Published in final edited form as:
Proteins. 2010 November 15; 78(15): 3166–3173. doi:10.1002/prot.22772.

Selection of Near-Native Poses in CAPRI Rounds 13-19

Sanbo Qin and **Huan-Xiang Zhou***
Department of Physics and Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306, USA

Abstract

In CAPRI rounds 13-19, we submitted models that are of acceptable or higher quality for 6 of the total of 13 targets. This success builds on our record in previous CAPRI rounds. The docking problem can be divided into two steps. In the first, translational/rotational and conformational space is searched to generate a pool of docked poses; the success of this search step is measured by whether near-native poses are included in the pool. In the second step, the pool is selected for near-native poses. In our previous assessment of CAPRI results, we suggested that the search problem is largely solved; a remaining problem is to select near-native poses. Our work in these new rounds of CAPRI was guided by this assessment. To solve the selection problem, we used an assortment of criteria on the interfaces of candidate poses. In one extreme, represented by T29, with very little known interface information, our criterion for top models was based on interface prediction. Poses in which the predicted interface residues occurred in interfaces were selected. Our model 1 for T29 was of medium quality. In the other extreme, represented by T40, with reliably known interface information, our selection was solely based on such information. Nine of the 10 models submitted for T40 were of high (3 models), medium (4 models), and acceptable (2 models) quality. Our strategy of mixing predicted and known interface information appears to be widely applicable for the selection of near-native poses.

Keywords

protein docking; interface prediction; protein-protein interaction

Introduction

The docking of two unbound proteins into their native complex involves the search in the 6-dimensional space of relative translation and rotation and the high dimensional space of internal conformations. For proteins that do not undergo large-scale conformational changes upon binding, the global search in the 6-dimensional translational/rotational space is largely uncoupled from the search in conformational space. Due to advances in global docking methods,[1-3] the search problem in the translational/rotational space has largely been solved.[4] After a global search, only poses with good geometric complementarity are retained. A successful global search means that at least one near-native pose is among the thousands poses retained. In an ideal case, near-native poses are highly enriched and form one of the largest clusters in the retained poses. Selecting near-native poses is then a simple task. However, in order to speed up the searching process, global docking methods have to use relatively simple scoring functions. These scoring functions usually do not rank near-native poses well. Then a further selection step is necessary to pick out near-native poses from all the poses retained in a global search.

*Correspondence information: hzhou4@fsu.edu; phone: (850) 645-1336; fax: (850) 644-7244.

More sophisticated scoring functions potentially can accomplish the task of selecting near-native poses. Although this approach has shown promises in some cases,[5-8] at present it is still short of providing a general solution. One problem is that these scoring functions are very sensitive to the rotamers and interactions of interfacial sidechains.

On the other hand, biochemical and structural information is often available and can be valuable for selecting near-native poses. For example, in an enzyme-substrate complex (in which the substrate is a protein), knowledge of residues involved in catalysis allows one to define the center of the interface between the two subunits. It is likely that a majority of the poses retained in a global search do not have the specific residues at the centers of their interfaces and can thus be eliminated. In CAPRI rounds 13-19, we took advantage of biochemical data for targets T32 and T40 and structural data for targets T32, T33/T34, T37, T40, and T41. Such information is useful even when it is only available for one of the two subunits, as in targets T35/T36. However, the use of literature data also carries risks. For targets T30 and T38/T39, we apparently incorrectly interpreted the available information, and as a result modeled the interfaces in wrong locations.

For our models submitted for CAPRI rounds 13-19, we used biochemical and structural information gathered from the literature as much as possible. We now present this use, in the hope that our experiences and lessons can be instructive for others.

Theoretical Methods

Interface Prediction

Our group pioneered protein-protein interface prediction at the residue level.[9] The prediction results have been shown to be very useful for the selection of near-native poses.[4,10] For CAPRI rounds 13-19, we used our meta-PPISP server[11] for interface prediction. This server is built on three individual methods: cons-PPISP,[9,12] PINUP,[13] and Promate.[14] Residues predicted with high scores by meta-PPISP were taken as candidates for interface residues. Interface prediction provided valuable information for our correct model selection for target T29, for which we could not find any useful information from the literature.

Global Search

We used ZDOCK2.3[1] to generate 2000 poses for all targets but T34 and T42. For the homodimeric target T42, M-ZDOCK[15] was used to generate poses. Most ZDOCK runs used a 15° sampling angle; occasionally a 6° sampling angle was used.

Local Enrichment

RosettaDock[6] was used as a local enrichment method to help refine poses for target T37. The program was started on poses selected from a ZDOCK run. The search was limited to a small range with the RosettaDock local perturbation protocol, specified by the option "-dock_pert 3 8 8", with default side chain repacking.

Scoring

In addition to biochemical and structural information, our selection of near-native poses was helped with scoring functions. Generally, we did not use the scoring functions for direct selection. Rather, we used them for negative selection. That is, poses with poor scores were eliminated.

One scoring function used was FastContact,[16] which gives rapid estimates of contact and binding free energies. To use it, hydrogens were added to the heavy atoms of each pose by a script from Multiscale Modeling Tools (MMTSB[17]) in CHARMM. The direct electrostatic

energy, desolvation free energy, and buried surface area were separately assessed in deciding whether or not to eliminate a pose. FastContact was used on targets T29, T32, T35, and T36.

Another scoring function used was ZRANK, which was designed to improving the [5] ranking of ZDOCK. This program was used on all but targets T29, T33, and T34.

RMSD-Based Clustering

When the number of retained poses was more than a few, a clustering method was used to reduce redundancy between poses. With tens of poses, the hierarchical clustering method in R program (http://www.r-project.org/) was used to find the separation and distribution of the poses. When there were more retained poses, a clustering program from HADDOCK[18] package was used. The clustering was based on either L_rmsd or I_rmsd. Typically a representative was selected from each cluster of interest.

Energy Minimization

To remove clashes, all submitted models were subjected to energy minimization by the AMBER program. Minimization consisted of 50 steepest-descent steps and 450 conjugate-gradient steps.

Results and Discussion

Our performance in CAPRI rounds 13-19 is summarized in Table 1. Overall we submitted models with acceptable or better quality for six of the 13 targets. Below we present details of our model selection.

Target 29: Trm82 and Trm8

Target T29 is the complex of Trm82 and Trm8.[19] Trm8 is an S-adenosyl-L-methionine (SAM) dependent tRNA (m^7G46) methyltransferase. Sequence analyses suggested the involvement of carboxylate residues (E69, E94, and D121 in *E. coli*) in the SAM binding site[20] and the sequence motif GxGxG for tRNA binding.[21] These residues of Trm8 should not be blocked by Trm82 in their complex, since the catalytic assays of Trm8 were carried out withTrm82 bound.[21,22] Truncation of the first 39 residues did not affect the catalytic activity of Trm8,[22] suggesting that the N-terminal is not in the interface between Trm8 and Trm82. No direct information was found from the literature on the interface between Trm8 and Trm82. We therefore turned to interface prediction.

Figure 1(A) displays the interface-prediction results (by the PINUP method) for Trm82 on the X-ray structure of the complex. The predicted interface residues are indeed mostly located in the interface with Trm8. The 2000 poses generated by ZDOCK were also enriched in the Trm8 binding site. Figure 1(B) displays this enrichment according to the frequencies of individual Trm82 residues appearing in the interfaces (defined as having at least one interfacial constant < 5 Å) of the 2000 poses. The convergence of the two approaches gave us confidence in choosing this site as the Trm8 binding site. Our model 1 turns out to be of medium quality [comparing Figure 1(A) and 1(B)].

It is of interest to note that model 1 was ranked 1648th by ZDOCK. Part of the reason this model was promoted to the top was that the electrostatic free energy calculated by FastContact was very favorable.

Target 30: Rnd1 and RBD-plexinB1 Dimer

Target T30 is a complex between the Ras-binding domain (RDB) of plexinB1 and Rnd1.[23] RDB-plexinB1 complexes with Rnd1 as a homodimer. There is evidence indicating that the RDB- plexinB1 binding site on Rnd1 is close to the GTP binding loop,[24] a common feature of the Ras family of proteins in their interactions with effectors. A genome-wide homology modeling effort has been made to build Ras-effector complexes, based on the large number of structures for such complexes that have been solved.[25] For target T30, we took a similar approach. Our submitted models were built on the structures of three Ras-effector complexes (PDB entries 2C5L, 1HE8, and 1LFD) in order to cover the structural variations. Two additional models from ZDOCK runs similar to these homology models were also among the models submitted.

In many previous structures of Ras-effector complexes, β-strands from the two sides form a cross-interface β-sheet. This feature is absent in the X-ray structure of the complex between Rnd1 and RDB-plexinB1. As a result, our submitted models are incorrect; all of them have I_rmsd > 10 Å from the X-ray structure. It is interesting that this target was difficult for the community as a whole, with just two of all the models submitted acceptable.

The variation in interface may be common in complexes formed by proteins from a large family with their interaction partners. This variation may be an evolutionary strategy for achieving specificity. Unfortunately this evolutionary strategy presents a roadblock for building protein complexes by homology modeling.

Target 32: Savinase and BASI

Target T32 is the complex of savinase and barley α-amylase/subtilisin inhibitor (BASI).[26] Savinase is a subtilisin-like serine proteinase. A serine, an aspartate, and a histidine form the catalytic triad in which the necleophilic serine O_γ atom attacks the carbonyl C atom of the peptide bond of the substrate to produce a tetrahedral covalent intermediate. According to the sequence of subtilisin BPN·, Asp32, His64, and Ser221 are the catalytic residues.[27] BASI is a bifunctional protein, inhibiting both the endogenous protein, α-amylase isozyme 2, and serine proteases of the subtilisin family from pathogens and pests.[28] The savinase binding kinetics of BASI and several mutants, including Y87A, T89A, and E95Q, have been studied.[29] The K_i for Y87A and T89A are significantly changed, but the K_i for E95Q is nearly the same as the wild-type protein. So Y87 and T89 are likely positioned in the interface between savinase and BASI, but E95 is probably not.

Residue Y87 is located in a loop of BASI. Our selection of poses hence was focused on models in which this loop was positioned in the interface with savinase. We noticed that an inhibitor of subtilisin BPN· also uses a loop in interacting with the enzyme.[30] However, other than the use of a loop in interacting with their respective enzymes, there is no structural similarity between BASI and the subtilisin inhibitor [see Figure 1(C); note that the two loops even trace opposite directions]. In any event, in selecting the models for target T32, we favored poses with the Y87-contiaining loop directed toward the enzyme in a similar fashion as the corresponding loop of the subtilisin inhibitor in the latter's complex with subtilisin BPN·. This turns out to be not a bad choice. However, the sidechain of Y87 adopts very different rotamers before and after binding the enzyme (χ_1 differing 150°). Consequently our best model is only of acceptable quality [see Figure 1(D)].

Targets 33 and 34: Methyltransferase and RNA

Targets T33 and T34 are the complex between a methyltransferase and an RNA transcript. The enzyme methylates a base in rRNA, and is found in many Gram-positive bacteria.[31,32]

The two targets differ in the structure provided for the RNA: a homology model for T33 and the bound structure for T34.

The T33/T34 enzyme is closely related to another methyltransferase of known structure; they share approximately 30% sequence identity and clearly have a common evolutionary origin. The known structure reveals the binding site for the substrate S-adenosyl-L-methionine (SAM). The residues that take part in SAM binding are either identical or of similar types to those in the T33/T34 enzyme; hence the SAM binding site is conserved. Because each enzyme transfers a methyl from SAM to a base in rRNA, the latter should be close to the bound SAM. In addition, a zinc binding domain is believed to be important in rRNA recognition. These features provided important information for us in docking the RNA substrate onto the enzyme.

With the interface so well defined, one of our submitted models for T34 is of medium quality and another three are of acceptable quality. For target T33, the structure for the RNA was a homology model, which we knew had to be very different from the bound structure. We tried to build the bound structure by hand and then dock it to the enzyme. However, that task proves to be too challenging. Indeed T33 is a challenge for the whole community: no acceptable model was submitted at all.

Targets 35 and 36: Intramolecular Complex of Xyn10B

Targets T35 and T36 are the complexes between two domains of Xyn10B: the catalytic domain and the carbohydrate-binding module.[33] Xyn10B is an endo-β-1,4-xylanase from *C. thermocellum*. The hydrolysis of xylan by xylanases, mostly in glycoside hydrolase (GH) families 10 and 11, is one of the key reactions in microbial degradation of plant cell walls. Xylananases often exhibit a modular structure, with the catalytic domain linked to one or more noncatalytic domains, the majority of which are carbohydrate-binding modules (CBM). Xyn10B is comprised the catalytic domain GH10 and the carbohydrate-binding module CBM22. The two domains are covalent linked, but the linker is disordered, suggesting the linker is not directly involved in the complex formation. For target T35, the structure of CBM22 was given as a homology model; for target T36, the bound structure was given.

For both targets, the structure for GH10 was a homology model with PDB entry 1N82 as the template. This structure has a clearly defined binding cleft for polysaccharides. We reasoned that, during catalysis, the two domains of Xyn10B bind the same polysaccharide molecule. We could then line up the two domains using the fact that a polysaccharide molecule is nearly linear. The complex of the GH10 and CBM15 domains of CjXyn10C[34] seemed to support our strategy. However, the implementation of the strategy was hampered by the fact that the polysaccharide binding site on CBM22 was not known. We could narrow the binding site, but not sufficiently so. In the end, our best submitted models have an I_rmsd of 7.3 Å from the X-ray structure for both T35 and T36. These two targets also prove to be challenging for the community as a whole, with only one acceptable model submitted for either target.

Target 37: ARF6 and LZ2 of JIP4

Target 37 is the complex of the human G-protein ARF6 and LZ2, the second leucine zipper motif of JIP4 (JNK-interacting protein 4).[35] LZ2 forms a symmetric coiled-coil and was given as a homology model.

Small GTP-proteins are regulators of cellular traffic. Many structures in this family have been solved, including those in complex with effectors. The effectors in several of these structures are helical, like LZ2, although the helices differ in length. These include PDB

entries 1J2J[36] and 1R4A.[37] We selected five ZDOCK poses based on their similarities to these two complexes. We further enriched these poses by running RosettaDock, generating 5000 new poses for each. The submitted models were based on both similarities to 1J2J and 1R4A and favorable binding energies according to ZRANK. One of our submitted models is of acceptable quality.

Targets 38 and 39: Centaurin-α1 and FHA of KIF13B

Targets 38 and 39 are the complex between Centaurin-α1 and the Forkhead-associated (FHA) domain of KIF13B. FHA domains are a class of ubiquitous signaling modules that appear to function through interactions with phosphorylated target molecules, to fulfill the role of a modular phosphoserine/threonine binding domain. The structure of an FHA and a phosphopeptide suggests a binding site on the FHA domain.[38] Centaurin-α1 has a phosphothreonine site at S87 and a phosphothreonine at T276 by PKC.[39] The interaction between Centaurin-α1 and KIF13B has been reported.[40] The experiments with truncates suggested that Centaurin-α1 interacts with KIF13B through the former's ARF GAP domain (residues 7-126). We were therefore led to believe that S87 was a site of interaction with the FHA domain of KIF13B, which turns out to be incorrect.

Target 40: Bovine Trypsin and API-A

Target T40 is a complex between bovine trypsin and the double-headed arrowhead protease inhibitor (API-A).[41] The bound structure of API-A was given. Each inhibitor binds two trypsin molecules. Information was given that Leu87 and Lys145 of the inhibitor interact with the trypsin. The two residues are on two loops located on opposite sides of the inhibitor, so we assumed that they each interact with a different trypsin molecule.

Bovine trypsin can be inhibited by several other inhibitors with similar structural features. In particular, Alzheimer's amyloid beta-protein precursor (APPI) binds to trypsin through a loop [see Figure 1(E)].[42] In the binding interface, an APPI Lys residue is found in the P1 position, which is the site defining the primary specificity of the inhibitor (or substrate). This Lys residue inserts deeply into the enzyme's binding pocket. ZDOCK poses with API-A Lys145 positioned in a similar fashion were exclusively selected. We also assumed that the API-A Leu87 residue would be located in the P1 position in the other complex with trypsin. We were less certain of this assumption, and hence submitted only one model (model 2) for that complex.

All our nine models with API-A Lys145 in the interface are of acceptable quality or better, including three with high quality [see Figure 1(F)]. We credit this excellent performance largely to the similarity in the positioning of the Lys-containing loops of API-A and APPI on trypsin [see Figure 1(E)], since we used this local feature of APPI in its interaction with trypsin as our guide. Our model 2 has an I_rmsd of 5.3 Å from the complex with API-A Leu87 in the interface.

Target 41: Colicin E9 and Im2

Target T41 is the complex between the DNase domain of colicin E9 and the Im2 immunity protein.[43] The structure of the DNase domain of colicin E9 complexed with its cognate immunity protein, Im9, has been solved previously (PDB entry 1EMV).[44] Four DNase colicins, E2, E7, E8, and E9 share ~80% sequence identity in the C-terminal DNase domain. The respective cognate immunity proteins, Im2, Im7, Im8, and Im9, share ~50% sequence identity; their structures consist of four helices. It has been proposed that the conserved residues of Im helix III act as the anchor of the endonuclease binding site while the variable residues of Im helix II confers the specificity for the recognition of the cognate partner.[45]

Between Im2 and Im9, the sequence identity rises to ~70%, and the conserved residues of helix III are identical. Yet their binding affinities with the DNase domain of colicin E9 differ by six orders of magnitude.[45] We nevertheless assumed that the E9-Im9 complex served as a good model for the E9-Im2 complex. Poses were generated from two ZDOCK runs, one with the actual Im2 structure and another with a homology model of Im2 built on the bound structure of Im9. ZDOCK poses with L_rmsd < 15 Å from 1EMV were clustered to remove redundancy. The remaining 22 models were energy minimized and again clustered. The 10 models with the best ZRANK scores in their respective clusters were submitted. Four of our models turn out to be of at least acceptable quality, including one with high quality.

Target 42: Homodimer of Designed TPR Repeat

Target T42 is a homodimer of a designed TPR repeat.[46] The TPR repeat was given as a homology model, with PDB entry 1NA3 recommended as the template. We instead used PDB entry 1NA0[47] as the template for the monomer, because it has a longer sequence than 1NA3. Except for an extra C-terminal helix on 1NA0, the two structures are very similar. The designed protein was intended to break polar interactions between two subunits by polar to apolar mutations.

We ran M-ZDOCK with both the full-length homology model and with the C-terminal helix removed or blocked. For this target we had no information on the dimer interface and very little time to work. In our model selection, we simply favored compact structures over more open ones. Our best model has an I_rmsd of 4.4 Å.

Conclusion

In CARPI rounds 13-19, we submitted “correct” predictions for 6 of the 13 targets. The models are of high quality for two of the targets and of medium quality for two other targets. This along with similar performances in previous CAPRI rounds[4,10] demonstrates our sustained success in the CAPRI exercise. Our success in large part is due to the use of biochemical and structural information, instead of a scoring function alone. When no such information is available, interface prediction can still be used, as in target T29.

For many of the targets (T32, T33/T34, T37, T40, and T41), we made use of homology models or critical structural motifs. These structural data were valuable in model selection. Biochemical data can be equally helpful. However, a successful model is hinged on correct interpretation of the data.

Our results demonstrate the relative usefulness of current docking methods for predicting structures of protein complexes. In our view, a docking strategy that is capable of producing high-quality models will require some information on the interface, either known or accurately predicted, and a versatile local search method.

Acknowledgments

This work was supported in part by NIH grant GM058187.

References

1. Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. Proteins. 2003; 52(1): 80–87. [PubMed: 12784371]
2. Kozakov D, Brenke R, Comeau SR, Vajda S. PIPER: an FFT-based protein docking program with pairwise potentials. Proteins. 2006; 65(2):392–406. [PubMed: 16933295]

3. Schneidman-Duhovny D, Nussinov R, Wolfson HJ. Automatic prediction of protein interactions with large scale motion. Proteins. 2007; 69(4):764–773. [PubMed: 17886339]
4. Qin S, Zhou HX. A holistic approach to protein docking. Proteins. 2007; 69(4):743–749. [PubMed: 17803232]
5. Pierce B, Weng Z. ZRANK: reranking protein docking predictions with an optimized energy function. Proteins. 2007; 67(4):1078–1086. [PubMed: 17373710]
6. Gray JJ, Moughon S, 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(1):281–299. [PubMed: 12875852]
7. Li L, Chen R, Weng Z. RDOCK: refinement of rigid-body protein docking predictions. Proteins. 2003; 53(3):693–707. [PubMed: 14579360]
8. Shen Y, Paschalidis I, Vakili P, Vajda S. Protein docking by the underestimation of free energy funnels in the space of encounter complexes. PLoS Comput Biol. 2008; 4(10):e1000191. [PubMed: 18846200]
9. Zhou HX, Shan Y. Prediction of protein interaction sites from sequence profile and residue neighbor list. Proteins. 2001; 44(3):336–343. [PubMed: 11455607]
10. van Dijk AD, de Vries SJ, Dominguez C, Chen H, Zhou HX, Bonvin AM. Data-driven docking: HADDOCK's adventures in CAPRI. Proteins. 2005; 60(2):232–238. [PubMed: 15981252]
11. Qin S, Zhou HX. meta-PPISP: a meta web server for protein-protein interaction site prediction. Bioinformatics. 2007; 23(24):3386–3387. [PubMed: 17895276]
12. Chen H, Zhou HX. Prediction of interface residues in protein-protein complexes by a consensus neural network method: test against NMR data. Proteins. 2005; 61(1):21–35. [PubMed: 16080151]
13. Liang S, Zhang C, Liu S, Zhou Y. Protein binding site prediction using an empirical scoring function. Nucleic Acids Res. 2006; 34(13):3698–3707. [PubMed: 16893954]
14. Neuvirth H, Raz R, Schreiber G. ProMate: a structure based prediction program to identify the location of protein-protein binding sites. J Mol Biol. 2004; 338(1):181–199. [PubMed: 15050833]
15. Pierce B, Tong W, Weng Z. M-ZDOCK: a grid-based approach for Cn symmetric multimer docking. Bioinformatics. 2005; 21(8):1472–1478. [PubMed: 15613396]
16. Camacho CJ, Zhang C. FastContact: rapid estimate of contact and binding free energies. Bioinformatics. 2005; 21(10):2534–2536. [PubMed: 15713734]
17. Feig M, Karanicolas J, Brooks CL 3rd. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. J Mol Graph Model. 2004; 22(5):377–395. [PubMed: 15099834]
18. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc. 2003; 125(7):1731–1737. [PubMed: 12580598]
19. Leulliot N, Chaillet M, Durand D, Ulryck N, Blondeau K, van Tilbeurgh H. Structure of the yeast tRNA m7G methylation complex. Structure. 2008; 16(1):52–61. [PubMed: 18184583]
20. Purta E, van Vliet F, Tricot C, De Bie LG, Feder M, Skowronek K, Droogmans L, Bujnicki JM. Sequence-structure-function relationships of a tRNA (m7G46) methyltransferase studied by homology modeling and site-directed mutagenesis. Proteins. 2005; 59(3):482–488. [PubMed: 15789416]
21. Alexandrov A, Martzen MR, Phizicky EM. Two proteins that form a complex are required for 7-methylguanosine modification of yeast tRNA. Rna. 2002; 8(10):1253–1266. [PubMed: 12403464]
22. Alexandrov A, Grayhack EJ, Phizicky EM. tRNA m7G methyltransferase Trm8p/Trm82p: evidence linking activity to a growth phenotype and implicating Trm82p in maintaining levels of active Trm8p. Rna. 2005; 11(5):821–830. [PubMed: 15811913]
23. Tong Y, Chugha P, Hota PK, Alviani RS, Li M, Tempel W, Shen L, Park HW, Buck M. Binding of Rac1, Rnd1, and RhoD to a novel Rho GTPase interaction motif destabilizes dimerization of the plexin-B1 effector domain. J Biol Chem. 2007; 282(51):37215–37224. [PubMed: 17916560]
24. Oinuma I, Katoh H, Harada A, Negishi M. Direct interaction of Rnd1 with Plexin-B1 regulates PDZ-RhoGEF-mediated Rho activation by Plexin-B1 and induces cell contraction in COS-7 cells. J Biol Chem. 2003; 278(28):25671–25677. [PubMed: 12730235]

25. Kiel C, Foglierini M, Kuemmerer N, Beltrao P, Serrano L. A genome-wide Ras-effector interaction network. J Mol Biol. 2007; 370(5):1020–1032. [PubMed: 17544445]
26. Micheelsen PO, Vevodova J, De Maria L, Ostergaard PR, Friis EP, Wilson K, Skjot M. Structural and mutational analyses of the interaction between the barley alpha-amylase/subtilisin inhibitor and the subtilisin savinase reveal a novel mode of inhibition. J Mol Biol. 2008; 380(4):681–690. [PubMed: 18556023]
27. Lange G, Betzel C, Branner S, Wilson KS. Crystallographic studies of Savinase, a subtilisin-like proteinase, at pH 10.5. Eur J Biochem. 1994; 224(2):507–518. [PubMed: 7925366]
28. Vallee F, Kadziola A, Bourne Y, Juy M, Rodenburg KW, Svensson B, Haser R. Barley alpha-amylase bound to its endogenous protein inhibitor BASI: crystal structure of the complex at 1.9. A resolution Structure. 1998; 6(5):649–659.
29. Bonsager BC, Nielsen PK, Abou Hachem M, Fukuda K, Praetorius-Ibba M, Svensson B. Mutational analysis of target enzyme recognition of the beta-trefoil fold barley alpha-amylase/subtilisin inhibitor. J Biol Chem. 2005; 280(15):14855–14864. [PubMed: 15657043]
30. Takeuchi Y, Satow Y, Nakamura KT, Mitsui Y. Refined crystal structure of the complex of subtilisin BPN· and Streptomyces subtilisin inhibitor at 1·8 Å resolution. J Mol Biol. 1991; 221(1):309–325. [PubMed: 1920411]
31. Lebars I, Husson C, Yoshizawa S, Douthwaite S, Fourmy D. Recognition elements in rRNA for the tylosin resistance methyltransferase RlmA(II). J Mol Biol. 2007; 372(2):525–534. [PubMed: 17673230]
32. Das K, Acton T, Chiang Y, Shih L, Arnold E, Montelione GT. Crystal structure of RlmAI: implications for understanding the 23S rRNA G745/G748-methylation at the macrolide antibiotic-binding site. Proc Natl Acad Sci U S A. 2004; 101(12):4041–4046. [PubMed: 14999102]
33. Najmudin S, Pinheiro BA, Prates JAM, Romao MJ, Fontes CMGA. Putting an N-terminal end to the Clostridium thermocellum xylanase Xyn10B story: Crystallographic structure of the CBM22-1-GH10 modules complexed with xylohexaose. 2010
34. Pell G, Szabo L, Charnock SJ, Xie H, Gloster TM, Davies GJ, Gilbert HJ. Structural and biochemical analysis of Cellvibrio japonicus xylanase 10C: how variation in substrate-binding cleft influences the catalytic profile of family GH-10 xylanases. J Biol Chem. 2004; 279(12): 11777–11788. [PubMed: 14670951]
35. Isabet T, Montagnac G, Regazzoni K, Raynal B, El Khadali F, England P, Franco M, Chavrier P, Houdusse A, Menetrey J. The structural basis of Arf effector specificity: the crystal structure of ARF6 in a complex with JIP4. Embo J. 2009; 28(18):2835–2845. [PubMed: 19644450]
36. Shiba T, Kawasaki M, Takatsu H, Nogi T, Matsugaki N, Igarashi N, Suzuki M, Kato R, Nakayama K, Wakatsuki S. Molecular mechanism of membrane recruitment of GGA by ARF in lysosomal protein transport. Nat Struct Biol. 2003; 10(5):386–393. [PubMed: 12679809]
37. Wu M, Lu L, Hong W, Song H. Structural basis for recruitment of GRIP domain golgin-245 by small GTPase Arl1. Nat Struct Mol Biol. 2004; 11(1):86–94. [PubMed: 14718928]
38. Durocher D, Taylor IA, Sarbassova D, Haire LF, Westcott SL, Jackson SP, Smerdon SJ, Yaffe MB. The molecular basis of FHA domain:phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms. Mol Cell. 2000; 6(5):1169–1182. [PubMed: 11106755]
39. Zemlickova E, Dubois T, Kerai P, Clokie S, Cronshaw AD, Wakefield RI, Johannes FJ, Aitken A. Centaurin-alpha(1) associates with and is phosphorylated by isoforms of protein kinase C. Biochem Biophys Res Commun. 2003; 307(3):459–465. [PubMed: 12893243]
40. Venkateswarlu K, Hanada T, Chishti AH. Centaurin-alpha1 interacts directly with kinesin motor protein KIF13B. J Cell Sci. 2005; 118(Pt 11):2471–2484. [PubMed: 15923660]
41. Bao R, Zhou CZ, Jiang C, Lin SX, Chi CW, Chen Y. The ternary structure of the double-headed arrowhead protease inhibitor API-A complexed with two trypsins reveals a novel reactive site conformation. J Biol Chem. 2009; 284(39):26676–26684. [PubMed: 19640842]
42. Scheidig AJ, Hynes TR, Pelletier LA, Wells JA, Kossiakoff AA. Crystal structures of bovine chymotrypsin and trypsin complexed to the inhibitor domain of Alzheimer's amyloid beta-protein precursor (APPI) and basic pancreatic trypsin inhibitor (BPTI): engineering of inhibitors with altered specificities. Protein Sci. 1997; 6(9):1806–1824. [PubMed: 9300481]

43. Meenan NAG, Sharma A, Fleishman SJ, MacDonald C, Morel B, Boetzel R, Moore GR, Baker D, Kleanthous C. The structural and energetic basis for high selectivity in a high affinity protein-protein interaction. Proc Natl Acad Sci U S A. 2010 in press.

44. Li W, Dennis CA, Moore GR, James R, Kleanthous C. Protein-protein interaction specificity of Im9 for the endonuclease toxin colicin E9 defined by homologue-scanning mutagenesis. J Biol Chem. 1997; 272(35):22253–22258. [PubMed: 9268373]

45. Li W, Hamill SJ, Hemmings AM, Moore GR, James R, Kleanthous C. Dual recognition and the role of specificity-determining residues in colicin E9 DNase-immunity protein interactions. Biochemistry. 1998; 37(34):11771–11779. [PubMed: 9718299]

46. Krachler AM, Sharma A, Kleanthous C. Self-association of TPR-domains: natural and engineered. Proteins. 2010 in press.

47. Main ER, Xiong Y, Cocco MJ, D'Andrea L, Regan L. Design of stable alpha-helical arrays from an idealized TPR motif. Structure. 2003; 11(5):497–508. [PubMed: 12737816]

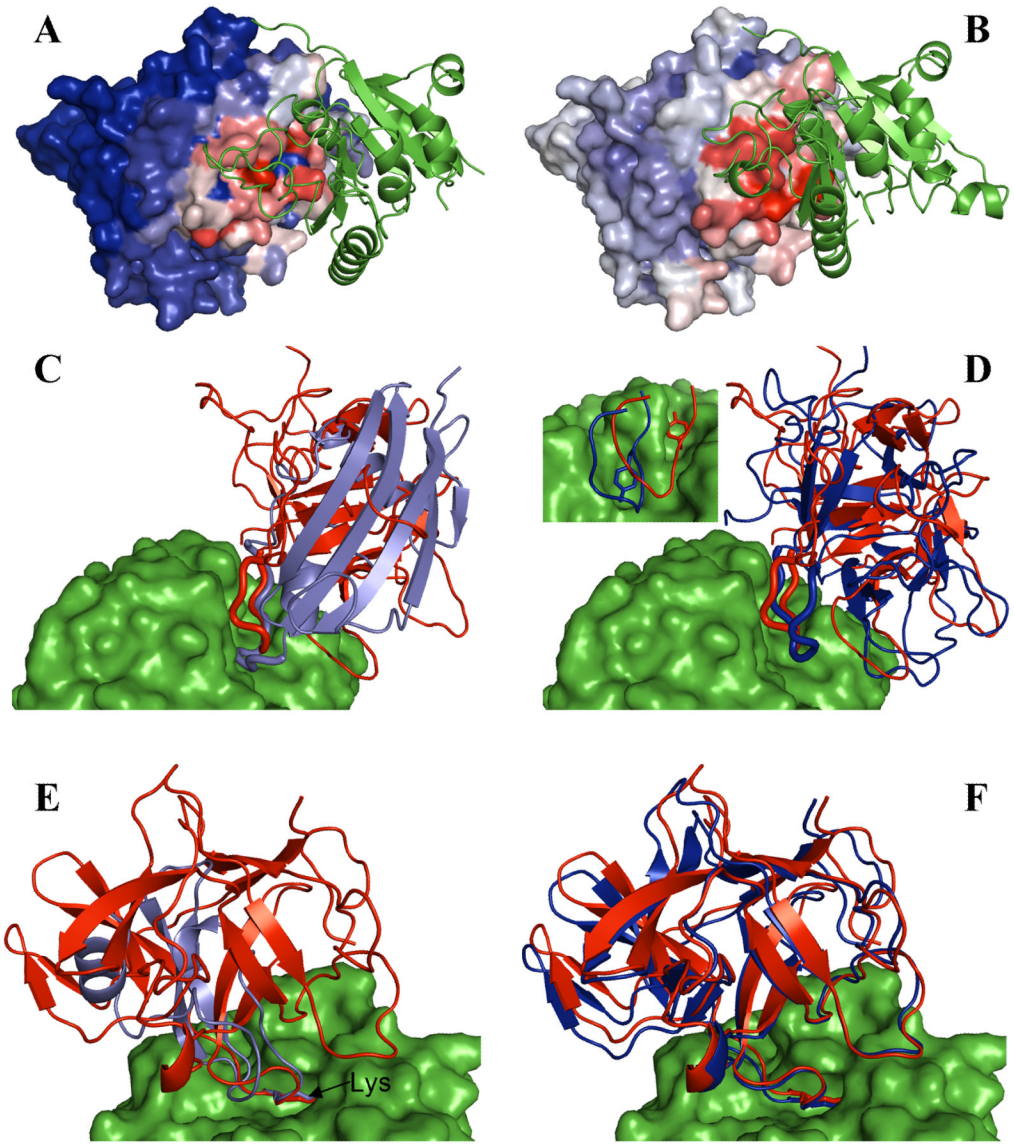

Figure 1.

Comparison of our best models for targets T29, T32, and T40 with their now released structures (see Table I for PDB names). The quality of our models for the three targets are medium, acceptable, and high, respectively.

(A) The X-ray structure of target T29. Trm82 is displayed as surface, with color coding indicating PINUP predictions of interface residues (red and blue for high and low interface probabilities, respectively); Trm8 is displayed as a ribbon representation.

(B) Best model (model 1) for target T29. Trm82 and Trm8 are presented in the same way as in (A), except that the color coding here displays the frequencies of individual Trm82 residues appearing in the interfaces of the ZDOCK poses. Red and blue indicate high and low frequencies, respectively. The quality of our model is shown by the resemblance in the position and orientation of Trm8 relative to Trm82 in (A) and (B).

(C) X-ray structure of target T32, highlighting the participation of a loop of BASI (thicker trace in the red ribbon representation) in the interaction with savinase (green surface). A similar loop of the subtilisin inhibitor (thicker trace in the light blue ribbon representation) is

involved in the interaction with subtilisin BPN· (PDB entry 2SIC). The two enzyme-inhibitor complexes are aligned by superimposing the enzymes; subtilisin BPN· is not shown. Except for the use of a loop in enzyme interaction, the two inhibitors do not have structural similarity.

(D) Comparison of our best model (model 6) for target T32 with the X-ray structure. The two structures are aligned by superimposing the enzyme (green surface); the inhibitor in the model and in the X-ray structure is shown as blue and red ribbons, respectively. Inset: difference in rotamers of BASI Y87 between the unbound and bound structures. Note that, in our model, Y87 in the unbound rotamter snugly fits into an alternative crevice on the surface of savinase.

(E) X-ray structure of target T40, highlighting the participation of a Lys-containing loop of API-A (red ribbon) with trypsin (green surface). A similar loop of APPI (light blue ribbon) is involved in the interaction with trypsin (PDB entry 1ATW). The two enzyme-inhibitor complexes are aligned by superimposing the enzyme. Except for the use of a Lys-containing loop in enzyme interaction, the two inhibitors do not have structural similarity.

(F) Comparison of our best model (model 6) for target T40 with the X-ray structure. The two structures are aligned by superimposing the enzyme (green surface); the inhibitor in the model and in the X-ray structure is shown as blue and red ribbons, respectively.

TABLE I

Summary of our performance in CAPRI rounds 13-19

Target	PDB name	Best L_rmsd	Best I_rmsd	Our group[a]	All groups[a]	Number of groups with correct predictions
T29	2VDU	6.4	1.4	1|1**	17|9**	7 of 39
T30		21.8	11.4		2|2*	2 of 35
T32	3BX1	6.7	3.0	1|1*	44|15***/13**	11 of 36
T33		30.5	18.8			0 of 28
T34		2.9	1.4	4|1**	65|25**	15 of 28
T35	2W5F	15.8	7.3		1|1*	1 of 34
T36		19.7	7.4		1|1*	1 of 32
T37	2W83	7.8	3.3	1|1*	21|1***/7**	11 of 39
T38	3FM8	52.7	19.0			0 of 40
T39		53.0	17.4		3|1***/2**	3 of 37
T40	3E8L	2.1	0.8	9|3***/4**	164|79***/54**	23 of 38
T41	2WPT	1.6	0.7	4|1***/2**	149|24***/58**	22 of 33
T42	2WQH	13.5	4.4		20|9***/5*	13 of 28

[a]Number before "|" denotes "correct" models; numbers after "|" indicate the best models, with "*", "**", and "***" denote number of models with acceptable, medium, and high quality, respectively.