



Published in final edited form as:

Proteins. 2010 November 15; 78(15): 3104–3110. doi:10.1002/prot.22764.

Performance of ZDOCK and ZRANK in CAPRI Rounds 13 - 19

Howook Hwang¹, Thom Vreven², Brian Pierce¹, Jui-Hung Hung¹, and Zhiping Weng^{1,2,3}

¹Bioinformatics Program, Boston University, Boston, Massachusetts 02215

²Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, Massachusetts 01605

³Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215

Abstract

We report the performance of the ZDOCK and ZRANK algorithms in CAPRI rounds 13-19, and introduce a novel measure Atom Contact Frequency (ACF). To compute ACF, we identify the residues that most often make contact with the binding partner in the complete set of ZDOCK predictions for each target. We used ACF to predict the interface of the proteins, which in combination with biological data available in the literature, is a valuable addition to our docking pipeline. Furthermore, we incorporated a straightforward and efficient clustering algorithm with two purposes: 1) to determine clusters of similar docking poses (corresponding to energy funnels), and 2) to remove redundancies from the final set of predictions. With these new developments, we achieved at least one acceptable prediction for targets 29 and 36, at least one medium quality prediction for targets 41 and 42, and at least one high quality prediction for targets 37 and 40; thus we succeeded for six out of a total of 12 targets.

1. Introduction

Since June 2001, the Critical Assessment of PRedicted Interactions (CAPRI)¹ experiment has provided the protein-protein docking community with opportunities to test algorithms and strategies on unknown targets. Our lab has participated since the first CAPRI round,^{2,3,4} using the ZDOCK⁵ and ZRANK⁶ docking programs. ZDOCK is a fast Fourier transform (FFT) based rigid-body docking algorithm, of which several versions have been released since its debut,^{5,7,8} and ZRANK reranks ZDOCK predictions with a more detailed scoring function. Compared with our previously reported strategy for solving CAPRI targets,⁴ we now include the Atom Contact Frequency (ACF) analysis, which integrates ZDOCK predictions to predict binding interfaces. In addition, we improved our method for clustering ZDOCK predictions.

2. Methods

Atom Contact Frequency

ZDOCK performs an exhaustive rigid body search in the 6-dimensional rotational and translational space. The three rotational angles are sampled with either a 6 degree or a 15 degree spacing, and the 3 translational degrees of freedom are sampled with a 1.2 Å spacing. For each set of rotational angles, only the best (based on ZDOCK score) translationally sampled prediction is retained. Generally for our CAPRI predictions, we use 6 degree sampling, resulting in 54,000 predictions.

The Atom Contact Frequency N_i for atom i is defined as the number of contacts n_{ik} (using a 6 Å cutoff) the atom makes with any atom of the binding partner, summed for a set of predictions k , and normalized:

$$N_i = M_i / \max \{M_1, \dots, M_j, \dots\}, \quad (1)$$

where

$$M_i = \sum_k n_{ik} \quad (2)$$

In this work, we used the 2000 highest-score ZDOCK predictions of each complex (i.e., $k=1, 2, \dots, 2000$). We predicted a residue to be in the interface if it has one or more atoms with $N_i > 0.7$. Performance assessment on rigid-body cases of the protein-protein docking benchmark 3.0⁹ shows that the method has a precision of 0.65, where the precision is defined as the number of correctly predicted interface residues (the number of true positives) divided by the total number of residues to predicted to be in the interface (sum of the true positives and false positives). Although the ACF provides only low resolution information, and no specific binding mode, it does allow the user to discard a large fraction of false-positive predictions, and improve the overall performance of the docking pipeline.

Clustering

We clustered the 54,000 ZDOCK predictions to retain only distinct predictions. In addition, the number of predictions in a cluster can help to identify false positive predictions.

For our clustering algorithm, the predicted complexes are rotated and translated to superimpose the receptors. We then express the similarity of two predictions as the root mean squares deviation (RMSD) between the C α atoms of the ligand structures. The complex with the highest ZDOCK score is taken to be the center of the first cluster. All predicted complexes that have RMSD's to this cluster center that are below a user-specified threshold are eliminated from the list of predictions, and these steps are repeated until no predictions remain in the list. The resulting cluster centers represent a pruned set of predictions that are spaced by at least the threshold, and with a bias towards high ZDOCK scores. For our CAPRI predictions, we generally consider only this pruned set.

General Docking Approach for CAPRI

As mentioned before, for the protein-protein targets we used ZDOCK 3.0 with 6 $^\circ$ rotational sampling, which produced 54,000 predictions per target. In addition to protein-protein docking targets, there were also protein-RNA targets, and we have not parameterized any of the ZDOCK versions for nucleic acids. Fanelli and Ferrari parameterized ZDOCK 2.3 for protein-DNA docking studies.¹⁰ Thus we used ZDOCK 2.3 with a modified version of their parameters for protein-RNA targets. We used top 30,000 ZDOCK predictions for clustering with a RMSD cutoff of 6 Å for all targets (unless stated otherwise).

In many cases, literature provides information regarding whether or not residues or domains are located in or near the binding interface. As described in our previous CAPRI papers,^{3,4} we used this information to place constraints ('blocking' and 'distance filtering') on the poses that ZDOCK considers. From the clusters we collected 10 predictions for submission to CAPRI. For this final selection we combine the ZDOCK scores with literature information, ACF results, cluster density, and visual assessment (which do not agree in all cases, as discussed below). Finally, the selected structures were refined using CHARMM¹¹ to reduce possible atom clashes. For the scoring challenge, in which predictions need to be picked from a set of predictions by us and other CAPRI teams, we used ZRANK, and followed the same overall strategy as for the docking, thus using the same biological information, clustering approach, and refinement step. The CAPRI evaluation team assign a

prediction to be *incorrect*, or of *acceptable*, *medium*, or *high* quality based on the evaluation methods.¹²

Results and Discussion

A summary of the prediction (ZDOCK) and scoring challenge (ZRANK) performance for the CAPRI targets of rounds 29 to 42 is shown in Table 1.

For the prediction challenge, we achieved at least one *acceptable* prediction for six of the targets out of 12 total number of targets. For two of these six targets, we achieved at least one medium quality prediction and for another two targets we achieved at least one *high* quality prediction. In Figure 1 and supplementary table S1 we show ACF results, as discussed in the text below.

We used ZRANK for the scoring challenge, and obtained at least one *acceptable* prediction for four targets. For all these targets, we were also successful in the prediction challenge. For two targets we were successful in prediction but not in scoring, and these were the result of either not participating (Target 36) or cancellation (Target 42).

In the remainder of this section we discuss the approach we followed for each target. Since our approach and results are very similar for prediction and scoring, we only discuss the former.

Target 29: tRNA methyltransferase 8 (TRM8) / tRNA methyltransferase 82 (TRM82)¹³

For Target 29 we were provided with the unbound (unpublished) structure of TRM8 and the bound structure of TRM82, both are tRNA guanine-N(7)-methyltransferases.¹³ Based on mutagenesis information and a conserved motif study by sequence alignment¹⁴, along with missing residue information of the provided structures, we blocked Pro73, Phe106, Arg128, Asn161, Pro185, His192, Glu260, Glu261 and Thr282 on TRM8. On TRM82 we blocked Ser2, Phe48, Pro100, Pro295, Ile305 and Leu436 based on the missing residue information. We submitted a structure from the 10 most populated clusters, which yielded an *acceptable* prediction with interface RMSD of 2.13 Å. From Figure 1 we can see that all the ACF identified residues lie in the interface.

Target 30: Rho family GTPase Rnd1 / Plexin-B1 dimer¹⁵

Target 30 is an unbound-unbound docking case of the Ras binding domain of the Plexin-B1 homodimer, and GTPase Rnd1 of the Rho family. An unpublished unbound homodimer of Plexin-B1¹⁵ was provided along with Rnd1 (PDB entry 2CLS chain A). Experimental results indicate that residues Leu1849, Val1850 and Pro1851 of Plexin-B1 are involved in the interaction with Rnd1,^{16,17} and mutagenesis studies indicate that Thr45 of the effector domain (residues 42-50) of Rnd1 is involved.¹⁷

The ACF results are consistent with the effector domain of Rnd1 being involved in binding, but not for Leu1849, Val1850 and Pro1851 of Plexin-B1. Instead, ACF predicts that residues 1809-1815 of Plexin-B1 are involved in binding. To deal with this discrepancy, we selected 5 clusters based on the ACF results, and 5 clusters based on distance filtering using literature information. From each cluster, we submitted the structure with the highest ZDOCK score. Our best prediction has an interface RMSD of 4.85 Å, which is classified as *incorrect*.

Target 32: Barley α -amylase – subtilisin inhibitor (BASI) / Subtilisin savinase¹⁸

Target 32 is an unbound/unbound docking target with the protease Subtilisin savinase (PDB entry 1SVN) binding to the bifunctional inhibitor BASI (PDB entry 1AVA chain C). In 1AVA, BASI is bound to Barley α -amylase isozyme AMY2.¹⁹ Literature information indicates that BASI, AMY2 and savinase form a ternary structure,²⁰ which implies that subtilisin and AMY2 do not share the interface on BASI. Furthermore, the complex of savinase and Chymotrysin Inhibitor 2 (CI2) (PDB entry 1LW2) indicates that CI2 inhibits savinase by blocking the catalytic triad Asp32-His64-Ser221.²¹ We therefore blocked the AMY2 binding site on BASI, and filtered the docking results for inclusion of the catalytic triad in the interface. The ACF prediction is consistent with the literature information. We submitted one structure each from the 10 most populated clusters.

While ACF predicted residues of savinase are located near the center of the interface, the ACF predicted residues of BASI are located at the edge of the interface area (Figure 1). Thus despite the accurate literature information, our predictions were *incorrect* for this easy target.

Target 33/34: Methyl transferase / RNA

Target 33 entailed docking homology models of methyl transferase and a 74 nucleotide-long RNA. For Target 34, the homology model of the RNA was replaced with the bound structure. For both targets we used a homology model of the methyl transferase built with insight II,²² based on PDB entry 1P91. For Target 33, the RNA was modeled based on the provided template (PDB entry 2AW4) and sequence alignment using the Rosetta²³ program. Literature information indicates that the methyl transferase transfers a methyl from a sadenosyl-methionine (SAM) molecule to N1 of a particular base of the RNA.²⁴ Therefore, SAM was incorporated in the homology modeled methyl transferase, based on a homologous protein that includes SAM.²⁵ Distance filtering was applied to the SAM binding site of the methyl transferase and the particular base of our interest of the RNA.

We only calculated the ACF for the protein binding partner, the methyl transferase. Although the ACF result is consistent with the literature information, the homology modeled RNA for Target 33 was not accurate enough to obtain a near native complex, and our best prediction has an interface RMSD of 16.21 Å. Using bound RNA in Target 34 still yields an *incorrect* prediction, although the interface RMSD was significantly improved to 5.00 Å.

Target 35/36: C-terminal Carbohydrate Binding Module 22 of Xylanase (CBM22-1) / Glycoside Hydrolysis domain of Xylanase (GH10)²⁶

Xylanase 10B has three domains: the GH10 catalytic domain, the CBM22-1 domain and the CBM22-2 domain.^{27,28} The CBM22-1 and CBM22-2 domains are covalently connected through two 7-residue linkers to the N-terminus and C-terminus of the GH10 domain, respectively. Target 35 is an intramolecule docking problem with homology models of the GH10 domain and the CBM22-1 domain. For Target 36 we were provided with a bound form of CBM22-1. Using Insight II,²² we obtained homology models of GH10 (using PDB entry 1N82 as a template), as well as CBM22-1 (using PDB entry 1DY0 as a template) for Target 35.

We applied distance filtering to mimic the linker that connects the two domains: between the N-terminus of the GH10 domain and the C-terminus of CBM22-1. ZDOCK predictions with distances > 25 Å were excluded. We submitted the structures with the highest ZDOCK scores of the 10 most populated clusters.

ACF calculations on Target 35 GH10 (Figure 1) show that the predicted residues are completely misplaced from the interface area of the complex. However, predicted residues of CBM22-1 are located near the center of interface of the complex. Our best prediction was *incorrect*. Since our set of submitted structures included both structures that agree with the ACF, and structures that do not agree, the interface RMSD of our best prediction is reasonably small, 4.61 Å.

Target 36 is the same complex as target 35, but with bound CBM22-1. Furthermore, additional literature searches identified homologous complexes (PDB entries 1WKY and 1PX8) that suggest that CBM22-2 possibly interacts with helices 7 (residues 282-307) and 8 (residues 342-350). Based on this additional literature information, we blocked helices 7 and 8 from the GH10 catalytic domain. In addition, we blocked residues 131 to 266 of the GH10 catalytic domain, which are inaccessible due to the 7-residue linker. We applied the same distance filtering as for Target 35 to mimic the covalent linkage between the two proteins. Again, we submitted the structures with the highest ZDOCK scores of the 10 most populated clusters.

From the ACF calculation we can see the dramatic improvement it makes to use a bound instead of a homology modeled structure of CBM22-1, especially for GH10. All ACF predicted residues are in the interface area of the complex. This led to an *acceptable* prediction, with an interface RMSD of 3.65 Å.

Target 37: Leucine Zipper 2 (LZ2) / ADP-ribonylation factor 6 (ARF6)²⁹

For Target 37 we docked unbound ARF6 (PDB entry 2A5D) and a homology model of the LZ2 domain of JIP4. The homology model is based on GCN4 leucine zipper domain (PDB entry 2ZTA), and provided by Alexandre Bonvin's group. We selected the second out of Bonvin's 16 homology models, since it displays the smallest amount of bending, and therefore has the most residues accessible by the binding partner.

The C-terminus and N-terminus of ARF6 were blocked during the docking. Based on a homologous complex between RhoA and ROCK1 (PDB entry 1S2C),³⁰ we concluded that ARF6 switch I (residues 43-50) and switch II (residues 67-77) are in the interface of the target complex, which is supported by our ACF calculation. We applied distance filtering using 4 hydrophobic residues (Leu16, Val25, Ile30 and Thr37) that have high ACF on LZ2, as well as switch I and switch II of ARG6. After clustering, we choose 10 clusters, and from those submitted the structure with the highest ZDOCK score.

Figure 1 shows that residues predicted by ACF for ARF6 are located near the center of the interface, and most of the ACF predicted residues for LZ2 are located in the interface of the complex. This resulted in our best prediction to be of *high* quality, with an interface RMSD of 0.99 Å.

Target 38/39: Centaurin- α 1 / Forkhead-associated domain of Kinesin-like protein (KIF13B FHA)

Target 38 is an unbound-homology model docking case of centaurin- α 1 (PDB entry 3FEH) and the FHA domain of KIF13B. However, it was soon announced to be an unofficial CAPRI target, and followed by Target 39, which involved the same proteins but replaces the homology model of the FHA domain with the bound structure (PDB entry 3FM8).

Centaurin- α 1 has three domains, GAP, N-PH and C-PH. Literature information suggests that the FHA domain of KIF13B is bound to the GAP domain.³¹ This is in contrast with our ACF results for centaurin- α 1. However, we decided to rely on the literature, and hence blocked the N-PH and C-PH domains of centaurin- α 1, along with the C and N-termini of

FHA. This resulted our best prediction to be *incorrect*, with an interface RMSD of 15.64 Å. The native complex, in fact, agrees with the ACF analysis: the predicted residues of centaurin- α 1 and KIF13B are all located in the interface area of the complex (Figure 1). Relying on the ACF calculation instead of the literature would have resulted in a much more accurate prediction.

Target 40: Bovine trypsin / Double-headed arrowhead protease inhibitor A (APIA)³²

Target 40 is the complex of an unbound x-ray structure bovine β -trypsin (PDB entry 1BTY) and the bound structure of double-headed arrowhead protease inhibitor A, APIA.

Mutagenesis information indicates that Arg76 of APIA plays a role in the binding process.³³⁻³⁵ However, the location of this residue in the given APIA structure prohibits it to contact trypsin directly. Rather, Arg76 stabilizes a surface loop that contains the potential P1 residue Lys145. In addition, we used BLAST³⁶ to find complexes of bovine trypsin with various other inhibitors (PDB entries 1TGS, 1SBW, 1F2S, 1TAB, 1OPH, 2FTM, 1TAW, 1ZR0, 2FI4 and 2FT1). We found that all these peptides contain a Lys or Arg P1 residue that contacts the S1 pocket of bovine trypsin. Furthermore, the x-ray structure of APIA and chymotrypsin (PDB entry 3BX1) suggests Leu87 to be a functional residue of the inhibitor. Lys145 and Leu87 correspond to different possible binding sites, and we applied two rounds of distance filtering. In the first run, we included the S1 pocket of bovine trypsin, and the surface loop (residue number 143-149) of APIA containing Lys145. In a second run, we applied distance filtering to include the S1 pocket of bovine trypsin and Leu87 of APIA.

The native complex of Target 40 shows that APIA has two alternative binding modes with trypsin, involving Lys145 or Leu87. The ACF calculation on APIA shows that ZDOCK prefers the binding through Leu87 (Figure 1). This resulted our best prediction to be of *high* quality, with an interface RMSD of 0.64 Å. Since we knew that there were two binding sites, we also submitted predictions for the interface that includes Lys145, of which the best is *medium* quality with an interface RMSD of 1.01 Å.

Target 41: Colicin E9 / Immunity protein 2 (IM2)

For Target 41, we docked the unbound structures of colicin E9 (PDB entry 1FSJ) and the IM2 immunity protein (PDB entry 2NO8). We examined complexes that were homologous to the target, namely colicin E7 and IM7 (PDB entry 7CEI),³⁷ and colicin E9 and IM9 (PDB entry 1EMV).³⁸ Along with the homologous complexes and additional literature information,³⁹ we concluded that Phe86 on colicin E9 and Tyr54 on IM2 play key roles in the interaction.

ACF results for both colicin E9 and IM2 are consistent with Phe86 and Tyr54 location in the interface, and distance filtering was used with the two key residues. This led to our best prediction to be of *medium* quality, with an interface RMSD of 1.73 Å.

Target 42: Tetrcopeptide repeat (TPR) symmetric homodimer

Target 42 is a symmetric homodimer, for which the monomeric structure required homology modeling. Although the target sequence and a template structure of TPR (PDB entry 1NA3) were provided, we decided to use a different template (PDB entry 1NA0). For our predictions we used M-ZDOCK 3.0⁴⁰, which was developed specifically for homomultimeric docking. In the absence of literature information, we relied entirely on the ACF analysis, which identified residues Tyr81, Tyr88 and Lys89 to be in the interface.

According to the CAPRI organizers, TPR can form two distinct symmetric homodimers. From Figure 1 we see that only one of the binding modes is reflected by our ACF results,

and we only made predictions for this mode. Our best prediction has an interface RMSD of 1.20 Å, and was classified as of *medium* quality.

Conclusions

Overall ZDOCK performed well for the current targets, providing a successful demonstration of ZDOCK 3.0 in the area of blind protein-protein docking. The performance of ZRANK for the scoring was similar to ZDOCK, in that there was a substantial overlap in the targets where successful predictions were made. Notably, for three out of four targets where both docking and scoring yielded at least an *acceptable* prediction, the top scoring prediction was better, highlighting the ability of ZRANK to identify high quality predictions with low RMSD from native.

We introduced the ACF analysis for these rounds, which correctly predicted interface residues in 17 out of 21 monomers (Figure 1 and Supplementary Table S1). For Target 39 the ACF analysis is not consistent with the available literature. We relied on the latter for our prediction, which was incorrect, and in retrospect the ACF turned out to be correct. It appears that the ACF analysis is reliable enough to be an integral component of the docking pipeline.

For the first time in CAPRI, Targets 33 and 34 involved a complex of a protein with a RNA molecule. Considering that the ZDOCK scoring function was developed specifically for protein-protein docking, it is encouraging that for Target 34 we obtained a prediction that was very close to being *acceptable*. As expected, ZDOCK performs better with higher quality input structures, which explains the poorer performance with homology models as input. Still, one of ZDOCK's strengths is its generality, obtaining correct predictions for a variety of target types and qualities of input structures. As in the previous CAPRI rounds, ZDOCK is among the best performing docking algorithms, in that we are among the groups that made at least one *acceptable* prediction for the largest number of targets (six targets). In addition, our predictions were close to the cutoff for being *acceptable* for four other targets (30, 32, 34, 35).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was funded by NIH grant R01 GM084884 awarded to ZW.

References

1. Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJ, Vajda S, Vakser I, Wodak SJ. CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins*. 2003; 52(1):2–9. [PubMed: 12784359]
2. Chen R, Tong W, Mintseris J, Li L, Weng Z. ZDOCK predictions for the CAPRI challenge. *Proteins*. 2003; 52(1):68–73. [PubMed: 12784369]
3. Wiehe K, Pierce B, Mintseris J, Tong WW, Anderson R, Chen R, Weng Z. ZDOCK and RDOCK performance in CAPRI rounds 3, 4, and 5. *Proteins*. 2005; 60(2):207–213. [PubMed: 15981263]
4. Wiehe K, Pierce B, Tong WW, Hwang H, Mintseris J, Weng Z. The performance of ZDOCK and ZRANK in rounds 6-11 of CAPRI. *Proteins*. 2007; 69(4):719–725. [PubMed: 17803212]
5. Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. *Proteins*. 2003; 52(1): 80–87. [PubMed: 12784371]
6. Pierce B, Weng Z. ZRANK: reranking protein docking predictions with an optimized energy function. *Proteins*. 2007; 67(4):1078–1086. [PubMed: 17373710]

7. Mintseris J, Pierce B, Wiehe K, Anderson R, Chen R, Weng Z. Integrating statistical pair potentials into protein complex prediction. *Proteins*. 2007; 69(3):511–520. [PubMed: 17623839]
8. Chen R, Weng Z. A novel shape complementarity scoring function for protein-protein docking. *Proteins*. 2003; 51(3):397–408. [PubMed: 12696051]
9. Hwang H, Pierce B, Mintseris J, Janin J, Weng Z. Protein-protein docking benchmark version 3.0. *Proteins*. 2008
10. Fanelli F, Ferrari S. Prediction of MEF2A-DNA interface by rigid body docking: a tool for fast estimation of protein mutational effects on DNA binding. *J Struct Biol*. 2006; 153(3):278–283. [PubMed: 16427316]
11. Brooks BR, Brooks CL 3rd, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: the biomolecular simulation program. *J Comput Chem*. 2009; 30(10):1545–1614. [PubMed: 19444816]
12. Mendez R, Leplae R, Lensink MF, Wodak SJ. Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures. *Proteins*. 2005; 60(2):150–169. [PubMed: 15981261]
13. 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]
14. 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]
15. 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]
16. Oinuma I, Ishikawa Y, Katoh H, Negishi M. The Semaphorin 4D receptor Plexin-B1 is a GTPase activating protein for R-Ras. *Science*. 2004; 305(5685):862–865. [PubMed: 15297673]
17. Oinuma I, Katoh H, Harada A, Negishi M. Direct interaction of Rnd1 with Plexin-B1 regulates PDZRhoGEF-mediated Rho activation by Plexin-B1 and induces cell contraction in COS-7 cells. *J Biol Chem*. 2003; 278(28):25671–25677. [PubMed: 12730235]
18. 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]
19. 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 Å resolution. *Structure*. 1998; 6(5):649–659. [PubMed: 9634702]
20. Abe J, Sidenius U, Svensson B. Arginine is essential for the alpha-amylase inhibitory activity of the alpha-amylase/subtilisin inhibitor (BASI) from barley seeds. *Biochem J*. 1993; 293(Pt 1):151–155. [PubMed: 8328958]
21. Radisky ES, Koshland DE Jr. A clogged gutter mechanism for protease inhibitors. *Proc Natl Acad Sci U S A*. 2002; 99(16):10316–10321. [PubMed: 12142461]
22. Accelrys. Insight II, version 2005. Accelrys. 2005
23. Rohl CA, Strauss CE, Misura KM, Baker D. Protein structure prediction using Rosetta. *Methods Enzymol*. 2004; 383:66–93. [PubMed: 15063647]
24. 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]
25. 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]
26. Najimudin S, Pinheiro BA, Prates JAM, Romao MJ, Fontes CM. Putting an N-terminal end to the *Clostridium thermocellum* xylanase Xyn10B story: Crystallographic structure of the CBM22-1-

- GH10 modules complexed with xylohexanase. *Acta Crystallographica section F*. 2008; (64):715–718.
27. Xie H, Gilbert HJ, Charnock SJ, Davies GJ, Williamson MP, Simpson PJ, Raghothama S, Fontes CM, Dias FM, Ferreira LM, Bolam DN. Clostridium thermocellum Xyn10B carbohydrate-binding module 22-2: the role of conserved amino acids in ligand binding. *Biochemistry*. 2001; 40(31):9167–9176. [PubMed: 11478884]
 28. Charnock SJ, Bolam DN, Turkenburg JP, Gilbert HJ, Ferreira LM, Davies GJ, Fontes CM. The X6 “thermostabilizing” domains of xylanases are carbohydrate-binding modules: structure and biochemistry of the Clostridium thermocellum X6b domain. *Biochemistry*. 2000; 39(17):5013–5021. [PubMed: 10819965]
 29. 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]
 30. Dvorsky R, Blumenstein L, Vetter IR, Ahmadian MR. Structural insights into the interaction of ROCK1 with the switch regions of RhoA. *J Biol Chem*. 2004; 279(8):7098–7104. [PubMed: 14660612]
 31. Venkateswarlu K, Hanada T, Chishti AH. Centaurin- α 1 interacts directly with kinesin motor protein KIF13B. *J Cell Sci*. 2005; 118(Pt 11):2471–2484. [PubMed: 15923660]
 32. 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 trypsin reveals a novel reactive site conformation. *J Biol Chem*. 2009; 284(39):26676–26684. [PubMed: 19640842]
 33. Xie ZW, Luo MJ, Xu WF, Chi CW. Two reactive site locations and structure-function study of the arrowhead proteinase inhibitors, A and B, using mutagenesis. *Biochemistry*. 1997; 36(19):5846–5852. [PubMed: 9153425]
 34. Li J, Chi CW, Ruan KC. Conformation nearby Trp residues of APIA and APIB modulates the inhibitory specificity of the protease. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)*. 2002; 34(4):494–497. [PubMed: 12098774]
 35. Li J, Ruan KC, Chi CW. The assignment of the reactive sites of the double-headed arrowhead proteinase inhibitor A and B. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)*. 2002; 34(5):662–666. [PubMed: 12198573]
 36. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215(3):403–410. [PubMed: 2231712]
 37. Ko TP, Liao CC, Ku WY, Chak KF, Yuan HS. The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. *Structure*. 1999; 7(1):91–102. [PubMed: 10368275]
 38. Kuhlmann UC, Pommer AJ, Moore GR, James R, Kleantous C. Specificity in protein-protein interactions: the structural basis for dual recognition in endonuclease colicin-immunity protein complexes. *J Mol Biol*. 2000; 301(5):1163–1178. [PubMed: 10966813]
 39. Li W, Hamill SJ, Hemmings AM, Moore GR, James R, Kleantous 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]
 40. 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]

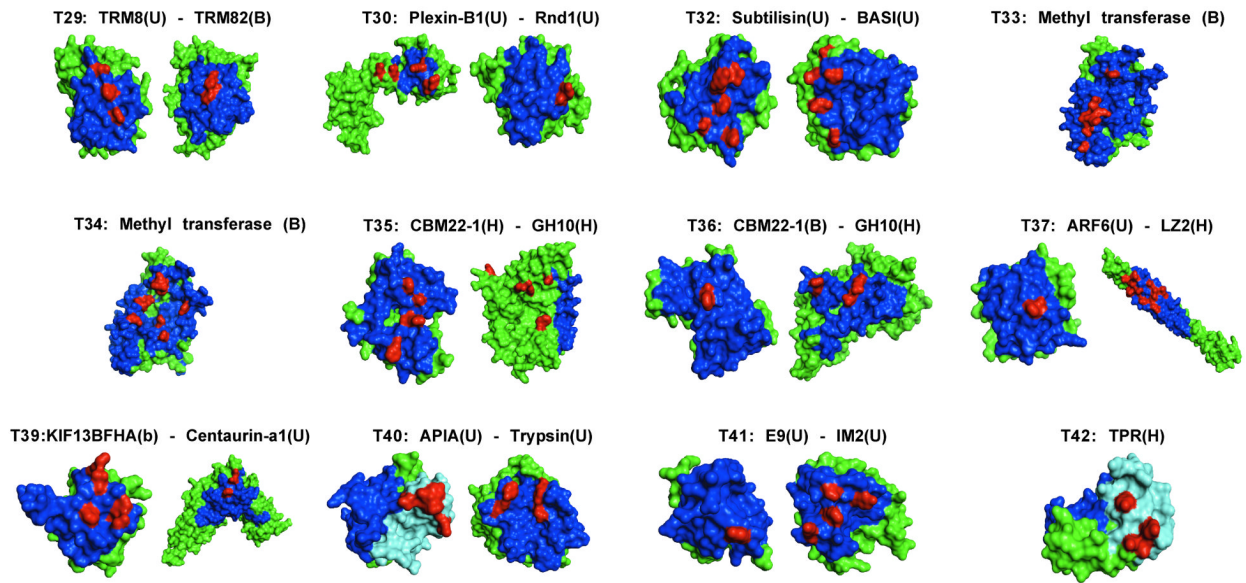


Figure 1.

Monomers of the Targets, with interface residues in blue (different shades for the interfaces corresponding to the distinct binding modes of Targets 40 and 42), and residues predicted by ACF to be in the interface in red. The origin of the monomer is in parenthesis (B, H, and U for bound, homology modeled, and unbound, respectively)

Table 1

Summary of ZDOCK and ZRANK performance for CARPI Rounds 13-19

| Target | Complex | Prediction (ZDOCK) | | | | Scoring (ZRANK) | | |
|--------|------------------------------------|--------------------|-----------------------------|------------------------|----------------------|-----------------------------|------------------------|----------------------|
| | | Type ^a | Classification ^b | f_{nat} ^c | $iRMSD$ ^c | Classification ^b | f_{nat} ^c | $iRMSD$ ^c |
| 29 | TRM8 / TRM82 | U/B | Acceptable | 49 | 2.131 | Medium | 69 | 1.120 |
| 30 | Rnd1 / Plexin-B1 | U/U | Incorrect | 16 | 4.849 | Incorrect | 0 | 10.039 |
| 32 | BASI / Subtilisin | U/U | Incorrect | 25 | 5.505 | Incorrect | 1 | 9.453 |
| 33 | Methyl transferase / RNA | H/H | Incorrect | 10 | 16.212 | Incorrect | 1 | 19.537 |
| 34 | Methyl transferase / RNA | H/B | Incorrect | 27 | 4.997 | Not participated | | |
| 35 | CBM22-1 / GH10 | H/H | Incorrect | 7 | 4.606 | Incorrect | 4 | 8.716 |
| 36 | CBM22-1 / GH10 | B/H | Acceptable | 37 | 3.649 | Not participated | | |
| 37 | LZ2 / ARF6 | H/U | High | 59 | 0.988 | High | 92 | 0.939 |
| 39 | Centaurin- α 1 / KIF13B FHA | U/B | Incorrect | 0 | 15.639 | Incorrect | 0 | 15.643 |
| 40 | Bovine Trypsin / APIA | U/U | Medium ^d | 92 | 1.007 | High ^d | 86 | 0.433 |
| | | | High ^d | 95 | 0.636 | Incorrect ^d | 0 | 19.879 |
| 41 | Colicin E9 / IM2 | U/U | Medium | 46 | 1.728 | Acceptable | 63 | 2.45 |
| 42 | TPR | H/H | Medium ^d | 100 | 1.199 | N/A | | |
| | | | Incorrect ^d | 15 | 4.135 | N/A | | |

^aU, B, and H denote unbound, bound, and homology modeling input structures, respectively.

^bBest classification of our 10 predictions, as assessed by the CAPRI evaluation team.

^c f_{nat} (%) and interface RMSD ($iRMSD$, Å) provided by the CAPRI evaluation team. In case of multiple predictions of the same classification, we report the prediction with the lowest interface RMSD.

^dResults are reported for the two distinct binding modes for targets 40 and 42.