ATTRACT: Protein–Protein Docking in CAPRI Using a Reduced Protein Model

Martin Zacharias*
International University Bremen, Computational Biology, School of Engineering and Science, Bremen, Germany

ABSTRACT **Protein–protein complex structures have been predicted for CAPRI Rounds 3 and 5 using a reduced protein model. Proteins are represented by up to 3 pseudoatoms per amino acid. The docking approach termed ATTRACT is based on energy minimization in translational and rotational degrees of freedom of one protein with respect to another protein. The reduced protein model allows one to perform systematic docking minimization of many thousand start structures in reasonable computer time. Flexibility of critical surface side-chains can be accounted for by a multiple conformational copy approach. The multicopy approach allows simultaneous adjustment of side-chain conformations and optimization of translational and rotational degrees of freedom of one protein with respect to the partner during docking. For 3 (Targets 8, 14, and 19) out of 5 CAPRI targets, the approach resulted in predictions in close agreement with experiment [root-mean-square deviation (RMSD) of backbone atoms within 10 Å of the protein–protein interface < 1.8 Å]. The comparison of predicted and experimental structures of the CAPRI targets indicates that besides local conformational changes (e.g., changes in side-chain conformations), global conformational changes of the protein backbone can be critical for complex formation. These conformational changes not accounted for during docking are a likely reason for the unrealistic predictions in 2 cases (Targets 9 and 18). Proteins 2005;60:252–256.**

Key words: protein–protein interaction; docking minimization; flexible docking; biomolecular modeling; mean field optimization

INTRODUCTION

Almost all biological processes involve protein–protein interactions. A detailed understanding of the function of these protein–protein interactions requires the knowledge of the 3-dimensional (3D) protein–protein complex structure. Since it is currently not possible to determine all protein–protein complex structures experimentally, the realistic prediction of protein–protein complexes (protein–protein docking) is of great importance. The CAPRI challenge[1] offers the opportunity to evaluate and compare different methods and protocols for protein–protein docking. We have participated in CAPRI Rounds 3 and 5, and predicted the complex structures for Targets 8, 9, 14, 18, and 19. The protein–protein docking approach, ATTRACT,[2] is based on energy minimization of protein partners and a reduced protein model to allow systematic docking with many thousands of start configurations. The interaction between amino acids in the reduced model considers differences in physicochemical character of the side-chains such that complex formation is driven not only by surface complementarity but also by the physicochemical character of the interacting protein surfaces. During Rounds 3 and 5, our docking approach included the possibility to include the flexibility of protein surface side-chains approximately by using a multiple conformational copy approach. Promising results were achieved in 3 out of 5 target structures. However, in 2 cases, none of the predicted structures were close to experiment, probably due to significant backbone conformational differences between bound and unbound protein conformations in these cases. A possibility to account for such changes using approximate normal mode analysis will be discussed.

MATERIALS AND METHODS

The ATTRACT docking approach and the reduced protein model have been described in detail in a previous publication.[2] In the following, a brief description of the general method and the model is given. In a first step, the protein partner coordinates are translated into a reduced protein presentation consisting of up to 3 pseudoatoms per amino acid residue. One pseudoatom represents the protein backbone (located at the Cα position). Small amino acid side-chains (Ala, Asp, Asn, Ser, Thr, Val, and Pro) are represented by 1 pseudoatom (geometric mean of side-chain heavy atoms), whereas larger and more flexible side-chains are represented by 2 pseudoatoms.[2] The effective interaction between pseudoatoms is described by a soft distance-dependent Lennard–Jones (LJ)-type potential (A/r^8-B/r^6-potential). The repulsive and attractive LJ-parameters describe approximately the size and physicochemical character of the side-chain chemical groups. Note that in contrast to purely residue-based potentials, the description of side-chains by 2 pseudoatoms allows us,

Grant sponsor: Deutsche Forschungsgemeinschaft.

*Correspondence to: Martin Zacharias, International University Bremen, Computational Biology, School of Engineering and Science, Campus Ring 6, D-28759 Bremen, Germany. E-mail: m.zacharias@iu-bremen.de

Received 16 January 2005; Accepted 1 February 2005

DOI: 10.1002/prot.20566

in principle, to account for the dual chemical character of some side-chains. For example, lysine can be considered as an amino acid with a nonpolar hydrophobic tail and a charged hydrophilic head group. Systematic tests of the model on "bound" protein partners indicate that rigid-body minimization of the experimental complex structures yields energy-minimized complex structures with an RMSD of the ligand protein from the experimental position of ~1–2 Å,[2] which is comparable to energy minimization using atomistic models. During docking, both partners are considered as rigid. Flexibility of the partner structures is taken into account by representing flexible surface side-chains (and also loops in a newer version) as multiple conformational copies. In the case of side-chains, the conformational copies represent possible rotameric states of the surface side-chains. The best fitting conformational copies are automatically selected during docking simultaneously with the minimization in translational and rotational degrees of freedom of the protein partners. Selection is based either on a switching method[2] or by using a weighted mean field method.[3]

For systematic docking studies, one of the proteins (usually the smaller protein, called the ligand protein) was used as probe and placed at various positions and various orientations on the surface of the second fixed (receptor) protein. A probe radius was chosen that was slightly larger than the maximum distance of any atom from the ligand center. At each starting position on the receptor protein, various initial ligand protein orientations were generated. The docking from each start position consisted of a series of energy minimizations in translational and rotational degrees of freedom of the ligand protein with respect to the receptor protein. During the first minimization, a harmonic restraint between the center of the fixed protein and the closest Cα-pseudoatom of the ligand protein was applied. This first minimization served to generate a close contact between the 2 proteins. For the subsequent energy minimizations, the ligand protein was free to move to the closest energy minimum. Approximately 10,000–15,000 complexes (in the case of medium-sized protein partners with < 200 residues) can be energy-minimized to low residual gradients in about 1 h on a 2 GHz Linux PC.

Experimental data and knowledge of possible residues involved in protein–protein interaction can be taken into account at various stages of the docking procedure. This includes the possibility to restrict the search to regions that are known to interact with the second protein partner or distance restraints that enforce a putative contact.

RESULTS AND DISCUSSION

Targets and Predictions

In CAPRI Rounds 3 and 5, we submitted predictions for Targets 8 and 9, 14, 18, and 19 (experimental structures for Targets 15–17 were published before the CAPRI submission deadline). A summary of some of the predictions is given in Table I.

TABLE I. Results of CAPRI Predictions

Target	Predictions closest to experiment (rank)	$Rmsd_I$ (Å)	$Rmsd_L$ (Å)	Fraction native contacts
08	1	0.9	7.0	0.39
09	1	9.5	12.1	0.18
14	2 (1)	0.58 (0.6)	1.2 (1.6)	0.6 (0.59)
18	1	22.5	69.0	0.0
19	9 (4)	1.8 (2.5)	5.3 (9.6)	0.64 (0.36)

$Rmsd_I$ is the root-mean-square deviation between prediction and experiment of protein backbone atoms within 10 Å of the protein–protein interface. $Rmsd_L$ indicates the deviation of the (smaller) ligand protein from experiment after superposition of receptor protein atoms. Numbers in parenthesis are for the second-best predictions.

Target 8: laminin–nidogen complex

Target 8 corresponds to the complex between a β-propeller domain of nidogen G3 and the laminin epidermal-growth-factor-like (LE) modules III3–5. The bound structure of nidogen and an unbound structure of the LE modules III3–5[4] were given. Mutagenesis studies were available that demonstrated the importance of 3 amino acids (Asp 800, Asn 802, and Val 804[5]) in the middle domain of the LE modules for binding to nidogen. During the systematic search, the nidogen G3 domain was treated as fixed receptor and the laminin modules as mobile ligand protein. Out of the generated low-energy complexes obtained by docking minimization, only those were retained with the 3 critical amino acids (Asp800, Asn802, and Val804) in the central laminin module located at the interface between laminin and nidogen. Representative complex structures out of 10 clusters of low-energy complexes were selected and submitted as target predictions. Of these, the top-ranking complex structure showed an interface RMSD (of all atoms within 10 Å of the protein protein interface) smaller than 0.9 Å compared to the experimental structure,[6] and a fraction of ~0.4 of native contacts (Table I). However, the average deviation of the complete predicted laminin ligand placement with respect to the nidogen protein was 7 Å compared to the laminin position in the experimental structure. This displacement is mainly due to a slightly different orientation of the elongated laminin molecule in the prediction compared to experiment (Fig. 1). Nevertheless, in terms of interface RMSD with respect to the experiment, our prediction for this target was among the best submitted solutions.

Target 9: wild-type LicT regulator dimer

The task was to predict the dimer structure of the wild-type antiterminator protein LicT involved in transcription regulation.[9] An experimental dimer structure of a double mutant (His207Asp and His269Asp) was available,[9] and the monomers out of this structure served as starting structures for modeling the wild-type protein and for docking. Each LicT monomer consists of 2 globular domains. During docking, the LicT monomers were treated as rigid structures assuming no significant conformational change due to replacement of Asp207 and Asp269 by His residues. It turned out that in the experimental wild-type LicT structure,[10] the interacting monomers adopt a domain arrangement that differs significantly from the ar-

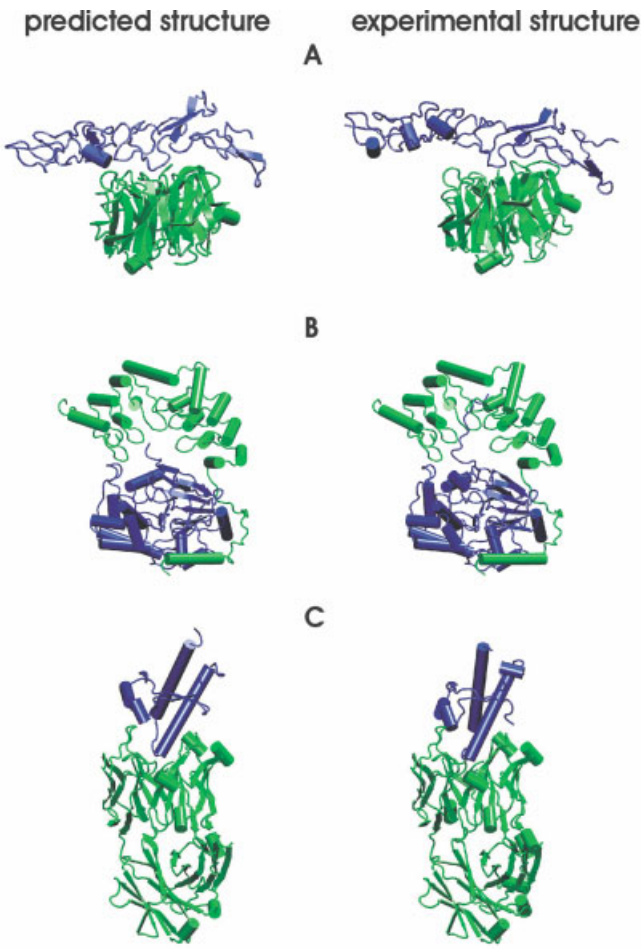

Fig. 1. Side-by-side comparison of predicted (**left panels**) and experimental (**right panels**) complex structures for 3 CAPRI targets (cartoon representation). (**A**) Target 8: Laminin modules III3—5 (blue) in complex with nidogen (green, experimental structure[6]). (**B**) Target 14: Protein phosphatase-1 (blue) in complex with myosin phosphatase–targeting subunit (green, experimental structure[7]). (**C**) Target 19: Ovine prion protein (blue) in complex with an antibody FAB fragment (green, experimental structure[8]). For Targets 8 and 14, top-ranking predictions are shown, and for Target 19, the prediction with rank 9 is shown.

rangement in the structure of the LicT mutant.[9] Not surprisingly, since this large conformational change was not considered during docking, all predicted complex structures deviated significantly from experiment (Table I).

Target 14: protein phosphatase-1 in complex with myosin phosphatase targeting subunit

For this target, the myosin phosphatase–targeting subunit (MYPT1) structure was given in the bound form.[7] For the second partner, the chicken protein phosphatase-1 (PP1), a homology model based on the high sequence similarity to the human PP1 with known structure[11] in the unbound form was built. In addition, experimental evidence was available concerning an RVxF sequence motif in MYPT1 for which the corresponding interaction region in PP1 was known.[12] This experimental information was used to restrict the search start structures for PP1 to the concave region between the N-terminal α-helix and the ankyrin repeats of MYPT1. Docking minimization of these start structures resulted in several low-energy docking complexes, of which the lowest energy structure and representative low-energy structures from several clusters were submitted as predictions. The top-ranked prediction and also several other submitted solutions showed small interface RMSD of <0.7 Å and a fraction of native contacts of >0.5 with respect to the experimental structure[7] (Table I and Fig. 1). A C-terminal tail of PP1 (residues 303–307) missing in the human unbound crystal structure,[11] and which contacts MYPT1 in the experimental complex structure, was not included in the predictions (affects the fraction of native contacts). The success in this case is also influenced by the fact that PP1 undergoes only little conformational change upon complex formation with MYPT1 and the use of MYPT1 in the bound form.

Target 18: xylanase–inhibitor (TAXI-I) complex

For Target 18, the *Aspergillus niger* xylanase was given in the unbound form,[13] and the xylanase–inhibitor from *Triticum aestivum* (TAXI-I) in the bound form.[14] Systematic docking searches were performed using the xylanase as receptor and the TAXI inhibitor as ligand protein. All submitted predictions showed significant deviation from the experimental docking geometry.[14] Most of the submitted solutions involved a TAXI-I surface loop region (residues 39–45) at a different location than the loop regions involved in binding of the experimental complex structure. After the experimental complex structure was published,[14] a systematic docking run was performed using both partners in the bound form. In this docking run, a docking geometry close to experiment (interface RMSD ~1 Å from experiment) was obtained as the top-ranking solution. Comparison of the bound[14] and unbound[13] xylanase structures indicated a significant conformational difference especially in a loop region near the active site, which is involved in complex formation [compare Fig. 2(A and B)]. In the unbound form, the active site cavity is ~1.5–2.0 Å more narrow than in the bound form. Therefore, using the unbound xylanase structure in systematic docking runs resulted in a docking geometry closest to experiment that had an RMSD of the ligand ~7 Å apart from the experimental placement (not among the top-ranking solutions). This result indicates that the differences in protein main-chain conformation between bound and unbound xylanase structures significantly affected the poor performance in this case.

In an effort to predict such main-chain conformational changes prior to docking, we explored the possibility to calculate soft collective degrees of freedom of unbound xylanase by applying an approach developed by Hinsen.[15] This approach is related to Gaussian-network models of protein motion,[15] where one basically assumes that the mobility of a different protein regions is determined by the local density (or the locally available free space). In the approach by Hinsen,[15] a quadratic energy function for the protein backbone using the experimental coordinates as reference is used. The force constant for any displacement of protein (Cα) atoms depends on the neighbor

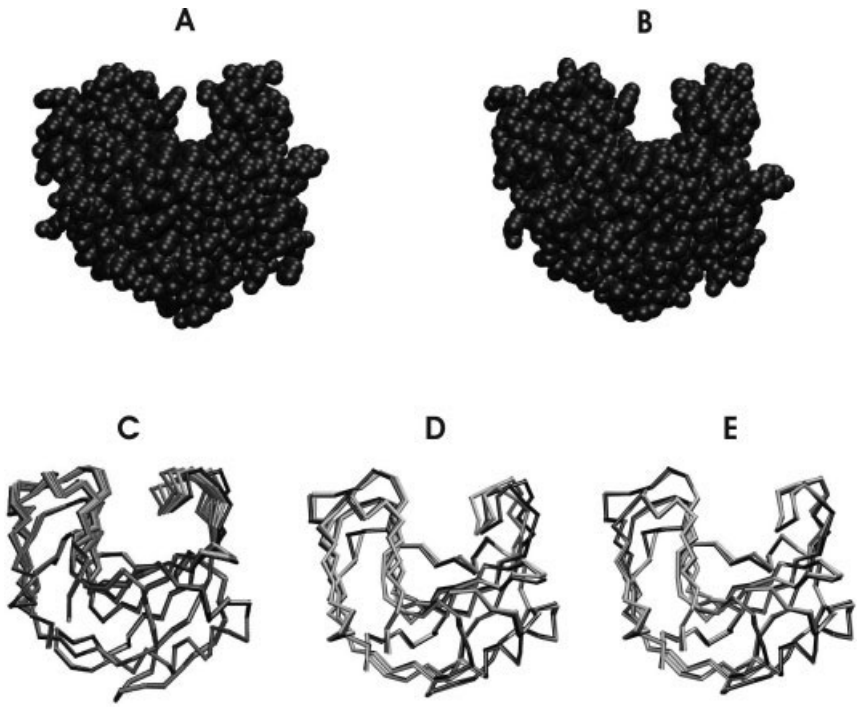

Fig. 2. Comparison of the unbound (**A**) and bound (**B**) structures of the xylanase from *A. niger* (Target 18) in van der Waals presentation. The TAXI inhibitor binding site corresponds to the cleft on the upper side of the xylanase. (**C**) Superposition of 4 structures (Cα trace; different gray levels) of the unbound xylanase structure deformed in the softest mode direction obtained by approximate normal mode analysis using the approach by Hinsen.[15] The view is approximately the same as in (A) and (B). In (*D*) the superposition of the bound (black) and unbound (light gray) xylanase structures (Cα trace; Cα-RMSD: 0.65 Å) is shown. In (**E**), the best possible approximation of the unbound form deformed in the softest mode (light gray) with respect to the bound form is superimposed on the bound xylanase structure (black, Cα-RMSD: 0.35 Å).

density of each atom. Harmonic mode analysis of this simple energy function allows us to identify possible flexible (soft) collective degrees of freedom (soft modes) of the protein.[15] Interestingly, the softest precalculated mode of the unbound form of *A. niger* xylanase shows significant overlap with the experimentally observed conformational difference relative to the bound form [illustrated in Fig. 2C–E)]. Deformation of the unbound form in a single softest mode (the mode with the smallest eigenvalue) can reduce the difference with respect to the bound form from 0.65 Å to 0.35 Å (Fig. 2). Such soft modes extracted from normal mode analysis at atomic resolution or as principal components of motion from molecular dynamics (MD) simulations have already been used successfully in ligand–receptor docking simulations.[16,17] The possibility to use such modes as additional variables during protein–protein docking in combination with the present reduced protein model will be explored in future docking simulations.

Target 19: ovine prion antibody (FAB) complex

In the case of Target 19, the antibody (FAB fragment) coordinates of the bound form were given,[8] and for the ovine prion protein a comparative model structure had to be built prior to docking based on sequence similarity to an NMR structure of a homologous bovine prion protein structure.[18] The start sites for systematic docking minimization runs of the prion model structure to the antibody FAB fragment were restricted to the approximately known ligand-binding regions of the FAB fragment. For the prion protein, no information on possible binding regions was available; hence, an exhaustive search over orientational degrees of freedom for each minimization start position was performed. Clusters of low-energy docking solutions were collected, and 10 representative complex structures were submitted. Of these, the top-ranked solution had an interface RMSD of ~5 Å compared to experiment,[8] and a fraction of 0.185 of native contacts (Table I). However, prediction number 9 showed an interface RMSD of 1.8 Å and a fraction 0.64 of native contacts, and can be counted as a good prediction. Also, prediction 4 was still acceptable, with an interface RMSD of ~2.5 Å (0.36 native contacts). After publication of the experimental complex structure, a systematic docking search using the bound partner structures and the same starting conditions was performed and yielded a solution close to experiment (RMSD of the prion ligand ~1.5 Å, interface RMSD $\leq$ 1 Å) as top-ranking solution. Comparison of the ovine prion model structure used as unbound ligand protein and the ovine protein in the experimental complex indicated significant conformational main-chain differences (>1 Å), particularly in the prion region involved in binding to the FAB fragment. These main-chain conformational differences between bound form and ovine protein model structure are likely responsible for the deviation of the predicted structures

from experiment and may contribute to the fact that the solution closest to experiment was ranked as number 9.

CONCLUSIONS

Results of protein–protein docking predictions for CAPRI Rounds 3 and 5 have been presented. The systematic docking minimization approach combined with a reduced protein model resulted in acceptable and good predictions for 3 out of 5 targets. The poor performance in 2 cases (Targets 9 and 18) can be attributed to a significant degree of main-chain conformational changes involved in protein–protein complex formation. A major current focus is to further develop the docking approach to efficiently accounting for such conformational changes during docking simulations.

ACKNOWLEDGMENTS

I thank the organizers of the CAPRI challenge for this opportunity, and the assessors for their hard work in evaluating the predictions. My thanks to all structural biologists who contributed target structures for the CAPRI experiment, and to K. Bastard and Drs. A. Barthels, C. Prevost, A. Pastor-Zacharias, and D. Roccatano for helpful discussions.

REFERENCES

1. Janin J. Welcome to CAPRI; a critical assessment of predicted interactions. Proteins 2002;47:257.
2. Zacharias M. Protein–protein docking with a reduced protein model accounting for side-chain flexibility. Protein Sci 2003;12: 1271–1282.
3. Koehl P, Delarue M. Mean-field minimization methods for biological macromolecules. Curr Opin Struct Biol 1996;6:222–226.
4. Stetefeld J, Mayer U, Timpl R, Huber R. Crystal structure of three consecutive laminin-type epidermal growth factor-like (LE) modules of laminin y1 chain harboring the nidogen binding site. J Mol Biol 1996;257:644–657.
5. 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 laminin y1 chain. EMBO J 1996;15:5154–5159.
6. Takagi J, Yang Y, Liu J, Wang J, Springer TA. Complex between nidogen and laminin fragments reveals a paradigmatic b-propeller interface. Nature 2003;424:969–974.
7. Terrak M, Kerff F, Langsetmo K, Tao T, Dominguez R. Structural basis of protein phosphatase 1 regulation. Nature 2004;429:780–784.
8. 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.
9. von 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.
10. Graille M, Zhou C-Z, Receveur-Bréchot 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.
11. Goldberg J, Huang HB, Kwon YG, Greengard P, Nairn AC, Kuriyan J. The three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1. Nature 1995; 376:745–753.
12. Cohen PTW. Protein phosphatase 1—targeted in many directions. J. Cell Sci 2002;115:241–256.
13. 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.
14. Sansen S, De Ranter CJ, Gebruers K, Brijs K, Courtin CM, Delcours JA, Rabijns A. Structural basis for inhibition of *Aspergillus niger* xylanase by *Triticum aestivum* xylanase inhibitor-I. J Biol Chem 2004;279:36022–36028.
15. Hinsen K. Analysis of domain motions by approximate normal mode calculations. Proteins 1998;33:417–429.
16. Zacharias M, Sklenar H. Harmonic modes as variables to approximately account for receptor flexibility in ligand-receptor docking simulations: application to a DNA minor groove ligand complex. J Comput Chem 1999;20:287–300.
17. Zacharias M. Rapid Protein–ligand docking including soft degrees of freedom from molecular dynamics simulations to account for protein flexibility: FK506 binding to FKBP binding protein as an example. Proteins 2004;54:759–767.
18. Lopez-Garcia F, Zahn R, Riek R, Wuthrich K. NMR structure of the bovine prion protein. Proc Natl Acad Sci USA 2000;97:8334–8339.