)}{\left(\frac{C_i - Con_i}{\sum_{k=0}^{n}(C_k - Con_k)}\right)}\right). \quad (1)$$

3. *Clustering and ranking by total score:* The most energetically stable protein–protein interaction structure was determined by giving scaled coefficients to the grid-scoring sum, the number of hydrogen bonds, and the pairwise interaction potential between 20 × 20 amino acid residue pairs. In CAPRI Rounds 3–5, the coefficients were determined depending on the biological information of each target. However in the grid-score calculations in the proposed models, consideration of the hydrogen bonds between proteins was not included in the case of Targets 8–12, because we had not yet developed this part of our docking procedure; that is, only the pairwise interaction potential of between 20 × 20 amino acids was used in these cases. The pairwise interaction potential was represented by Eq. (2).[2] $P(i,j)$ is the probability of amino acid residues i and j being in contact in the interaction site. $P(i)$ and $P(j)$ represent the probabilities of residues i and j having at least 1 contact with any residue across the interface. Accordingly, the matrix elements correspond to the Gibbs free energy–like quantity:

$$M(i,j) = -\log\frac{P(i,j)}{P(i)P(j)N}. \quad (2)$$

4. *Repacking of side-chains:* In the previous steps of the energetic optimization processes, in order to search for an energetically stable configuration, only Cα atoms and some Cβ atoms were included in the calculation. After side-chain packing, shorter distances than the allowed experimental values and clashes between side-chains of amino acid residues are consequently observed. To remove the side-chain clashes, we used our software FAMS Complex, a fully automated homology modeling system for protein complex structure. The modeling method of FAMS Complex was developed to be basically the same as that of our previously developed FAMS for monomer protein. The FAMS Complex was used for all targets in CAPRI Rounds 3–5. We also used FAMS in the 6th Community-Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (CASP6) experiment.

5. *Normal mode analysis (NMA):* Hinge regions that connect to domains are very important for considering domain movements of protein structures. In Target 9, in order to predict hinge regions, we used NMA. The energy optimization and NMA were calculated by programs developed by our laboratory.[7,8] Energy optimization was performed with a slightly modified AMBER united atom force field,[9] and NMA was performed with torsion angles. We

detected hinge regions that were defined by ϕ and ψ around Cα atoms having torsionally large fluctuations.

RESULTS

Here we report the best models of the proposed results for each target.

Target 08: Nidogen-G3 Domain–Laminin Epidermal Growth Factor (EGF) Modules 3–5 Complex[10] (Unbound–Bound)

Three BCs were generated on nidogen (chain B), and 3 BCs were generated on laminin (chain A) after MD calculations. We generated 216 ($24 \times 3 \times 3$) initial structures as shown in the Materials and Methods section; as generated in the BC fitting step. These were filtered manually according to the biological information,[11–13] shape complementarity, and number of clashes. After the docking and scoring process using the pairwise potential of 20×20 residues, 139 structures were obtained, and the 10 best-scoring structures were selected. We used FAMS Complex to refine the packing of side-chains in the complexes and submitted them to CAPRI.

From the *BC* fitting, the best predictions of residues in the interaction site (predicted–native) were 3.7% for nidogen-G3 and 8.3% for laminin, respectively. The fraction of predicted contacts over the native structure (f_{nat}) of our best prediction was only 0.015.

Target 09: LicT Dimer[14] (Unbound–Unbound)

In Target 09, the native structure was a homodimer with 2 domains. Our methods did not originally consider the large main-chain and domain movement; therefore, NMA[7,8] was also used for domain movement.

First, 5 BCs were generated on LicT (chain A, B) after MD calculations. From these clusters we derived 600 ($24 \times 5 \times 5$) initial structures in the BC fitting step. These were filtered according to symmetric properties, biological information,[15] and the number of clashes. In this step, clashes in the C-terminal domain were ignored, although this movement was considered later. Second, hinge regions were determined from NMA, as shown in the Materials and Methods section, and these were assigned at the ϕ of Glu168 and ψ of Met169. Third, the C-terminal domain was moved around the above hinge region in fixing the N-terminal domain, and the 10 structures with the best pairwise potential scores were selected.

The predicted interaction sites of LicT were acceptable. Two of the BCs predicted over 20% of the native interaction sites (Table I). The fraction of predicted contacts over the native (f_{nat}) of our best prediction was 0.244. However, since the orientation of the C-terminal domain was not predicted correctly, the value of ligand-RMSD was 11.881 Å. The reasons for this result were that there were too few hinge residues, and the searching process was done at too low a resolution.

Target 10: Tick-Borne Encephalitis Virus (TBEV) Glycoprotein E Trimer[16]

Our programs can only deal with 2 proteins at a time. For Target 10, we generated 520 trimers using 3-fold symmetry and calculation of the pairwise potential of amino acid residues.

Although, we did not use BC fitting for this target, some interaction sites were correctly predicted by the BCs (24% and 34%, as shown in Table I). None of our 10 models predicted correct contacts (Table II) due to the coarse search in the symmetry operation that corresponds to the use of longer distance interval and wide angle intervals. Accordingly, a high-resolution search should be done. The result is similar to that for Target 09.

Target 11: Cellulosome Cohesin–Dockerin Complex[17] (Unbound–Homology Model)

In Target 11, the dockerin structure was a homolog of entry 1DAQ. This prediction was an unbound–homology model docking problem. To make the dockerin structure, we used the FAMS homology modeling program.

Two BCs on the dockerin model obtained from the homology modeling and 2 BCs on cohesin (chain A; as receptor) were generated after MD calculations. We generated 96 ($24 \times 2 \times 2$) initial structures in the BC fitting step. No biological information was used for this target. After the docking and scoring process using the pairwise potential of the 20×20 residues, 2939 solutions were generated. The 10 best scoring structures were selected. Side-chain packing was refined with FAMS Complex and the resulting structures were submitted to CAPRI.

The interaction sites of cohesin and dockerin were correctly predicted by the BCs. As shown Table I, the first (largest) and second BCs of the dockerin model (chain B) predicted 57% and 36%, respectively, of the native interaction sites. The first (largest) benzene cluster of cohesin predicted 71% of the native interaction sites. Accordingly, the initial interaction structures obtained using the BC fitting methods were close to the native complex structure. For Target 11, the BC fitting and pairwise potential refinement combined well to give a better prediction structure.

Target 12: Cellulosome Cohesin–Dockerin Complex[17] (Unbound–Bound)

Target 12 is the same complex as Target 11, but the dockerin structure of Target 12 was determined from an X-ray structure as bound. One BC was generated on dockerin (chain B; as ligand), and 2 BCs were generated on cohesin (chain A; as receptor) after MD calculations. We generated 48 ($24 \times 2 \times 1$) initial structures in the BC fitting step. After the docking and scoring process using the pairwise potential of 20×20 residues, 428 solutions were generated, and the 10 best solutions were selected. Side-chain packing was refined with FAMS Complex and the results were submitted to CAPRI.

As shown in Table I, the first (largest) BC of dockerin predicted 79% of the native interaction sites, and the first BC of cohesin predicted 71% of the native interaction sites. These results were similar to those for Target 11. The fraction of predicted contacts over the native (f_{nat}) of our best submitted prediction was 0.527. The RMSD in the protein–protein interacting area was 2.4 Å, which was in good agreement with the experiment.

TABLE I. Summary of the PPI Site Predictions Obtained by Benzene Clusters[a]

				Cluster ID							
	Target	ch_ID		1st	2nd	3rd	4th	5th	6th	7th	8th
Round 3	08	A	*N*	24	19	17					
			Rate	0.083	0.000	0.000					
		B	*N*	14	12	12					
			Rate	0.037	0.000	0.000					
	09	A	*N*	18	17	16	15	10			
			Rate	0.293	0.171	0.000	0.000	0.244			
		B	*N*	18	17	16	15	10			
			Rate	0.146	0.171	0.220	0.244	0.146			
Round 4	10	A	*N*	40	30	19	17	14	14	12	10
			Rate	0.029	0.240	0.048	0.048	0.038	0.135	0.010	0.067
		B	*N*	39	34	24	19	11	11		
			Rate	0.143	0.339	0.071	0.000	0.107	0.161		
	11	A	*N*	21	19						
			Rate	0.710	0.000						
		B	*N*	21	17						
			Rate	0.571	0.357						
	12	A	*N*	21	18						
			Rate	0.710	0.000						
		B	*N*	25							
			Rate	0.786							
	13	A	*N*	27	13	11	10				
			Rate	0.600	0.000	0.000	0.000				
		B	*N*	27	2	11	10				
			Rate	0.683	0.000	0.000	0.000				
		F	*N*	24	20						
			Rate	0.160	0.600						
Round 5	14	A	*N*	18	14	11					
			Rate	0.069	0.224	0.017					
		B	*N*	35	31	18	10				
			Rate	0.299	0.299	0.000	0.030				
	15	A	*N*	11	10						
			Rate	0.080	0.000						
		B	*N*	16	10						
			Rate	0.150	0.150						
	18	A	*N*	36	17	14	12	11	10		
			Rate	0.792	0.000	0.000	0.000	0.000	0.000		
		C	*N*	20	12	11					
			Rate	0.000	0.000	0.688					
	19	A	*N*	36	20	12	10				
			Rate	0.636	0.000	0.000	0.045				
		H	*N*	46	22	19	15				
			Rate	0.000	0.667	0.000	0.000				
		L	*N*	46	22	19	15				
			Rate	0.077	0.692	0.000	0.000				

[a]The rate was computed as the ratio of the number of amino acid residues in the protein–protein interaction area that contact BCs within at least 5 Å to one atom of the interaction amino acid residues, divided by the number of contacts in the target. *N* is the number of benzene molecules determined as a BC in accord with our definition.

Target 13: *Toxoplasma gondii* Surface Antigen (SAG-1)–FAB Complex[18] (Unbound–Bound)

Four BCs were generated on FAB (chains A, B), and 2 BCs were generated on SAG-1 (chain F) after MD calculations. We generated 192 ($24 \times 4 \times 2$) initial structures in the BC fitting step. These were filtered according to the complementarity determining region (CDR) contacts and number of clashes. The 192 initial structures were reduced to 48 structures. As a result of docking and scoring for the 48 structures using the pairwise matrix of the 20×20 residues, 1133 solutions were generated. Side-chain packing was refined with FAMS Complex and the results were submitted to CAPRI.

Because of human error and program bugs, the BC fitting step was not performed correctly, but the interaction sites were correctly predicted by BCs. The prediction ratios were 60% and 68% for FAB and 60% for SAG1, as

TABLE II. Our Best Predictions for Targets 08–19[a]

				fIR				
	Predictions	f_{nat}	$f_{non\text{-}nat}$	Ligand	Receptor	Distance	L_rmsd	I_rmsd
Round 3	T08_P09.10.A	0.015	0.989	0.154	0.733	12.43	17.799	5.319
	T09_P09.8.B	0.244	0.828	0.688	0.708	10.46	11.881	13.929
Round 4	T10_P03.1.B	0	1	0.119	0.122	50.292	66.161	—
	T11_P03.7.B	0.345	0.725	0.842	0.914	4.238	6.179	2.681
	T12_P03.1.B	0.527	0.678	0.789	0.857	4.281	7.281	2.401
	T13_P03.8.Hl	0.014	0.989	0.281	0.68	20.424	30.248	15.792
Round 5	T14_P68.7.B	0.045	0.919	0.263	0.412	29.325	36.622	7.223
	T15_P68.1.B	0.054	0.936	0.308	0.577	4.162	17.526	9.376
	T18_P68.10.A	0.074	0.898	0.41	0.485	10.475	26.852	12.842
	T19_P68.1.A	0.049	0.951	0.36	0.724	10.177	22.312	9.728

[a]Measured in fractions of the predicted contacts over native (f_{nat}). $f_{non\text{-}nat}$ represents the fraction of non-native predicted contacts over prediction. fIR stands for the interface residue ratios over native. Distance (in Angstroms) is between the geometric centers of predicted and target ligand molecules. L_rmsd is the RMSD values calculated between the ligand's backbones once the corresponding receptors are superimposed. I_rmsd refers to the RMSD values when superimposing the backbones of the residues at the interface on the prediction of the counterpart in the target.

shown in Table I. The best solution after correction of the BC fitting step gave a ligand-RMSD of 15.12 Å (122th/311), but this solution was not submitted for evaluation.

Target 14: Protein Phospatase 1-β (PP1-β)–Myosin Phosphatase Targeting Subunit 1 (MYPT1) Complex[19] (Unbound Homology Model–Bound)

In Target 14, PP1-β had to be homology-built from PP1-α [Protein Data Bank (PDB) entry: 1FJM]. To make the PP1-β structure, we used the FAMS homology modeling program.

Three BCs were generated on the PP1-β model (chain A; as receptor), and 4 BCs were generated on MYPT1 (chain B; as ligand) after MD calculations. We generated 288 ($24 \times 3 \times 4$) initial structures in the BC fitting step. These were filtered manually according to shape complementarity and the number of clashes. The 288 initial structures were reduced to 43. After docking the 43 structures using the grid scoring sum (see Materials and Methods section), 2908 solutions were generated. We used FAMS Complex to refine the packing of the side-chains in the complex of the 10 best models as determined by the total score (grid-scoring sum + residue–residue pairwise potential + number of H-bonds), and submitted them to CAPRI.

One BC (second) on the PP1-β model predicted 22% of the native interaction sites, and 2 BCs (first, second) on MYPT1 predicted over 20% of the native interaction sites. These results were better for the prediction of protein interaction sites.

In the best submitted prediction, the position of the helix and loop (1–36) in the N-terminal of MYPT1 (ligand) was reasonably good (Fig. 2). However, the relative position of the C-terminal moiety was not predicted correctly. The fraction of the predicted contacts over the native (f_{nat}) of our best prediction was only 0.045.

Target 15: Colicin D Nuclease–Immunity D (ImmD) Complex[20]

Target 15 was canceled, but we will describe our result. In this target, the surface side-chains were shaved to the Cβ, and the orientation was random. In the first step, we generated side-chains of both Target 15 structures using FAMS.

Two BCs on the catalytic domain of colicin D (chain A; as receptor), and 2 BCs on the ImmD protein (chain B; as ligand) were generated after MD calculation. We generated 96 ($24 \times 2 \times 2$) initial structures in the BC fitting step. These were filtered according to the shape complementarity and the number of clashes. The 96 initial structures were reduced to 55 structures. After docking the 55 structures by grid-scoring sum, 10,592 solutions were generated. The 10 best models were selected based on total score (grid score + residue–residue pair potential + number of H-bonds). Side-chain packing was refined with FAMS Complex, and the results were submitted to CAPRI.

The largest BC on the catalytic domain of colicin D predicted only 8% of the native interaction sites, and the 2 BCs (first, second) on the ImmD protein predicted 15% of the native interaction sites. After colicin D structures were superimposed, the distance between the geometric centers of predicted and target ligand molecule was 4.162 Å; however, the relative position was not correct. The reason for this result was that only the interaction sites of ImmD were predicted correctly. As shown in Figure 2, the predicted colicin D (blue) existed at the ImmD interaction site of the native. This result suggests that the interaction sites of both the receptor and ligand have to be predicted correctly.

Target 18: *A. niger* Xylanase–*T. aestivum* Xylase Inhibitor-I (TAXI) Complex[21]

Six BCs on xylanase (chain C; as ligand), and 3 BCs on TAXI (chain A; as receptor) were generated after MD calculations. We generated 432 ($24 \times 6 \times 3$) initial structures in the BC fitting step. These were filtered according to the shape complementarity and the number of clashes. The 432 initial structures were reduced to 48. After the docking process of the 48 structures using the grid-scoring sum, 3906 solutions were generated. The 8 best models were selected based on total score (grid

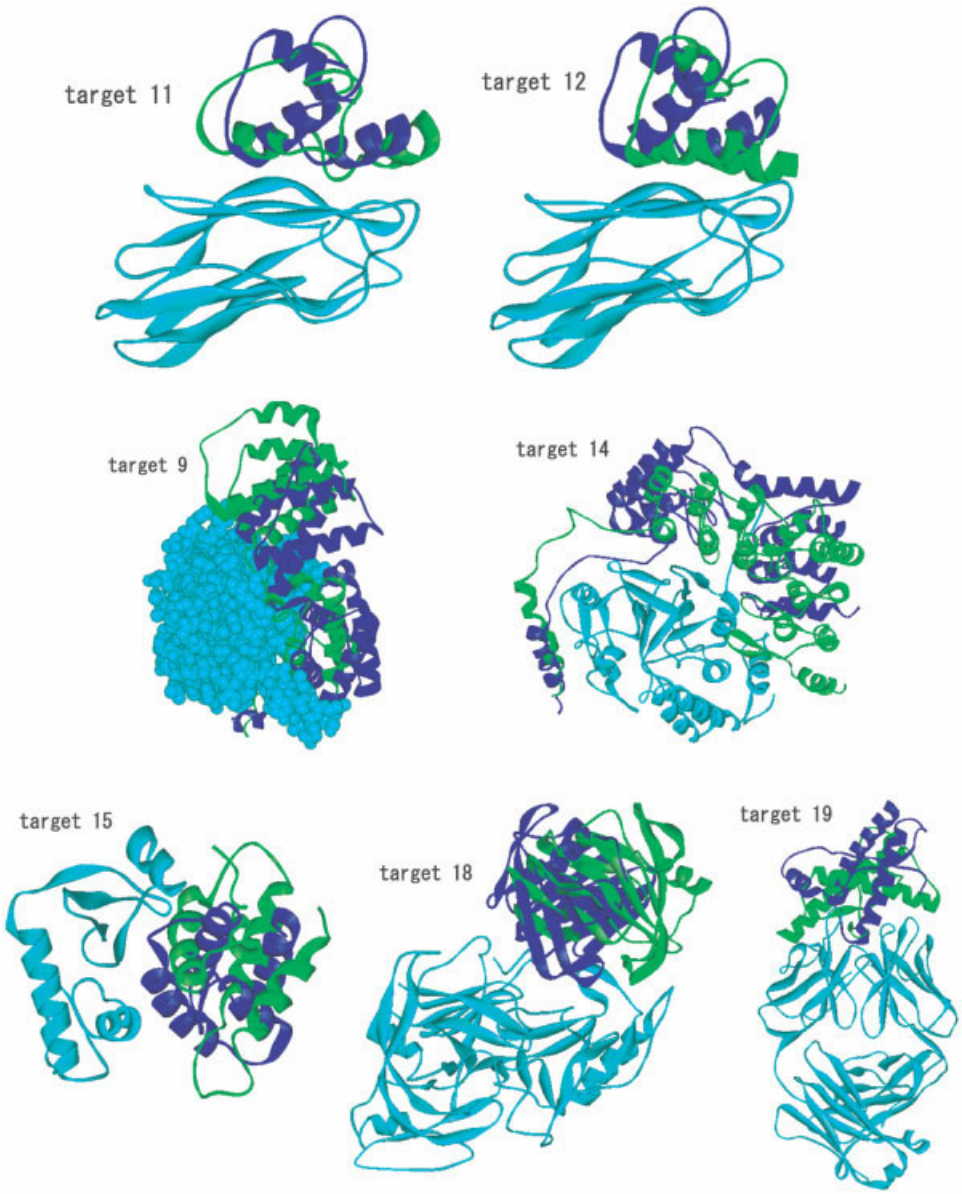

Fig. 2. Our best models (green) and native structures (blue: ligand; light blue: receptor).

score + residue–residue pair potential + number of H-bonds), and we also selected the best grid-scoring model and the best pairwise potential model. Side-chain packing was refined with FAMS Complex, and the results were submitted to CAPRI.

Xylanase and TAXI interaction sites were correctly predicted by the BCs. As Table I shows, the third BC of xylanase predicted 69% of the interaction sites, and the first BC (largest) of TAXI predicted 79% of the native interaction sites.

When observing the relative position of the ligand to the receptor, the position of the center of gravity of the best result was good (Fig. 2), but a worse relative position was observed due to changes in translation and rotation. The fraction of the predicted contacts over native (f_{nat}) of our best submitted prediction was only 0.074. The ligand-RMSD was 26.852 Å. But in 3906 solutions of the result of docking based upon grid-scoring sums, the best model near-native solution produced a ligand-RMSD of 3.23 Å for the 444th of the 3906 structures. This suggests that we have to improve the scoring function.

Target 19: Ovine prion–FAB Complex[22] (Homology Model–Bound)

In Target 19, first we generated the ovine prion structure from homology modeling (FAMS) with PDB entry 1DWY. Four BCs were generated on the ovine prion model (chain A; as ligand), and 4 BCs were generated on FAB (chains H, L; as receptor) after MD calculations. We generated 384 ($24 \times 4 \times 4$) initial structures in the BC fitting step. These were filtered according to the CDR contacts and the number of clashes. The 384 initial

structures were reduced to 72. After docking the 72 structures using the grid-scoring sum, 1961 solutions were generated. The 10 best models were selected based on total score (grid score + residue–residue pair potential + number of H-bonds). Side-chain packing was refined with FAMS Complex, and the results were submitted to CAPRI.

The interaction sites of the ovine prion model and FAB were correctly predicted by the BCs. As Table I shows, the first (largest) BC of the ovine prion model predicted 64% of the native interaction sites, and the second BC of FAB predicted 67% and 69% of the native for the H and L chains, respectively. But the fraction of predicted contacts over the native (f_{nat}) of our best prediction was only 0.049. Ligand-RMSD was 22.312 Å.

DISCUSSION

As shown in Table I, in CAPRI Rounds 3–5, almost all interaction sites for receptors and ligands were correctly predicted in the BC fitting step. The interaction sites were predicted with fewer than 20% of the native interface residues for only 2 (Targets 08 and 15) of the 10 targets. Moreover, for Targets 11 and 12, both of the determined structures were close to the experimental or native structure after refinements such as the pairwise potential calculations following the BC fitting step.

Nevertheless, even if the predicted interaction sites were better due to BC fitting, there were many targets with bad predictions of the protein–protein interaction structure. These results indicate that the binding direction and binding angles between the receptor and ligand were worse. Our present problems are the coarse samplings for rotation angles and translation distances, and the strong dependence upon the accuracy of the BC fittings; we cannot get the correct docking structure if the initial docking structures generated are wrong. The final total-score function is still insufficiently accurate in the last ranking step for many models. Additionally, a docking method that considers a more revised protein fluctuation and flexibility than those performed in this article was needed in the case of Targets 9 and 10.

In the next CAPRI experiments, a theoretically careful flexible docking method will be needed for improved predictions, in addition to the development of a method for determination of a precise initial docking structure, accurate packing of amino acid residues in the complex, and the useful final-score function for selecting the correct docking structure.

ACKNOWLEDGMENTS

We thank the CAPRI organizers and the assessors for this experiment. We are greatly thankful to Professor Lynn F. Ten Eyck of University of California at San Diego for kind and conscientious revision of our article. We thank Dr. Youji Kurihara and Hirokazu Tanaka of Kitasato University for helpful discussion.

REFERENCES

1. Komatsu K, Kurihara Y, Iwadate M, Takeda-Shitaka M, Umeyama H. Evaluation of the third solvent clusters fitting procedure for the prediction of protein–protein interactions based on the results at the CAPRI blind docking study. Proteins 2003;52:15–18.
2. Glaser F, Steinberg DM, Vakser IA, Ben-Tal N. Residue frequencies and pairing preferences at protein–protein interfaces. Proteins 2001;43:89–102.
3. Henrick K, Thornton JM. PQS: a protein quaternary structure file server. Trends Biochem Sci 1998;23:358–361.
4. Yoneda S, Umeyama H. Free energy perturbation calculations on multiple mutation bases. J Chem Phys 1992;97:6730–6736.
5. Takeda-Shitaka M, Takaya D, Chiba C, Tanaka H, Umeyama H. Protein structure prediction in structure based drug design. Curr Med Chem 2004;11:551–558.
6. Ogata K, Umeyama H. An automatic homology modeling method consisting of database searches and simulated annealing. J Mol Graph Model 2000;18:258–272, 305–306.
7. Adachi M, Kurihara Y, Nojima H, Takeda-Shitaka M, Kamiya K, Umeyama H. Interaction between the antigen and antibody is controlled by the constant domains: normal mode dynamics of the HEL–HyHEL-10 complex. Protein Sci 2003;12:2125–2131.
8. Kamiya K, Sugawara Y, Umeyama H. Algorithm for normal mode analysis with general internal coordinates. J Comput Chem 2003;24:826–841.
9. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S Jr, Weiner P. A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 1984;106:765–784.
10. 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.
11. Stetefeld J, Mayer U, Timpl R, Huber R. Crystal structure of three consecutive laminin-type epidermal growth factor-like (LE) modules of laminin gamma1 chain harboring the nidogen binding site. J Mol Biol 1996;257:644–657.
12. Jeon H, Meng W, Takagi J, Eck MJ, Springer TA, Blacklow SC. Implications for familial hypercholesterolemia from the structure of the LDL receptor YWTD–EGF domain pair. Nat Struct Biol 2001;8:476–478.
13. Mayer U, Nischt R, Poschl E, Mann K, Fukuda K, Gerl M, Yamada Y, Timpl R. A single EGF-like motif of laminin is responsible for high affinity nidogen binding. EMBO J 1993;12:1879–1885.
14. Graille M, Zhou CZ, Receveur-Brechot V, Colinet 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(15):14780–14789.
15. van Tilbeurgh H, Le Coq D, Declerck N. Crystal structure of an activated form of the PTS regulation domain from the LicT transcriptional antiterminator. EMBO J 2001;20:3789–3799.
16. 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.
17. 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.
18. Graille M, Stura E, Bossus M, Muller BH, Letourneur O, Battail-Poirot N, Sibai G, Gauthier M, Rolland D, Le Du MH, Ducancel F. Structure of the immunodominant epitope displayed by the surface antigen 1 (SAG1) of *Toxoplasma gondii* complexed to a monoclonal antibody. Submitted.
19. Terrak M, Kerff F, Langsetmo K, Tao T, Dominguez R. Structural basis of protein phosphatase 1 regulation. Nature 2004;429:780–784.
20. 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.
21. 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.
22. 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.