PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Improved flexible refinement of protein docking in CAPRI rounds 22–27

Yang Shen*

Toyota Technological Institute at Chicago, Chicago, Illinois 60637

ABSTRACT

Since the fourth evaluation for critical assessment of prediction of interactions (CAPRI), we have made improvements in three major areas in our refinement approach, namely the treatment of conformational flexibility, the binding free energy model, and the search algorithm. First, we incorporated backbone flexibility into our previous approach, which only optimized rigid backbone poses with limited side-chain flexibility. Here, we formulated and solved the conformational search as a hierarchical optimization problem (involving rigid-body poses, backbone flexibility, and side-chain flexibility). Second, we used continuum electrostatic calculations to include solvation effects in the binding free energy model. Finally, we eliminated sloppy modes (directions in which the free energy is essentially constant) to improve the efficiency of the search. With these improvements, we produced correct predictions for 6 of the 10 latest CAPRI targets, including one high, three medium, and two acceptable accuracy predictions. Compared to our previous performance in CAPRI, substantial improvements have been made for targets requiring homology modeling.

Proteins 2013; 00:000–000.

Key words: protein docking; structural refinement; protein flexibility; normal modes; free energy minimization; energy funnel; sloppy modes; hierarchical optimization; continuum electrostatics.

INTRODUCTION

Protein docking provides an alternative to classical experimental methods (such as X-ray crystallography, nuclear magnetic resonance, and cryo-electron microscopy) by predicting atomic-level structural details of protein complexes and guiding our understanding of driving forces.[1–3] It plays an especially important role for complexes whose structures could not be analyzed easily under experimental conditions. Over the past 12 years, critical assessment of prediction of interactions (CAPRI) has provided a unique opportunity for our community to assess and advance protein docking methods on unknown targets.[4–7]

Protein docking methods generally incorporate a multistage scheme,[8] where more accurate free energy (or scoring function) models and more protein flexibility are gradually introduced. We have focused on refinement methods that start with initial hits and aim for high-accuracy predictions. Previously, we developed a refinement method named semidefinite underestimation (SDU) that minimizes a funnel-like free energy function in a medium range.[9,10] For a set of locally minimized conformational samples in a given search region, SDU first constructs an optimal quadratic underestimator by solving a semidefinite programming problem. It then shifts and shrinks the search region, discards previous samples outside, and biases new sampling within, all based on the optimal underestimator. The two processes of underestimating and sampling are iterated until convergence.[9] We also found that the parameterization of conformational space was critical to its success. Therefore, we devised an exponential-coordinate-based parameterization of the space of rigid-body motions (a Riemannian manifold *SE(3)*) and treated interface side-chain flexibility with local minimization.[10] This method was applied to problems from previous CAPRI rounds at different stages during its development and had considerable success.[11,12] However, the exclusion of backbone flexibility limited its success, especially for targets requiring homology modeling.

In this article, we summarize our recent improvements in flexible refinement methods and their

Grant sponsor: Toyota Technological Institute at Chicago.
*Correspondence to: Yang Shen, Toyota Technological Institute at Chicago, Chicago, IL 60637. E-mail: yangshen@ttic.edu
Received 17 June 2013; Revised 15 August 2013; Accepted 21 August 2013
Published online 28 August 2013 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24404

performance in the latest rounds of CAPRI. Improvements were made modeling protein flexibility, calculating binding free energy, and searching conformational space for the minimum. First, we incorporated backbone flexibility by adding nested optimization for internal flexibility to our previous optimization, which only treated rigid-body poses with limited side-chain flexibility. We solved the nested optimization problem with either generic minimization methods in the space of Cartesian coordinates or with free energy funnel-driven SDU in the space of backbone normal modes (with side chains locally optimized). Second, we improved the binding energy model by considering solvation effects with continuum electrostatic calculations. Finally, we improved the search efficiency by avoiding directions along which binding free energy is relatively insensitive (termed "sloppy modes"). With the methodological advances, we have produced 1 high-accuracy, 3 medium-accuracy, and 2 acceptable predictions for all 10 targets since the fourth CAPRI (7 of the 10 required homology modeling); of the set, three predictions had the lowest ligand root mean square deviations (RMSDs) among all submissions.

METHODS

Homology modeling

All targets but T48, T49, and T58 required homology modeling, which was achieved with webservers I-TASSER,[13] Robetta,[14] and ModWeb.[15] Structural templates or sequence alignments provided to participants were deliberately disregarded for generalizability concerns.

Initial rigid-body protein docking

For all targets but T57 (which involves a polysaccharide), we used ClusPro 2.0,[16] a webserver for rigid protein docking, to generate initial models. ClusPro ranks fast Fourier transform solutions based on four independent sets of binding energy coefficients (default, electrostatics favored, hydrophobics favored, and van der Waals and electrostatics only) and clusters the top 1000 for each set. Multiple pairs of monomer structures were used as inputs when multiple homology models existed. Throughout the article, the monomer being fixed in position during docking is termed a "receptor" and the other a "ligand".

We used AutoDock4[17] for T57. The protein was treated as rigid and the polysaccharide was allowed to have rotatable bonds. 1000 independent Lamarckian Genetic algorithm trajectories with random starting points were run on a grid with 0.6-Å spacing to generate a local minimum in each. The 1000 local minima were clustered with a 9-Å radius in a greedy approach provided by AutoDock4 and the centers of the first 100 clusters were retained as initial models.

Flexible refinement

Binding free energy models

Two types of binding free energy models were used hierarchically. Each conformational sample was locally minimized with the CHARMM22 energy model[18] (with the exception of T57, for which CHARMm22[19] was used). A distance-dependent dielectric constant of 4r was used for electrostatics. No distance cutoff for nonbonded interactions was used. The 1–2 and 1–3 interactions were excluded for the calculation of van der Waals terms. In addition, when the CHARMm22 energy model was used, 1–4 interactions were scaled down by half.

Conformational search based on the resulting local minima was performed using a more accurate binding free energy model. The electrostatic contribution, including receptor and ligand desolvation terms and screened receptor–ligand electrostatic interactions, was computed with continuum electrostatic calculations[20,21] using the computer program DelPhi.[22] Partial atomic charges and atomic radii for proteins were from the PARSE parameter set.[23] For large receptors (such as T48 and T49), partial charges for noninterfacial atoms (>9 Å away from any ligand atom) were set to zero to save computational cost. Heparin was geometrically optimized at the restricted Hartree–Fock 6–31 G* level with the program Gaussian 03[24] and its partial atomic charges were derived though restrained electrostatic potential fitting. Nonelectrostatic contributions included intermolecular van der Waals interactions (CHARMM22) and a nonpolar contribution to solvation that was dependent on the solvent-accessible surface area.[23]

Hierarchical optimization of rigid-body poses, backbone flexibility, and side-chain flexibility

The rigid-body pose of a ligand relative to its receptor was optimized by SDU in an outer loop. For each given rigid-body pose, backbone conformations were additionally optimized together with side chains in an inner loop (see Fig. 1).

We incorporated the modeling of backbone flexibility using two methods. In the first method used for earlier targets, we used a combination of rigid-body local moves and all-atom local minimization as implemented in CHARMM.[25] The conformational space, parameterized in Cartesian coordinates of movable atoms, is thus of high-dimensionality (typically hundreds to thousands) and very redundant in describing collective motions of proteins. The success of any generic optimization method in such a space would be limited. This concern is addressed in the second method, in which the backbone conformational space was separated from that of side chains, parameterized with normal modes, and searched by SDU. Specifically, normal modes of either monomer were calculated with an anisotropic network model using the program

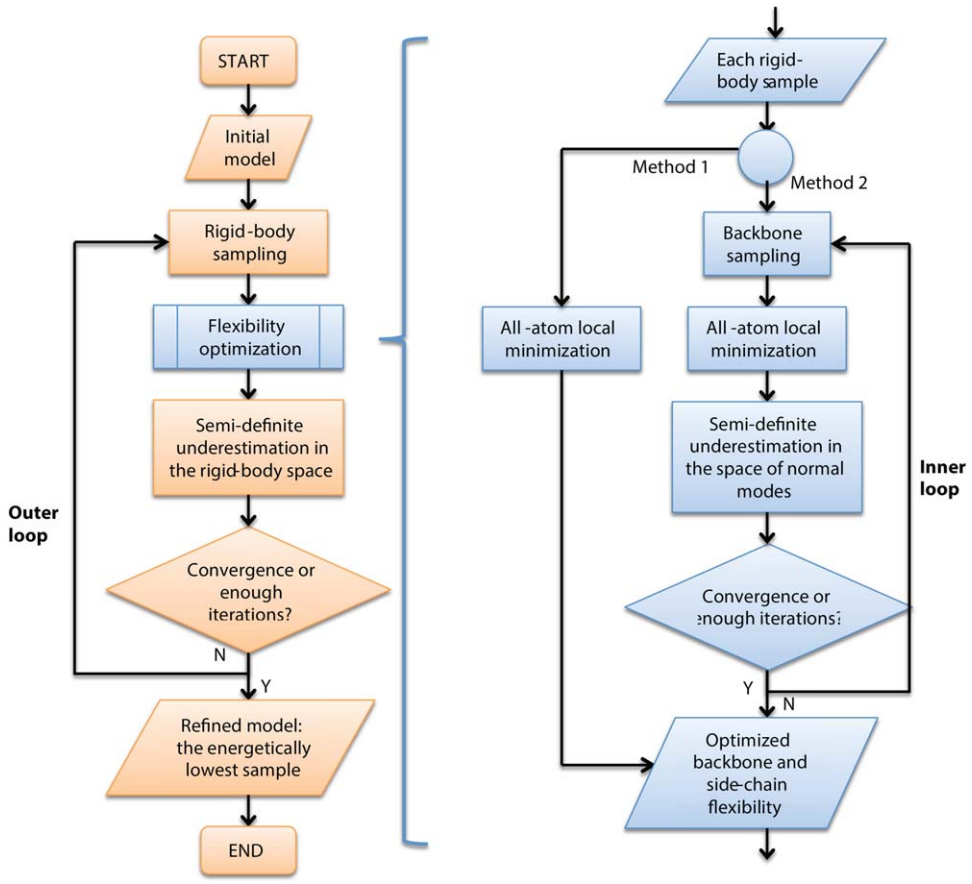

Figure 1
Flowchart of the refinement method. Details on the step of flexibility optimization are given on the right.

ProDy.[26] Nodes in the network were C_α atoms only. The default cutoff distance (15 Å) and force constant (1.0) were used. The coarse-grained normal modes were then extended from every C_α atom to all other atoms in the same residue by assuming that they share the same components of normal modes. For any given rigid-body pose of the ligand, SDU was used to optimize in the six-dimensional (6D) space spanned by three slowest nontrivial normal modes for receptor and three for ligand. Each sample in the inner loop was allowed to deform from its starting conformation by at most 2 Å in backbone RMSD. The optimized backbone and side-chain conformations for all rigid-body poses sampled then provided inputs for SDU to optimize the pose in the outer loop. The energetically lowest structure after at most five iterations was retained as a refined model.

Avoiding searches along sloppy modes of free energy funnels

When constructing optimal quadratic underestimators, sloppy modes were often found in the form of eigenvectors corresponding to small eigenvalues of Hessian matrices. Biased sampling was only performed in the subspace spanned by the top principal components of the Hessian matrix (eigenvectors whose absolute values of eigenvalues were at least 1% of the maximum such value). Coordinates along the other eigenvectors (sloppy modes) were adopted from those of the current lowest energy sample.

The refinement of each initial model was performed on a local computer cluster consisting of Intel Xeon X5650 processors (2.66 GHz CPU with 4 G of memory per core) and took about 1.5–4 CPU hours for the first method and 16–40 CPU hours for the second. The majority of the running time was spent on calculating continuum electrostatics.

Scoring

All refined models were first ranked energetically using the more accurate energy model. The first 10 (enthalpically favorable) nonredundant models originating from large clusters (the top 5 or 10 in cluster membership for each set of energy coefficients, thus entropically favorable) were usually our final models submitted. For T57, cluster sizes from AutoDock4 were not used. Structurally close models were sometimes observed, which reflected

Table I
Performance on Seven Regular, Pairwise Protein–Protein Targets

Target index	Name	Type	Our performance[a]	Community performance[a]
T46	Mtq2–Trm112	Homology–homology	0*	13*
T47	E2 DNase–Im2	Homology–unbound	3*** + 4** + 2*	95*** + 108** + 13*
T48	Hydroxylase–ferredoxin	Unbound–unbound	3** + 3* (hexamer)[b] 3** + 2* (trimer)[b]	4** + 17* (hexamer) 11** + 39* (trimer)
T49	Hydroxylase–ferredoxin	Unbound–unbound	1** + 3* (hexamer)[b] 1** + 2* (trimer)[b]	1** + 11* (hexamer) 1** + 35* (trimer)
T50	HA-HB36.3	Bound[c]–homology	2*	17** + 35*
T53	Rep4–Rep2	Unbound–homology	1**	1*** + 11** + 31*
T54	Neocarzinostatin–Rep16	Unbound–homology	1*[b]	6*

[a]***, **, and * indicate predictions classified to be of high, medium, and acceptable accuracy, respectively, in official assessments.
[b]Our predictions included one with the lowest ligand-RMSD values among all submissions for specified targets.
[c]HA (hemagglutinin) was given as bound structure with another protein (not HB36.3).

consensus of initial-stage energy models with different weighting schemes or/and convergence of refinement trajectories originating from different initial models. For such refined models, one or two low-energy representatives were chosen and given a higher rank. No biological information was used here (or in earlier docking steps) due to resource limitations.

RESULTS

The 10 CAPRI targets can be classified into two sets: one with seven regular targets involving pairwise protein–protein interactions and another with four new types of targets (Target 47 involved both regular predictions and predictions for interface water). The seven regular targets include five cases where monomer structures need to be built by homology modeling (Table I), and are thus potentially difficult. Nevertheless, our flexible refinement approach led to correct predictions for all seven except T46. In contrast, we did not perform well for the new types of targets (Table II) where our pairwise protein–protein docking methods do not readily apply and only provided fair-quality water predictions for T47.

Target 46: methyl transferase Mtq2–Trm112

T46 is a two-subunit eRF1 (eukaryotic release factor 1) methyl transferase.[27] Both subunits were provided as sequences and therefore 3D structural models were first built using Robetta and I-TASSER. For either subunit, the highest ranked model from either server was retained, which resulted in four combinations of input-structure pairs for rigid-body docking and flexible refinement.

None of our 10 models submitted was correct and even our best model M03 had a ligand RMSD of 22.7 Å. Only the Bonvin group and their HADDOCK[28] webserver provided acceptable predictions. Posterior analysis indicated that the failure in T46 could be partially attributed to the Trm112 homology models. Even though the overall backbone RMSD for I-TASSER-modeled Mtq1 and Trm112 were only 2.3 Å and 5.0 Å, respectively, that for the modeled C-terminal region (residues 117–125) of Trm112 was 18.1 Å. Therefore, the modeled Trm112 C-term would have blocked part of the Trm112 interface and clashed with the bound Mtq2 even if the Trm112 homology model were structurally aligned to the bound structure [see Fig. 2(A)]. A similar story held for the

Table II
Performance on New Types of Targets

Target index	Name	Type	Our performance[a]	Community performance[a]
(T47 water)[b]	Mtq2–Trm112	Interface water prediction (homology–unbound)	7*	11***+60**+75*
T51	[GH5–CBM6][c]–CBM13–Fn3	Multi-domain assembly (unbound–homology–unbound)	0* (T51.1) 0* (1* not selected for final 10) (T51.2) 0* (T51.3)	3* (T51.1) 1* (T51.2) 0* (T51.3)
T57	BT4661–Heparin	Protein–polysaccharide (unbound–sequence[d])	0*	5** + 26*
T58	PliG–SalG	Additional SAXS data (unbound–unbound)	0*	15** + 18*

[a]Except for T47 water prediction, ***, **, and * indicate predictions classified to be of high, medium, and acceptable accuracy, respectively, in official assessment. For T47 water prediction, ***, ***, **, and * indicate predictions classified to be of outstanding, excellent, good, and fair accuracy, respectively.
[b]Community-wide T47 water prediction will be summarized in a separate manuscript. Therefore, our performance on this target was provided but not counted toward the overall statistics for 10 targets in this study.
[c]Both domains of GH5 and CBM6 were provided in an unbound structure.
[d]The heparin was provided in a random conformation. We did not have tools to perform homology modeling for this polysaccharide.

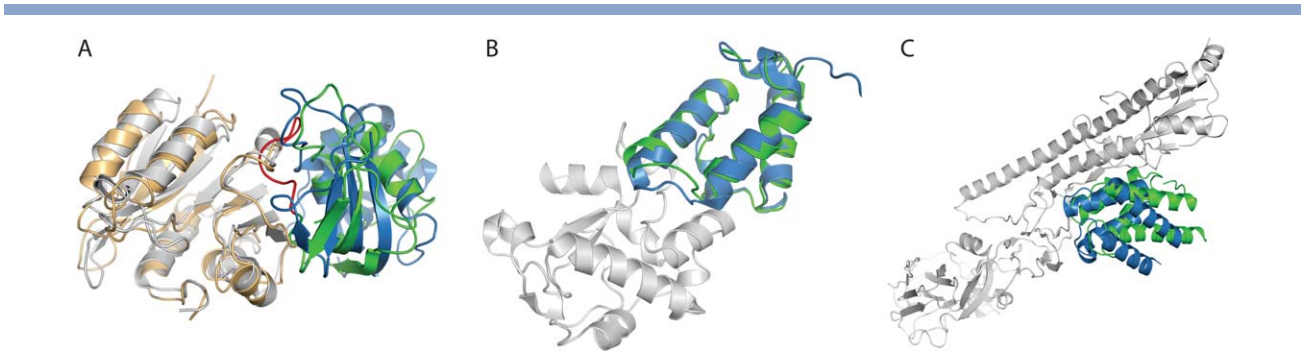

Figure 2

Our predictions for three regular targets in comparison with bound, crystal structures in cartoon representations. We also made acceptable or better predictions for the four other regular targets (T48, T49, T53, and T54) that were not officially published. (**A**) Target 46: methyl transferase Mtq2–Trm112. Bound Mtq2 and Trm112 are shown in gray and blue, respectively, and their aligned homology models (from I-TASSER) in wheat and green, respectively. The incorrectly modeled C-term of Trm112 is shown in red, differing from the rest of the Trm112 homology model (in green). We had no correct prediction for this target. (**B**) Target 47: colicin E2 DNase–Im2 complex. Bound E2 DNase is shown in gray, and bound Im2 and our high-accuracy prediction M05 are in blue and green, respectively. (**C**) Targets 50: influenza hemagglutinin (HA) and a designed protein HB36.3. Bound HA is shown in gray, and bound HB36.3 and one of our acceptable predictions (M09) are in blue and green, respectively.

Robetta-modeled Trm112 C-terminal region. The low sequence similarity to templates (around 12% for Mtq2 and 20% for Trm112) could be responsible.

Target 47: colicin E2 DNase–Im2 complex

For T47,[29] colicin E2 DNase domain (E2 in short) was provided as a sequence and its cognate immunity protein Im2 was given as its unbound structure. An E2 homology model was first built with I-TASSER. In addition, with crystal structures of E9–IM9 and E9–IM2 [Protein Data Bank (PDB) accession codes 1EMV and 2WPT, respectively] structurally aligned to the E2 model, only five interfacial waters that were conserved across both complexes and hydrogen bonded with E9 regions conserved in E2 were retained and attached to the E2 model before docking.

All of our models submitted except the last (M10) were correct in regular assessment, including three high-accuracy (M04, M05, and M06) and four medium-accuracy predictions. The best model (M05) had a ligand RMSD of 0.95 Å (improved from a 3.4-Å ClusPro model) and 61% native contacts predicted successfully [Fig. 2(B)]. Only two other groups had predictions with ligand RMSD <1 Å. However, our water predictions were only fair for seven models and their native, interfacial water fractions were all below 30%. We believe that "soaking" procedures to introduce additional interfacial waters after docking would improve the performance, especially for the high-accuracy regular predictions.

Targets 48 and 49: hydroxylase–ferredoxin complexes

Both targets involve binding of a heterohexamer hydroxylase and a ferredoxin given in unbound 3D coordinates. The difference lies in the hydroxylase coordinates which were biological units constructed from PDB entry 3DHH for T48 but new coordinates for T49. The ferredoxin structure was as in PDB entry 1VM9. 6 (5) of our 10 models for T48 were correct in hexamer (trimer) assessment, including three medium-accuracy ones (M03, M04, and M08). The best model (M04) had the lowest ligand RMSD of 3.6 Å (3.4 Å) in hexamer (trimer) assessment among all models submitted and predicted 47.9% of native contacts. M04 was refined from the second largest cluster generated by ClusPro with default energy coefficients. For T49, 4 (3) of our 10 models were at least acceptable in hexamer (trimer) assessment, including the only medium-accuracy model (M05) from the community. M05 (4.5-Å ligand RMSD and 47.9% native contacts) was refined from the fourth largest cluster using hydrophobics-favored energy coefficients. Two acceptable models (M01 and M03) were also refined from large clusters generated by ClusPro. Although originating from different energy coefficients, they all had low binding-energy values after refinement.

Targets 50: influenza hemagglutinin and a designed protein (HB36.3)

HB36.3 was a protein designed by the Baker group to bind hemagglutinin (HA) based on APC36109 from *Bacillus stearothermophilus* (PDB accession code 1U84).[30] The sequence of HB36.3 was provided along with its alignment to APC36109, whereas HA was in its bound structure with another protein (PDB accession code 3GBN). Five homology models of GB36.3 were built using I-TASSER without the information provided on template or alignment and then used for rigid-body docking and flexible refinement.

We had two models with acceptable accuracy: M02 and M09. In particular, M09 came from the fifth largest cluster (first homology model of HB36.3 and default energy coefficients in ClusPro) whose center had a 7.0-Å ligand RMSD. The refinement slightly increased the ligand RMSD to 7.5 Å, which implies that even our improved binding energy models may not discriminate structures within 0.5 Å. Compared to the native structure, it was rotated by roughly 28° around an axis perpendicular to the interface [Fig. 2(C)] and predicted 57.1% native contacts.

Targets 51: multidomain assembly of GH5–CBM6–CBM13–Fn3

T51 is the only multi-domain target. It consists of a glycoside hydrolase family five catalytic domain (GH5), a carbohydrate binding module family 6 and 13 (CBM6 and CBM13), and fibronectin type III-like module (Fn3)[31] (another domain, CBM62, is mobile and no prediction was requested). GH5–CBM6 was provided as an unpublished structure, CBM13 as a sequence with a template (PDB accession code 1KNL), and Fn3 as an unbound structure (PDB accession code 3MPC).

Five homology models of CBM13 were first built with I-TASSER, each of which was randomly rotated, attached to Fn3 at a terminus, and energetically minimized by CHARMM. This was repeated 100 times and the energetically lowest structure among all 500 was refined by CHARMM using ten 10-ns implicit-solvent molecular dynamics simulations. Electrostatics was computed with the EEF1 model.[32] The energetically lowest CBM13–Fn3 model at 10 ns was then rigidly docked to GH5–CBM6 using ClusPro and refined with10-ns MD simulations. The ten energetically lowest 10-ns structures were chosen as the final 10 models submitted. The simple approach of incrementally assembling multiple domains did not produce a correct model. No other groups were successful either, although two produced acceptable models for one interface (T51.1) and one did for another (T51.2). In hindsight we missed an acceptable prediction for T51.2 due to scoring. Among the 100 energetically lowest 10-ns structures, M68 was assessed to be an acceptable prediction for T51.2 (complex of GH5–CBM6 and CBM13).

Targets 53: designed Rep4–Rep2 α-repeat complex

Rep4 was provided as an unbound structure and Rep2 as a sequence. 3 homology models were built using ModWeb with default parameters. The first two homology models were based on PDB entry 3LTJ and the last on 3L9T. We produced a medium-accuracy model M04. This model had the second-lowest ligand RMSD (3.1 Å) and the highest fraction of native contacts (76.9%) among all submissions. It was generated by refining the seventh largest cluster from ClusPro (using the second homology model of Rep2 and the hydrophobics-favored energy coefficients). It was the fourth lowest energetically refined model from a top 10 (thus considered large) cluster and agreed well with the seventh lowest such model; it was chosen as our fourth submission.

Targets 54: designed neocarzinostatin–Rep16 α-repeat complex

T54 involves neocarzinostatin provided as an unbound structure (PDB accession code 2CBO) and Rep16, another α-repeat protein, as a sequence. ModWeb generated three homology models for Rep16 that, like those for Rep2 in T53, were based on template 3LTJ, 3LTJ, and 3L9T, respectively. Our docking approach produced an acceptable model M07, whose ligand RMSD was 5.1 Å (the lowest among all submissions) with the fraction of native contacts being 28.0%. M07 was refined from the second largest cluster from ClusPro when the first homology model of Rep16 and energy coefficients only for van der Waals and electrostatics were used. Four groups in total had acceptable predictions.

Targets 57: a BT4661–heparin complex

T57 is the first protein–polysaccharide target in CAPRI. BT4661 was provided as an unbound structure, and a six-unit heparin as a random one. None of our submissions was correct. Our best model M09 was nearly acceptable, with 14.2-Å ligand RMSD and 41.2% native contacts.

Targets 58: PliG–SalG lysozyme complex with small-angle X-ray scattering data

T58 is the first CAPRI target with low-resolution experimental data provided. During the 3-week experiment, we had no time to develop methods that involve small-angle X-ray scattering (SAXS) data in protein docking and thus did not use those data. None of our submissions was correct. After examining all 74 refined models from which we selected final submissions, we found an acceptable model with a ligand RMSD of 9.0 Å. However, it was not chosen as it neither originated from a large cluster (16th largest from ClusPro using default energy coefficients) nor ranked high in energy values.

DISCUSSION

The 10 targets in the latest rounds of CAPRI have presented at least two types of challenges to protein docking. The first is in the difficulty of regular, pairwise protein docking—five of seven such targets required homology modeling for one or both monomers. We produced correct predictions for all seven except T47, including one high, three medium, and two acceptable

Table III
Cosine Similarities (in Absolute Values) Between the Normal Modes Calculated for I-TASSER Homology Models and the Deviations of the Models From Target Crystal Structures

Target index	Protein	Backbone RMSD (Å)	Sim_{all}^{max} [a]	$rank_{all}$	Sim_{top3}^{max} [b]	$rank_{top3}$
T46	Mtq2	2.86	0.31	12	0.17	3
T46	Trm112	6.83	0.36	3	0.36	3
T47	E2 DNase	1.02	0.36	3	0.36	3
T50	HB36.3	0.65	0.47	6	0.36	3

[a] Sim_{all}^{max} is the highest cosine similarity (in the absolute value) between any normal mode and the deviation.
[b] Sim_{top3}^{max} is the highest cosine similarity (in the absolute value) between any of the slowest three normal modes and the deviation.

accuracy predictions. These results represent a substantial improvement from our previous performances for homology-docking targets (one medium-accuracy prediction for seven targets).[12] It is also noteworthy that they were produced with "blind" homology modeling without the use of structural templates or sequence alignment provided and "blind" protein docking without the use of biological information from the literature. That being said, (1) finding correct templates and addressing sequence alignment uncertainty especially for sequences of distant homologs, which are important to homology-based docking, have progressed but remain unsolved[33]; (2) docking two protein sequences without building individual structures first is increasingly gaining attention[34] as high-throughput sequencing techniques are determining protein sequences with unprecedented speed; and (3) biological information on protein interactions would boost the accuracy and the efficiency of protein-docking methods if used properly.[28] The second challenge brought here is in the diversity of targets. Three targets represent new types of docking tasks: multidomain assembly (T51), protein–polysaccharide interactions (T57), and protein–protein interactions with low-resolution information such as SAXS data (T58). We did not succeed in these targets as our methods were not ready to address such challenges but we expect to continue the development of corresponding methods.

Our improved refinement approach decomposes conformational space into that of rigid-body poses, backbone flexibility, and side-chain flexibility, which are treated hierarchically during conformational search. Correspondingly, the free energy minimization problem is projected into lower dimensional spaces and thus easier to handle. The three types of conformational variables act on different spatial ranges during protein–protein association, which provides a potential biophysical rationale for the decomposition approach. Questions remain in the speed of convergence that we will explore in future. As to the parameterization of conformational space, normal modes provide a biophysically sound and numerically efficient way to represent backbone flexibility. This parameterization was extensively studied for protein interactions[35–37] and applied to protein docking.[38,39] We found that normal modes of homology models correlated to varying extents with the deviations of these models from target crystal structures (see Table III).

Another improvement involves avoiding sloppy modes of free energy funnels: free energy values are relatively insensitive to conformational perturbations along these directions. These directions are observed in our method as eigenvectors corresponding to small eigenvalues (in absolute values) of the Hessian matrices of our free energy funnel underestimators (quadratic functions). The existence of such sloppy modes indicates that not all conformational coordinates are equally important in protein association and is reminiscent of association pathways in a subspace of all conformational coordinates.

ACKNOWLEDGMENTS

We thank CAPRI organizers, assessors, and other committee members as well as structural biologists who provided the targets, for providing a unique opportunity to advance protein docking methods. We appreciate that many people have made their computer programs or webservers available, which were very helpful to this work. In particular, we thank Barry Honig for DelPhi, Sandor Vajda for ClusPro, David Baker, Andrej Sali, and Yang Zhang for Robetta, ModBase, and I-TASSER, respectively, Ahmet Bakan and Ivet Bahar for ProDy, and Martin Karplus for CHARMM. We also thank Bruce Tidor, Sandor Vajda, and Jinbo Xu for providing comments on the manuscript and Adam Bohlander for providing administrative help on a computer cluster.

REFERENCES

1. Wodak SJ, Janin J. Computer analysis of protein-protein interaction. J Mol Biol 1978;124:323–342.
2. Smith GR, Sternberg MJE. Prediction of protein–protein interactions by docking methods. Curr Opin Struct Biol 2002;12:28–35.
3. Wiehe K, Peterson MW, Pierce B, Mintseris J, Weng Z. In: Zaki MJ, Bystroff C, editors. Methods in Molecular Biology, Protein Struct. Predict. Vol. 413. Humana Press., Totowa, NJ; 2008. pp 283–314.
4. 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.
5. Méndez R, Leplae R, Lensink MF, Wodak SJ. Assessment of CAPRI predictions in rounds 3–5 shows progress in docking procedures. Proteins 2005;60:150–169.
6. Lensink MF, Méndez R, Wodak SJ. Docking and scoring protein complexes: CAPRI 3rd edition. Proteins 2007;69:704–718.
7. Lensink MF, Wodak SJ. Docking and scoring protein interactions: CAPRI 2009. Proteins 2010;78:3073–3084.
8. Vajda S, Kozakov D. Convergence and combination of methods in protein–protein docking. Curr Opin Struct Biol 2009;19:164–170.
9. Paschalidis ICh, Shen Y, Vakili P, Vajda S. SDU: a semidefinite programming-based underestimation method for stochastic global optimization in protein docking. IEEE Trans Autom Control 2007; 52:664–676.
10. Shen Y, Paschalidis ICh, Vakili P, Vajda S. Protein docking by the underestimation of free energy funnels in the space of encounter complexes. PLoS Comput Biol 2008;4:e1000191.

11. Shen Y, Brenke R, Kozakov D, Comeau SR, Beglov D, Vajda S. Docking with PIPER and refinement with SDU in rounds 6–11 of CAPRI. Proteins 2007;69:734–742.
12. Kozakov D, Hall DR, Beglov D, Brenke R, Comeau SR, Shen Y, Li K, Zheng J, Vakili P, Paschalidis ICh, Vajda S. Achieving reliability and high accuracy in automated protein docking: ClusPro, PIPER, SDU, and stability analysis in CAPRI rounds 13–19. Proteins 2010; 78:3124–3130.
13. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 2008;9:40.
14. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 2004;32:W526–W531.
15. Eswar N, John B, Mirkovic N, Fiser A, Ilyin VA, Pieper U, Stuart AC, Marti-Renom MA, Madhusudhan MS, Yerkovich B, Sali A. Tools for comparative protein structure modeling and analysis. Nucleic Acids Res 2003;31:3375–3380.
16. Comeau SR, Kozakov D, Brenke R, Shen Y, Beglov D, Vajda S. ClusPro: performance in CAPRI rounds 6–11 and the new server. Proteins 2007;69:781–785.
17. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 2009;30:2785–2791.
18. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck J, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith J, Stote R, Straub J, Watanabe M, Wiorkiewicz–Kuczera J, Yin D, Karplus MJ. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.
19. Momany FA, Rone R. Validation of the general purpose QUANTA 3.2/CHARMm force field. J Comput Chem 1992;13:888–900.
20. Lee L-P, Tidor B. Optimization of binding electrostatics: charge complementarity in the barnase-barstar protein complex. Protein Sci 2001;10:362–377.
21. Shen Y, Gilson MK, Tidor B. Charge optimization theory for induced-fit ligands. J Chem Theory Comput 2012;8:4580–4592.
22. Nicholls A, Honig BH. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson–Boltzmann equation. J Comput Chem 1991;12:435–445.
23. Sitkoff D, Sharp KA, Honig B. Accurate calculation of hydration free energies using macroscopic solvent model. J Phys Chem 1994; 98:1978–1988.
24. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA. Gaussian 03, Revision C.02. Wallingford, CT: Gaussian, Inc.; 2004.
25. Brooks BR, Brooks CL III, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: the biomolecular simulation program. J Comput Chem 2009;30:1545–1614.
26. Bakan A, Meireles LM, Bahar I. ProDy: protein dynamics inferred from theory and experiments. Bioinformatics 2011;27: 1575–1577.
27. Liger D, Mora L, Lazar N, Figaro S, Henri J, Scrima N, Buckingham RH, van Tilbeurgh H, Heurgué-Hamard V, Graille M. Mechanism of activation of methyltransferases involved in translation by the Trm112 'hub' protein. Nucleic Acids Res 2011;39:6249–6259.
28. Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc 2003;125:1731–1737.
29. Wojdyla JA, Fleishman SJ, Baker D, Kleanthous C. Structure of the ultra-high-affinity colicin E2 DNase–Im2 complex. J Mol Biol 2012; 417:79–94.
30. Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch EM, Wilson IA, Baker D. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. Science 2011;332:816–821.
31. Najmudin S, Pinheiro BA, Prates JA, Gilbert HJ, Romão MJ, Fontes CM. Putting an N-terminal end to the Clostridium thermocellum xylanase Xyn10B story: crystal structure of the CBM22-1–GH10 modules complexed with xylohexaose. J Struct Biol 2010;172:353–362.
32. Lazaridis T, Karplus M. Effective energy function for proteins in solution. Proteins 1999;35:133–152.
33. Mariani V, Kiefer F, Schmidt T, Haas J, Schwede T. Assessment of template based protein structure predictions in CASP9. Proteins 2011;79:37–58.
34. Mukherjee S, Zhang Y. Protein–protein complex structure predictions by multimeric threading and template recombination. Structure 2011;19:955–966.
35. Petrone P, Pande VS. Can conformational change be described by only a few normal modes? Biophys J 2006;90:1583–1593.
36. Bakan A, Bahar I. The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. Proc Natl Acad Sci 2009;106:14349–14354.
37. Dobbins SE, Lesk VI, Sternberg MJE. Insights into protein flexibility: the relationship between normal modes and conformational change upon protein–protein docking. Proc Natl Acad Sci 2008;105: 10390–10395.
38. Mashiach E, Nussinov R, Wolfson HJ. FiberDock: flexible induced-fit backbone refinement in molecular docking. Proteins 2010;78: 1503–1519.
39. Moal IH, Bates PA. SwarmDock and the use of normal modes in protein–protein docking. Int J Mol Sci 2010;11:3623–3648.