WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

A holistic approach to protein docking

Sanbo Qin[2,3] and Huan-Xiang Zhou[1,2,3]*

[1] Department of Physics, Florida State University, Tallahassee, Florida 32306

[2] Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306

[3] School of Computational Science, Florida State University, Tallahassee, Florida 32306

ABSTRACT

Docking of unbound protein structures into a complex has gained significant progress in recent years, but nonetheless still poses a great challenge. We have pursued a holistic approach to docking which brings together effective methods at different stages. First, protein-protein interaction sites are predicted or obtained from experimental studies in the literature. Interface prediction/experimental data are then used to guide the generation of docked poses or to rank docked poses generated from an unbiased search. Finally, selected models are refined by lengthy molecular dynamics (MD) simulations in explicit water. For CAPRI target T27, we used information on interaction sites as input to drive docking and as a filter to rank docked poses. Lead candidates were then clustered according to RMSD among them. From the clustering, 10 models were selected and subject to refinement by MD simulations. Our Model 7 is rated number one among all submissions according to L_rmsd. Six of our other submissions are rated acceptable. As scorer, eight of our submissions are rated acceptable.

Proteins 2007; 69:743–749.

Key words: protein docking; interface prediction; protein–protein interaction.

INTRODUCTION

Protein–protein interactions are at the center of protein functions. Interactions occur through the formation of complexes. The complexes may be either transient or more long-lasting and may involve two partners or multiple components. In all cases the structure of the complex is the basis for molecular understanding of protein interaction and function. Docking is a general methodology for building the structures of complexes from the structures of unbound proteins. Several docking methods[1–3] developed in the past few years have been found to be very effective, as demonstrated in the previous CAPRI evaluation meeting.[4,5] However, given the enormous challenges faced by docking, docking methods all have their limitations. We thus pursued a holistic approach. Here we report the rationale of this approach, its application to CAPRI target T27, and project its potential.

The holistic approach consists of three main steps. First, residues that likely form the interface in the native complex are predicted from the unbound structures of binding partners. We have developed an interface prediction method called cons-PPISP[6,7]; there are also a large number of alternative methods and several web servers are now available (for a review, see Zhou and Qin[8]). The interface prediction is complemented by biochemical data found in the literature. In the second step, predicted interface and biochemical data are used to guide the generation of docked poses of the binding partners. There are two complementary ways of using the prediction/data. In front-end use, the prediction/data are used to narrow the search space for docked poses.[9] In back-end use, after an unbiased search, the prediction/data are used to assist the ranking of docked poses.[10–13] While front-end use can potentially focus attention on important regions of search space, inaccuracies in the prediction/data can also mislead the search. On the other hand, in back-end use, the prediction/data can be combined with other scores, such as those based on interaction energy, and are thus more tolerant of inaccuracies. Either way, candidate poses are selected. In the final step, the candidate poses are subjected to refinement by lengthy molecular dynamics (MD) simulations in explicit solvent. We have shown that an initial approximate pose can undergo conformational rearrangements on a nanosecond time scale and approach the native complex.[14]

We had participated in CAPRI rounds 3 to 5 through a collaboration with the Bonvin group, achieving considerable success.[9] We provided predictions for interface residues by using cons-PPISP[6,7]; these predictions, along with any available biochemical information for interfaces, were then used by the Bonvin group to guide the search for docked poses by their HADDOCK program.[2] Starting with

The authors state no conflict of interest.
Grant sponsor: NIH; Grant number: GM058187.
*Correspondence to: Huan-Xiang Zhou, Department of Physics, Florida State University, Tallahassee, FL 32306.
E-mail: zhou@sb.fsu.edu
Received 29 May 2007; Revised 18 July 2007; Accepted 20 July 2007
Published online 5 September 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21752

target T27, we submitted our own models. Our Model 7 for T27 is rated number one among all submissions according to L_rmsd. Six of our other submissions are rated acceptable. As scorer, eight of our submissions are rated acceptable. For the homodimeric target T28, we and all other groups failed to submit any acceptable models, due to large inter-domain movements not predicted in the homology model for the monomer. (For the latest target, T29, our Model 1 is rated medium accuracy.)

To explore the potential of the holistic approach more thoroughly, we obtained interface predictions from the cons-PPISP web server (http://pipe.scs.fsu.edu/ppisp.html and generated 2000 docked poses by running ZDOCK[1] for each of 24 CAPRI targets (T01 through T27 except for T03, T09, and T10). Near-native poses were found for 23 of the 24 targets, but the poses with the lowest L_rmsd's were ranked among the top 100 only for seven of the targets. It appears that the search problem is largely solved by programs such as ZDOCK, but the ranking problem remains a formidable challenge. For nine of the CAPRI targets, cons-PPISP interface prediction can improve the ranking of near-native poses.

THEORETICAL METHODS

Implementation of holistic approach on T27

Target T27 is the complex formed by the SUMO-1-conugating enzyme UBC9 and the ubiquitin-conjugating enzyme E2-25K (also known as HIP2).[15] UBC9 and E2-25K are referred to ligand and receptor, respectively (based on sequence length). The unbound structures were from Protein Data Bank (PDB) entries 1a3s and 1yla, respectively.

Prediction of interface residues for UBC9 and E2-25K

Interface residues were predicted by submitting the unbound structures to the cons-PPISP web server (http://pipe.scs.fsu.edu/ppisp.html). cons-PPISP is a neural network-based predictor, taking PSI-Blast generated sequence profiles and solvent accessibilities as input.[7] Interface predictions from 17 neural network models[16] were clustered and filtered.

Biochemical data on interaction of UBC9 and E2-25K

Experimental data were found from the following papers. (1) Pichler *et al.*'s structural and biochemical studies[17] of UBC9-catalyzed SUMOylation of E2-25K, indicating that residue K14 of E2-25K is the SUMO acceptor site; (2) Reverter and Lima's structure determination[18] for a quaternary complex involving UBC9, SUMO-1, SUMOylation target RanGAP1, and E3 ligase RanBP2, placing the product isopepetide bond near residue C93 of UBC9; and (3) Yunus and Lima's structural and biochemical structures[19] identifying the active site of UBC9 at around residues C93 and D127. In addition, our study was guided initially by the structure of the complex between UBC13 and Mms2, a ubiquitin-conjugating E2 enzyme variant.[20]

Generation of docked poses

Docked poses were generated by three programs. (1) ZDOCK (version 2.3).[1] Three separate runs were performed, each generating 2000 poses. In two runs, the orientations of UBC9 and E2-25K were as downloaded; the sampling angle was set to 15° and 6°, respectively. In the third run, the orientations of UBC8 and E2-25K were fitted to those of UBC15 and Mms2, respectively, in the latter's complex (PDB entry 2gmi). For later reference, these runs are designated zd1, zd2, and zd3. (2) ClusPro.[21] 10 poses were obtained from the web server (http://nrc.bu.edu/cluster/) using default options. One of these poses passed the filtering by biochemical data (see below) and is designated "clu1" for later reference. (3) HADDOCK (version 1.3).[2] Two separate runs were performed, each generating 200 poses. In one run (designated hd1), the active residues were chosen as K14 of E2-25K and C93 and D127 of UBC9; residues V16-F36 of E2-25K and residues E122-Q130 of UBC9, close to the presumed interface, were treated as fully flexible. In the second run (designated hd2), the active residues were chosen as F13 and K14 of E2-25K and Y87, C93, and D127 of UBC9; backbones were fixed. For both ha1 and ha2, passive residues were identified from the clu1 pose, and included G-1, S0, L17, K18, S23-N25, K28, D30-V32, E34, and R55 of E2-25K and K65, D67, S70, K74, P88, S89, E99-K101, I125, Q126, P128, A131, E132, T135, and Q139 of UBC9. hd1 also included F13 of E2-25K and Y87 of UBC9 as passive residues.

Screening by biochemical data

All the initial poses were screened for the contacts (defined as within 6 Å) between K14 of E2-25K and C93 and D127 of UBC9. A total of 38 poses passed this screening process. Of these, zd1, zd2, and zd3 accounted for 9, 10, and 8 poses, respectively; hd1 and hd2 accounted for eight and two poses, respectively; the remaining pose is clu1.

RMSD clustering and model selection

The remaining 38 poses were clustered using the R program (http://www.r-project.org/) according to C_α RMSD between poses. The final 10 models were selected manually. The assignments of Model 1 to through Model 10 were: hd2_11, zd2_117, hd1_5, hd1_12, hd1_21, zd2_271, zd2_570, zd1_714, zd3_1024, and zd1_612. The

DOI 10.1002/prot

numbers after the underscore are the rankings of these poses in the original ZDOCK or HADDOCK runs.

Refinement by MD simulations

The 10 models were further subjected to refinement by MD simulations in explicit solvent. Given the time limitation in the CAPRI exercise, lengthy simulations were not possible. MD simulations were therefore limited to only 40 ps, run with the AMBER program.

After the MD simulations, some side chains were further adjusted manually. These included R8, K10, R11, F13, and K14 of E2-25K and D100 of UBC9. The aims were to allow for better ion pair interactions between R8, K10, and K11 of E2-25K and D100 of UBC9, better aromatic-aromatic interaction between F13 of E2-25K and Y87 of UBC9, and closer contact between K14 of E2-25K and C93 of UBC9. For Model 7, none of these interactions seemed possible through side chain adjustments, and so no adjustments were made. Finally, the models were subjected to 5000 steps of energy minimization in AMBER. The models were checked for clashes in the interface region by tleap in AMBER; no significant clashes were detected.

Scoring

Scoring of 1489 downloaded models for T27 started with screening by the three key residues identified from biochemical data (see earlier). The 293 models that passed the screening process were fitted to our submitted Model 1; the 50 models with the lowest C_α RMSDs were retained. These were then clustered according to C_α RMSD between themselves. The final 10 models were selected manually, with consideration for diversity among different predictor groups (which could be easily identified from the different file formats used by the predictor groups).

Implementation of holistic approach on T28

Target T28 is a homodimer of NEDD4-like E3 ubiquitin-protein ligase.[22] The monomer structure was generated by homology modeling with SWISS-MODEL (http://swissmodel.expasy.org//SWISS-MODEL.html)[23]; four templates (with PDB entries 1zvd_A, 1d5f_A, 1d5f_C, and 1c4z_C)[24,25] were used. Thousand five hundred docked poses were obtained by running M-ZDOCK.[26] These poses were clustered according to the distribution of interface residues along the sequence (due to symmetry only one subunit was considered). Interface of a pose was defined as consisting of residues making <5 Å contacts with the partner protein. The sequence was divided into 20-residue blocks, which were labeled as 1 if at least three interface residues were present and 0 otherwise. Each pose was thus translated into a binary sequence. Poses with the same binary sequence were collected into one cluster; a total of 257 clusters were obtained. For each cluster, the top-ranked pose according to M-ZDOCK was inspected by eye and 11 clusters were selected. Criteria for selection included the following: (1) the active site and the E2 binding site have to be exposed; and (2) a sufficient amount of surface area has to be buried in the interface. Out of all the 58 poses within the 11 selected clusters, 8 were retained to cover a variety of potential sites around one monomer. For each retained pose, all poses among the original 1500 that had L_rmsd within 20 Å were further considered, and one or two representative poses were chosen according to electrostatic interaction energy and desolvation energy calculated by FastContact[27] and buried surface area calculated by NACCESS.[28] The final 10 chosen poses had M-ZDOCK rankings of 800, 70, 333, 26, 23, 88, 86, 908, 800, and 624 (ordered according to submitted model number). Only Models 1, 3, 4, and 5 were refined by 1 ns of MD simulations; the other models were simply energy minimized (note that pose 800 resulted in two models: Model 1 after MD refinement and Model 9 with energy minimization).

ZDOCK runs on T01-T27

ZDOCK runs were performed on CAPRI targets T01 through T27, except for T09 and T10. For T09, the large conformational changes upon complex formation preclude a rigid-body docking program like ZDOCK from generating near-native poses; this target was not considered further. T10 is a homo-trimer, which is not suitable for running ZDOCK. In addition, the ZDOCK program crashed on running T03. For each of the remaining 24 targets, 2000 poses were collected with a sampling angle of 15° (corresponding to the zd1 run described earlier). The 24 targets are listed in Table I.

Among the 2000 poses for each target, near-native poses were identified as those with L_rmsd <10 Å from the native complex. In calculating L_rmsd, only ligand C_α atoms within 10 Å of the receptor in the native complex were used. The performance of ZDOCK in ranking near-native poses was assessed by counting N_{nn}, the number of near-native poses, among the first N_{cut} poses according to ZDOCK scores.

Ranking by cons-PPISP interface prediction

Interface predictions for the 24 CAPRI targets, using the unbound structures of the binding partners, were obtained from the cons-PPISP web server (http://pipe.scs.fsu.edu/ppisp.html). Only the unfiltered clusters of positively predicted residues were used for ranking the poses generated by ZDOCK.

To rank poses for a target, the predicted interface residues for both binding partners were combined. The combined set was treated as a benchmark. The fraction of benchmark residues found in the interface of a pose was taken as the score of the pose.

Table I
ZDOCK Ranking of 2000 Poses for Each of 24 CAPRI Targets

Target	Complex[a]		Binding partners[b]		Lowest L_rmsd (Å)	Rank	Total N_{nn}[c]
T01	HPrK:HPr	1kkl	1jb1	1sph	9.0	1086	3
T02	Rotavirus VP6:FAB		1qhd	bound	8.8	1866	1
T04	α-amylase:Ab AM-D10	1kxv	1pif	bound	6.0	180	7
T05	α-amylase:Ab AM-07	1kxt	1pif	bound	4.6	41	7
T06	α-amylase:Ab AM-D9	1kxq	1pif	bound	2.8	6	37
T07	TCRβ:SpeA	110x	1bec	1blz	15.2	551	0
T08	Nidogen G3:laminin	1npe	bound	1klo	6.6	1895	5
T11	Cohesin:dockerin	1ohz	1anu	1daq[d]	6.2	315	11
T12	Cohesin:dockerin	1ohz	1anu	bound	1.6	32	34
T13	SAG1:FAB	1ynt	1kzq	bound	1.4	8	11
T14	phosphatase-1:MYT1	1s70	1fjm[d]	bound	5.7	443	4
T15	Colicin D:ImmD	1v74	bound	bound	1.7	1	31
T16	*A. nidulans* xylanase:XIP-1	1ta3	1bg4[d]	1om0	2.3	37	15
T17	*P. funiculosum* xylanase:XIP-1	1te1	1ukr[d]	1om0	5.6	919	6
T18	*A. niger* xylanase:TAXI	1t6g	1ukr	bound	3.8	1176	7
T19	Ovine prion:FAB	1tpx	1dwy[d]	bound	1.1	351	11
T20	eRF1:HemK	2b3t	1gqe[d]	1t43	9.0	1070	1
T21	Orc1:Sir1	1zhi	1m4z	1z1a	2.9	1192	21
T22	U5-15K:U5-52K	1syx	1qgv	1gyf	3.9	14	35
T23	GBP1 dimer	2b8w	1f5n		9.5	594	1
T24	Arf1:ARHGap10	2j59	1o3y	1btn[d]	7.1	1646	5
T25	Arf1:ARHGap10	2j59	1o3y	bound	9.9	1616	1
T26	TolB:Pal	2hqs	1c5k	1oap	4.5	350	9
T27[e]	E2-25K:UBC9	2o25	1yla	1a3s	5.6	1013	9

[a]Names of binding partners (separated by ":") in each complex is followed by the PDB entry of the complex.
[b]PDB entries used for docking; "bound" means the corresponding protein is taken from the native complex.
[c]Number of near-native poses out of the 2000 poses generated by ZDOCK for each target.
[d]Templates for building homology models.
[e]ZDOCK was evaluated against the T27.2 solution.

RESULTS AND DISCUSSION

Interface residues of T27

Interface predictions for UBC9 and E2-25K were obtained from the cons-PPISP web server (http://pipe.scs.fsu.edu/ppisp.html). For UBC9, two clusters of residues were predicted to be in interfaces. As Figure 1(a) shows, one cluster largely overlaps with the binding site for the E3 ligase RanBP2 (as found in PDB entry 1z5s).[18] The second cluster largely overlaps with the binding site for SUMO-1 found in 1z5s. The crystal structure for the complex between UBC9 and E2-25K (PDB entry 2o25)[15] defines two possible solutions for T27, designated T27.1 and T27.2. The binding site for E2-25K on UBC9 in T27.2 overlaps almost perfectly with the binding site for RanGAP1 (as found in 1z5s). Unfortunately, besides Q126 (as part of the second cluster), no other UBC9 residues forming the RanGAP1/E2-25K(T27.2) binding site were predicted by cons-PPISP.

cons-PPISP predicted three clusters of residues for E2-25K. As shown in Figure 1(b), the largest cluster centers around the active site cysteine (residue C92), and seems to define a binding site for ubiquitin. The other two clusters overlap with the binding sites for UBC9 found in T27.1 and T27.2, respectively.

Like RanGAP1, E2-25K is a target for SUMOylation by UBC9. We thus assumed at the very beginning that E2-25K occupies a binding site on UBC9 close to the binding site for RanGAP1. Since cons-PPISP did not predict the binding site for RanGAP1, we decided to abandon cons-PPISP predictions all together for docking UBC9 with E2-25K. Instead, we sought experimental data from the literature that would indicate important residues that are involved in the interaction between UBC9 and E2-25K.

On the UBC9 side, based on studies of Lima and coworkers,[18,19,29] we selected C93 and D127. On the E2-25K side, based on mutational data of Pichler *et al.*[17] indicating that residue K14 of E2-25K is the SUMO acceptor site, we selected that residue. These three residues were used for screening docked poses [Fig. 2(a); see Theoretical Methods].

Submitted models for T27

Our Model 1 reflected our working hypothesis for the structure of target T27 and is representative of most of our 10 submitted models [Fig. 2(a)]. In this model, K14 of E2-25K is close to C93 and D127 of UBC9 (at 3.1 and 5.1 Å, respectively). In addition, R8, K10, and F13 of E2-

DOI 10.1002/prot

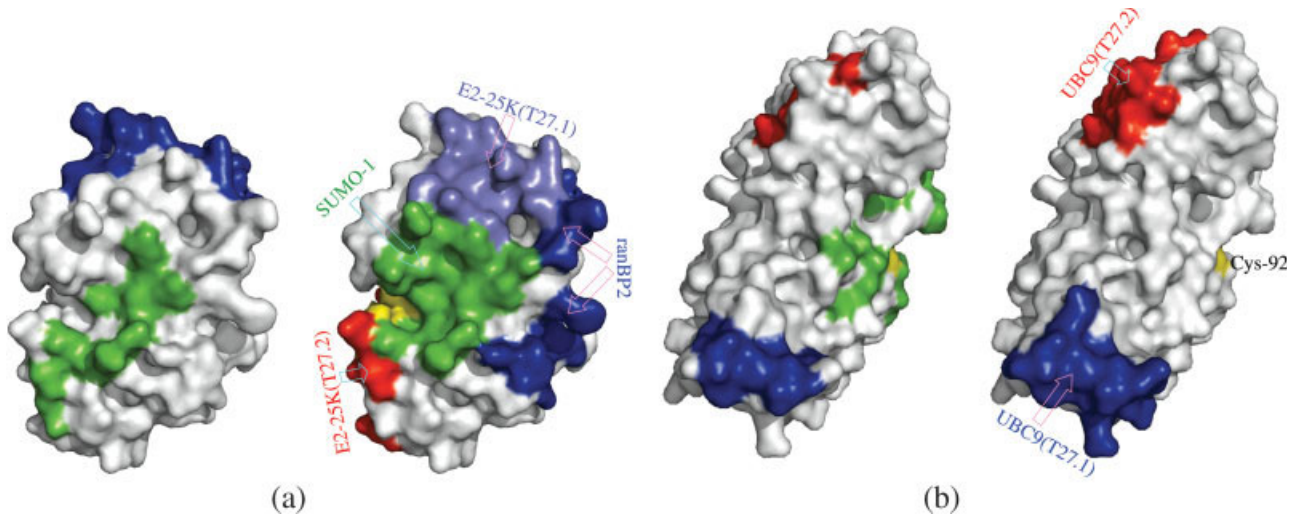

Figure 1

*Comparison of predicted interface residues (left panels) and binding sites found in X-ray structures (right panels) for (**a**) UBC9 and (**b**) E2-25K. In the right panel of (a), the RanBP2, SUMO-1, and RanGAP1 binding sites on UBC9, found in PDB 1z5s, are shown in blue, green, and red, respectively. The RanGAP1 binding site is identical to the E2-25K binding site found in solution T27.2. A subset of the RanBP2 binding site, shown in a lighter shade of blue, forms the E2-25K binding site in solution T27.1. Residues that are both in the SUMO-1 binding site and the RanGAP1/E2-25K(T27.2) binding site are shown in yellow. In (b), the active-site cysteine (C92) is shown in both panels in yellow. In the right panel, the UBC9 binding sites found in T27.1 and T27.2 are shown in blue and red, respectively.*

25K are positioned in the interface, which was motivated by Pichler *et al.*'s work[17] showing that mutations of these residues reduced the SUMOylation of E2-25K by UBC9. Model 1 is rated acceptable, as are Models 2, 4, 6, 8, and 9.

Our Model 7 is the best of all submissions, with an L_rmsd of 6.18 Å [Fig. 2(b)]. It is one of the only two submissions in the medium accuracy category. Compared with Model 1, E2-25K in Model 7 is moved along the N-terminal helix (residues 2–19) for about one turn of helix, positioning K10 at close contact (at 3.8 Å) with C93 of UBC9. Model 7 is marked by a surprisingly high value of f_{nat} (at 0.73), the fraction of native contacts that are found in the interface of the model. This f_{nat} is even higher than what is obtained, 0.63, if the unbound structures are superimposed to the native complex.

Solution T27.2 (but not T27.1) seems to be in general agreement with available biochemical data, though not in complete accord. In particular, in T27.2, K10 of E2-25K is placed near C93 of UBC9, but the mutational data of Pichler *et al.* showed that residue K14 of E2-25K is the

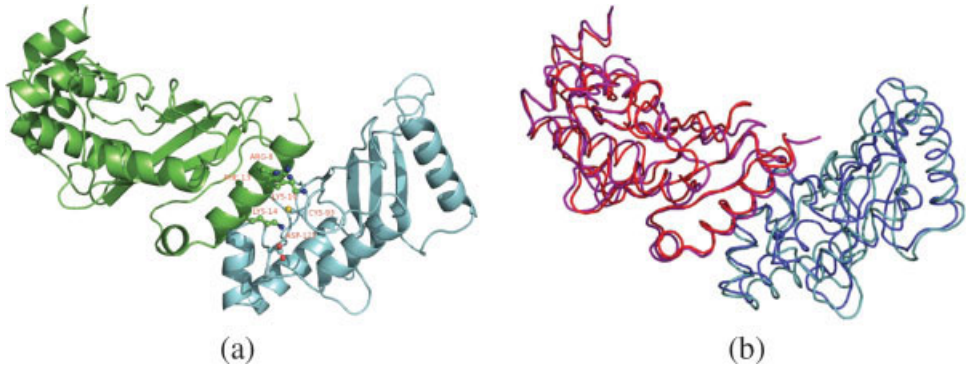

Figure 2

*Structures of (**a**) Model 1 and (**b**) Model 7. In (a), residues that were assumed to be key in the protein–protein interaction are shown as ball-and-stick. In (b), E2-25K and UBC9 are shown in red and blue for the model and in purple and cyan for the T27.2 solution. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]*

SUMO acceptor site. It is possible that the difference in SUMO acceptor site is due to the difference in experimental conditions. K10, unlike K14, is part of a consensus motif (ΦKXE) found in most SUMO targets, and Pichler *et al.* in the same study found that K10 is the SUMOylation site in an unstructured peptide corresponding to the N-terminal helix of E2-25K.

In the scoring exercise for T27, we used our submitted Model 1 as the benchmark. Eight of our 10 submissions are rated acceptable.

Submitted models for T28

Human NEDD4-like E3 ubiquitin-protein ligase belongs to the HECT class of E3s, which share a conserved C-terminal catalytic domain (known as the HECT domain). The HECT domain can be divided into an N-lobe and a C-lobe; the N-lobe can be further divided into a large subdomain and a small subdomain.[24,25] The active site is located in the interface between the N-lobe large subdomain and the C-lobe. The N-lobe small subdomain defines the E2 binding site. The relative position and orientation between the N- and C-lobes are known to be significantly different among different HECT domains.[25] When the homology model for the NEDD4-like HECT domain is compared actual the crystal structure (PDB entry 2oni)[22] by superimposing the N-lobe large subdomain, the C-lobes in the two structures are found to be rotated by ~90° and translated by ~20 Å between each other.

The target structure for T28 is the homodimer of the NEDD4-like HECT domain. Because of the large-scale inter-lobe motion, we and all other groups failed to submit any acceptable models. It is still not known whether dimer is the biologically relevant oligomeric state for the NEDD4-like HECT domain.[22] In lieu of any positive information on the dimer interface, our model selection used the "negative" assumption that the active site and the E2 binding site should not be buried in the dimer interface (see Theoretical Methods). Our best submission for T28 is Model 4. Although no native contacts are present in this model, the fraction of native interface residues in the interface of Model 4 is a respectable 0.36. The L_rmsd and I_rmsdBB of Model 4 are 27.8 and 7.2 Å, respectively. In comparison, the best results for these three parameters among all groups are 0.41, 21.1, and 4.3 Å, respectively. Without specifically modeling the substantial inter-lobe movement, it would be impossible to produce an acceptable model for the dimer. Indeed, superimposing the unbound monomers onto the dimer structure gives severe clashes.

Is the search problem largely solved?

The lowest L_rmsd's of the ZDOCK-generated poses for 24 CAPRI targets are listed in Table I. Except for T07, these values are all <10 Å, indicating that ZDOCK is capable of generating near-native poses in almost all cases. The problem is that these near-native poses are not ranked well (see Table I). For only seven targets, ZDOCK ranked the poses with the lowest L_rmsd's among the top 100 (out of a total of 2000). For T27, 6 of the 10 models that we submitted were initially generated by ZDOCK; none of the 6 was ranked in the top 100 by the original ZDOCK score. Model 7, which is the best among all submissions, was ranked 570th by ZDOCK.

The number of near-native (N_{nn}) poses ranked among the first N_{cut} poses by ZDOCK provides a measure on the quality of the ranking. A high N_{nn} increases the chances that near-native poses will be kept in the final selection of models. Among the first 100 poses, 9, 5, and 5 CAPRI targets, respectively, have 0, 1, and 2 near-native poses; the remaining five targets have 3, 4, 5, 6, and 8 near-native poses, respectively.

How much can predicted interface residues help ranking?

The con-PPISP-predicted interface residues can also be used to rank the docked poses and generate N_{nn} versus N_{cut} curves (see Theoretical Methods). For nine of the 24 targets, the curves moved upward compared with the ZDOCK counterparts (not shown), meaning that more near-native poses were given high rankings. These are T06, T11, T12, T13, T15, T16, T19, T22, and T26. In each of the successful cases, the predicted binding sites for both partner proteins partially overlap with the respective binding sites found in the native complex. Simulated data confirm that interface prediction can achieve improvement in the ranking of near-native poses when predictions are partially correct for the binding sites on both partner proteins.[8]

Given that current interface prediction can improve ranking for only ~40% of targets, it become crucial to decide whether to trust predicted interface residue in model selection. The decision is easier when there are biochemical data available from the literature, like what happened for T27.

Another important issue is whether the near-native poses can be refined by lengthy MD simulations. Can MD simulations move near-native poses toward the native complex? If so, how close to the native complex does the starting pose has to be? A systematic study to address these questions is underway.

Significant progresses are being made both on interface prediction[8] and on force-field improvement for better modeling of protein motions.[30] Along with these developments, it is hopeful that the holistic approach described here for protein docking will become more and more successful in building structural models for protein complexes.

DOI 10.1002/prot

REFERENCES

1. Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. Proteins 2003;52:80–87.
2. 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.
3. Gray JJ, Moughan SE, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. J Mol Biol 2003;331:281–299.
4. Janin J. Sailing the route from Gaeta, Italy, to CAPRI. Proteins 2005;60:149.
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. Zhou H-X, Shan Y. Prediction of protein interaction sites from sequence profile and residue neighbor list. Proteins 2001;44:336–343.
7. Chen H, Zhou H-X. Prediction of interface residues in protein-protein complexes by a consensus neural network method: test against NMR data. Proteins 2005;61:21–35.
8. Zhou H-X, Qin SB. Interaction-site prediction for protein complexes: a critical assessment. Bioinformatics, in press.
9. van Dijk ADJ, de Vries SJ, Dominguez C, Chen H, Zhou H-X, Bonvin AMJJ. Data-driven docking: HADDOCK's adventures in CAPRI. Proteins 2005;60:232–238.
10. Tress M, de Juan D, Grana O, Gomez MJ, Gomez-Puertas P, Gonzalez JM, Lopez G, Valencia A. Scoring docking models with evolutionary information. Proteins 2005;60:275–280.
11. Heuser P, Bau D, Benkert P, Schomburg D. Refinement of unbound protein docking studies using biological knowledge. Proteins 2005;61:1059–1067.
12. Chelliah V, Blundell TL, Fernandez-Recio J. Efficient restraints for protein-protein docking by comparison of observed amino acid substitution patterns with those predicted from local environment. J Mol Biol 2006;357:1669–1682.
13. Tjong H, Qin SB, Zhou H-X. PI^2PE: protein interface/interior prediction engine. Nucleic Acids Res 2007;35:W357–W362.
14. Huang X, Dong F, Zhou H-X. Electrostatic recognition and induced fit in the κ-PVIIA toxin binding to *Shaker* potassium channel. J Am Chem Soc 2005;127:6836–6849.
15. Walker JR, Avvakumov GV, Xue S, Newman EM, Mackenzie F, Weigelt J, Sundstrom M, Arrowsmith CH, Edwards AM, Bochkarev A, Dhe-Paganon S. A novel and unexpected complex between the SUMO-1-conjugating enzyme UBC9 and the ubiquitin-conjugating enzyme E2-25 kDa, to be published.
16. Qin SB, Zhou H-X. Meta-PPISP: a meta web server for protein-protein interaction site prediction, in press.
17. Pichler A, Knipscheer P, Oberhofer E, van Dijk WJ, Korner R, Olsen JV, Jentsch S, Melchior F, Sixma TK. SUMO modification of the ubiquitin-conjugating enzyme E2-25K. Nat Struct Mol Biol 2005;12:264–269.
18. Reverter D, Lima CD. Insights into E3 ligase activity revealed by a SUMO-RanGAP1-Ubc9-Nup358 complex. Nature 2005;435:687–692.
19. Yunus AA, Lima CD. Lysine activation and functional analysis of E2-mediated conjugation in the SUMO pathway. Nat Struct Mol Biol 2006;13:491–499.
20. Eddins MJ, Carlile CM, Gomez KM, Pickart CM, Wolberger C. Mms2-Ubc13 covalently bound to ubiquitin reveals the structural basis of linkage-specific polyubiquitin chain formation. Nat Struct Mol Biol 2006;13:915–920.
21. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 2004;20:45–50.
22. Walker JR, Avvakumov GV, Xue S, Butler-Cole C, Weigelt J, Sundstrom M, Arrowsmith CH, Edwards AM, Bochkarev A, Dhe-Paganon S. NEDD4-like E3 ubiquitin-protein ligase, to be published.
23. Chwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 2003;31:3381–3385.
24. Huang L, Kinnucan E, Wang G, Beaudenon S, Howley PM, Huibregtse JM, Pavletich NP. Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. Science 1999;286:1321–1326.
25. Ogunjimi AA, Briant DJ, Pece-Barbara N, Le Roy C, Di Guglielmo GM, Kavsak P, Rasmussen RK, Seet BT, Sicheri F, Wrana JL. Regulation of Smurf2 ubiquitin ligase activity by anchoring the E2 to the HECT domain. Mol Cell 2005;19:297–308.
26. Pierce B, Weng Z. M-ZDOCK: a grid-based approach for Cn symmetric multimer docking. Bioinformatics 2005;21:1472–1476.
27. Camacho CJ, Zhang C. FastContact: rapid estimate of contact and binding free energies. Bioinformatics 2005;21:2534–2536.
28. Hubbard SJ, Thornton JM. NACCESS. Department of Biochemistry and Molecular Biology, University College London; 1993.
29. Knipscheer P, Sixma TK. Divide and conquer: the E2 active site. Nat Struct Mol Biol 2006;13:474–476.
30. Yang W, Nymeyer H, Zhou H-X, Berg B, Brüschweiler R. Quantitative computer simulations of biomolecules: a snapshot. J Comput Chem, in press.