Published in final edited form as:

Proteins. 2017 March ; 85(3): 408–416. doi:10.1002/prot.25186.

Performance of ZDOCK and IRAD in CAPRI rounds 28–34

Thom Vreven*, Brian G. Pierce#, Tyler M. Borrmann, and Zhiping Weng*

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

Abstract

We report the performance of our protein-protein docking pipeline, including the ZDOCK rigid-body docking algorithm, on 19 targets in CAPRI rounds 28–34. Following the docking step, we reranked the ZDOCK predictions using the IRAD scoring function, pruned redundant predictions, performed energy landscape analysis, and utilized our interface prediction approach RCF. In addition, we applied constraints to the search space based on biological information that we culled from the literature, which increased the chance of making a correct prediction. For all but two targets we were able to find and apply biological information and we found the information to be highly accurate, indicating that effective incorporation of biological information is an important component for protein-protein docking.

Keywords

Protein-protein interaction; Docking; Complex; Structure; ZRANK

Introduction

For more than ten years, the Critical Assessment of PRedicted Interactions (CAPRI) experiment has been a catalyst for the development of methods for the prediction of protein-protein complex structures, with more recent expansions to the prediction of mutation energies, protein design, and protein-peptide binding.¹ Our laboratory has been a participant from the inception of CAPRI, using the ZDOCK and M-ZDOCK algorithms, a series of reranking functions, and several refinement programs.^{2–6} Here we report our results from CAPRI rounds 28 to 34, which were held from 2013 to 2015.

ZDOCK performs an exhaustive, grid-based search for the binding modes of two component proteins.^{7–9} The proteins are kept rigid in their unbound structures and the search space of their relative orientations is exhaustively explored in three rotational and three translational degrees of freedom, with the search of the translational degrees of freedom substantially sped up using a fast Fourier transform (FFT) approach. For each angle combination only the

*Thom Vreven, Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, ASC-5th floor room 1079, 368 Plantation St., Worcester, MA 01605, Phone: 508-856-2272; Fax: 508-856-2392, Thom.Vreven@umassmed.edu. Zhiping Weng, Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, ASC-5th floor room 1069, 368 Plantation St., Worcester, MA 01605, Phone: 508-856-8866; Fax: 508-856-2392, Zhiping.Weng@umassmed.edu.

#Present address: University of Maryland Institute for Bioscience and Biotechnology Research, 9600 Gudelsky Drive, Rockville, MD 20850

best scoring translation is kept, resulting in 3,600 or 54,000 predictions for 15° or 6° angular sampling respectively. Because the FFT approach requires that ZDOCK scoring functions be in the form of convolutions, we also developed a series of scoring functions for subsequent ranking of ZDOCK predictions.^{10–12}

In Table 1 we list the CAPRI rounds and targets covered in this paper. New to this series of rounds is a large number of protein-peptide targets. The results of round 31, the CASP-CAPRI experiment, will be published separately as a community-wide paper and is not discussed here.

Methods

This section outlines our general strategy for generating CAPRI predictions. Target-specific modifications or additions to the general strategy are described in the respective sub-sections in Results. For most targets we used the ZDOCK program for rigid-body docking with a 6° angular sampling, resulting in 54,000 predictions. ZDOCK needs the structures of the component proteins as inputs, and when unbound x-ray or Nuclear magnetic resonance (NMR) structures were not available we used Modeller to generate the structures from homologous proteins (called templates).¹³

When information regarding the interface location was available from the literature, we used the information prior to or post ZDOCK runs. In some cases we blocked the residues that were unlikely to be in the interface by assigning them an unfavorable atom type so that ZDOCK would be less likely to predict complex structures with these residues at the binding interface; alternatively, we filtered out the ZDOCK predictions with these residues at the interface. In other cases with experimental information on interacting residues at the interface, we identified predictions with sufficiently short distances between these residues.

We then clustered ZDOCK predictions to identify regions of the search space that had high densities of predictions and were more likely to contain a correct prediction. We applied the Residue Contact Frequency (RCF) method to identify residues that were most often found in the interface in the collection of ZDOCK predictions.¹⁴ We pruned predictions to remove highly similar structures.^{5, 15, 16}

From the remaining predictions we used the ZDOCK score, IRAD score,¹¹ and prediction density to make a short list of structures to inspect manually. We relied on RCF or experimental information to select the final ten models for submission. The selected models were then refined using Rosetta to remove possible steric clashes between the component proteins.¹⁷

For the CAPRI scoring rounds, we used the IRAD score¹¹ and manual inspection for the selection of ten models, followed by refinement using Rosetta.

Results

Target 59

Target 59 was a complex between EDC3, a component of the mRNA decapping machinery, and the ribosomal protein RPS28B. For EDC3 we used the unbound structure (PDB ID 4A53; NMR structure model 1) suggested by CAPRI. We generated a homology model for RPS28B using the template suggested by CAPRI, an NMR structure (PDB ID 1NE3) with 53% sequence identity to RPS28B. We identified an x-ray structure with a much higher sequence identity (85%) to RPS28B (PDB ID 3U5G); thus we also made predictions using a homology model built from this template.

To guide our selections of the ten final complex structures, we considered the following experimental information: (1) RPS28B binds RNA through two loops with residues 21-25 and 46-50 (PDB IDs 3U5B and 3U5C). We assumed these loops would not be involved in RPS28B-EDC3 binding; (2). Fromm et al.¹⁸ speculated that RPS28B bound EDC3 through an HML sequence (residues 51-58); (3) The peer reviews of the publication by Fromm et al.,¹⁸ which were available online, suggested that based on NMR experiments, the residues that were involved in EDC3 binding to the decapping enzyme DCP2 (residues 3 and 52-57), were also involved in EDC3-RPS28B binding.

We achieved a single *acceptable* prediction based on the homology model of RPS28B built from the NMR structure template suggested by CAPRI. The N-terminus of RPS28B was in the interface, and the structures of the N-terminus differed substantially between the NMR structure template and the x-ray structure template, with the latter being more distant from the bound structure despite its higher sequence identity (85% vs. 53%). In hindsight, this might be the reason that the CAPRI team recommended the NMR structure as the template. From our predicted model, it appears that all the experimental information we used to guide our selection was correct.

Targets 60–64

Targets 60–64 were peptides (13–15 aa in length) bound to mouse importin- α . The crystal structure of mouse importin- α was supplied by CAPRI (PDB ID 1EJL). We found that all five peptides contained the KRX(W/F/Y)XXAF motif, which was found in the Class 3 nuclear localization signal (NLS) family members that bound the minor site of importin;¹⁹ thus, we focused our effort on this region of the receptor. We used Modeller to produce two homology models for each peptide using two mouse importin- α /peptide complex crystal structures as templates, which contained distinct peptides bound to the minor NLS site (PDB IDs 2C1M²⁰ and 3UKW²¹).

We used Rosetta's FlexPepDock, a local protein-peptide docking protocol,²² to flexibly dock the peptides to importin using each of the two homology models for each peptide (flags: “-pep_refine -ex1 -ex2aro -unboundrot 2C1M.pdb/3UKW.pdb”), generating 200 models per run. For each starting peptide structure we also ran FlexPepDock with and without the “-lowres_preoptimize” flag, thus giving a total of 800 models per target. We then filtered out the models in which the protein did not have sufficient contacts with the peptide. We required the Lys and Arg residues at the N-terminus of the peptide motif to be within 4.5

Å of at least 15 non-hydrogen atoms of importin residues 328 and 396 and the Ala and Phe residues at the C-terminus of the peptide motif to be within 4.5 Å of at least 9 non-hydrogen atoms of importin residues 328 and 396. We then used the ZRANK2 scoring function¹² to rank the remaining structures and selected top structures for submission. We also submitted one model per target generated using a FlexPepDock model for Target 64 exhibiting strong protein-peptide hydrophobic stacking as template, as well as two models using just the N-terminal contact filter but not the C-terminal filter. Additionally, to ensure non-redundancy, only FlexPepDock models with >0.5 Å root mean square deviation (RMSD) from all previously selected models were added to the set selected for submission.

Predictions for these targets were successful overall. As it turned out, these peptides bind to both the minor and major sites of importin, and CAPRI provided assessments using all peptide residues for both the major and minor NLS sites, as well as using only the core six residues (XKRXXX) of the peptide for the minor NLS site. Using the core six residues and the minor NLS site, we achieved 9–10 *medium accuracy* predictions per ten submitted models for targets 60–62 and Target 64, and for Target 63 all submitted models were classified *acceptable*. When assessed using all peptide residues, our predictions were *incorrect*. This was likely due to lack of conformational sampling outside of the core region of peptide. Because we only performed docking on the minor site of importin, none of our predictions were correct when assessed using the major NLS site.

Target 65 and 66

Targets 65 and 66 were the complexes of two different proteins (RNase H from *E. coli* and PriA DNA helicase from *K. pneumonia*) with the same ten-residue peptide derived from the C-terminus of the single-stranded DNA binding protein (SSB). Because CAPRI stated that only the most C-terminal residues of the peptide were resolved in the complex x-ray structure, we also considered shorter peptides in our calculations. Because no information regarding the binding sites was available, our strategy was to use ZDOCK and RCF analysis using several rigid structures of the peptide. The manual selection of the submitted predictions was guided by the RCF results.

Unbound structures were available for the proteins of both targets (PDB entry 2RN2²³ for Target 65, and a CAPRI supplied the structure for Target 66). For the peptide we identified three fragments in the PDB with identical or near-identical sequences and a carboxy terminal group. Two fragments only had the four C-terminal residues (PDB IDs 3C94 and 3UF7), and one covered the entire peptide sequence (PDB ID 2XO8). The complex structures of both targets are now available in the PDB (4Z0U and 4NL8²⁴). The peptide models we used for the RCF analysis and docking overlapped well with the peptide conformations in these complex structures.

For Target 65, the RCF results were consistent among the three peptide structures we used, identifying a single binding pocket. Encouraged by this, we selected ZDOCK predictions according to the binding pocket determined by RCF. However, from the bound structure we see that RCF identified the wrong binding site, resulting in the *incorrect* assessment for all of our submissions.

For Target 66, the RCF analysis identified two sites, which we denoted as ‘major’ and ‘minor’ based on the signal strength. We chose six predictions for the major site, two for the minor site, and two based on the location of a DNA binding site.^{25, 26} The major site identified by RCF agreed with the bound complex structure; nevertheless, all of our predictions were still classified as *incorrect*, because the ligand RMSD exceeded the cutoff.

Target 67

This target was a complex between the WW3 domain of NEDD4, an E3 ubiquitin-protein ligase, and a synthetic 13-mer peptide containing the first PPXY motif (RPEAPPSYAEVVT) of ARDC3, an adapter protein that plays a role in regulating cell-surface expression of G-protein coupled receptors. We modeled the peptide using Modeller based on the structure of a related peptide bound to NEDD4 (the PPXY peptide from PDB ID 2M3O²⁷; NMR model 1), and fitted this peptide model to the CAPRI provided unbound NEDD4 WW3 domain based on the PPXY peptide position in 2M3O. We also performed homology modeling and fitting using another complex structure (the PPXY peptide bound to Ebola VP30, PDB ID 2KQ0, NMR model 1). These two models were used as input to Rosetta FlexPepDock²² with peptide refinement (“-pep_refine” flag), and each run generated 1200 models. From the 2400 FlexPepDock models, we selected forty based on ZRANK2 score,¹² FlexPepDock score, minimal RMSD between the modeled and the bound structures of the Tyr residue in 2M3O, and minimal RMSD between the modeled and the bound structures of the Tyr residue in 2KQ0 (ten models selected per criterion). These forty models were reviewed individually and ten non-redundant models were manually selected from this set (four by 2M3O Tyr RMSD, three by 2KQ0 Tyr RMSD, two by ZRANK score, and one by FlexPepDock score).

The evaluation of these models indicated that our strategy was largely successful. Evaluated using the central peptide residues (sequence “PPSY”), nine models were *acceptable*, and one model was of *medium accuracy*, with a ligand RMSD of 1.93 Å. The *medium accuracy* prediction is shown in Figure 1 compared with the crystallized complex structure, highlighting the correct positioning of the PPSY backbone and side chains. Evaluated using the entire peptide, five of our submitted models were rated *acceptable*; the slightly lower performance judged by this metric appeared to be due to insufficient modeling of the peptide regions outside of the PPSY core, in particular the turn following the tyrosine (Figure 1). In line with our results for Targets 60–64, this suggests that more extensive conformational sampling of peptides may improve results.

Target 95

Target 95 was a complex of the Ring1B/Bmi1/UbcH5c PRC1 ubiquitylation module, an enzyme, bound to the nucleosome core particle. The enzyme transfers the carboxy terminus of ubiquitin (not present in the target complex) from the UbcH5c catalytic cysteine residue to lysine 119 of histone H2A in the nucleosome core particle. The unbound structures were available for this target (PDB IDs 3RPG²⁸ and 3LZ0²⁹). Because the latest version of ZDOCK was parametrized for proteins only, and the nucleosome core particle includes DNA, we used ZDOCK 2.3 augmented with the nucleic acid parameters from Fanelli et al.³⁰

We assumed that for successful ubiquitin transfer the target lysine and the UbcH5c catalytic cysteine residue needed to be in proximity; thus, we filtered the 54,000 ZDOCK predictions using a distance cutoff of 8 Å between Lys118-eN of histone H2A (we used residue 118 as lysine 119 was not present in the unbound structure) and Cys85-γS of UbcH5c. The 41 retained predictions were manually inspected and reduced to ten for submission, which all were evaluated as *incorrect* by the CAPRI team. The target complex structure has been released (PDB ID 4R8P³¹), which allowed us to analyze our performance. First, the conformational changes between the bound and unbound structures were modest, which made this target suitable for rigid body algorithms such as ZDOCK. Second, the distance between the catalytic residue and target lysine is about 8 Å in the solved complex structure, thus our filtering was appropriate. Indeed, upon closer inspection of the 41 predictions retained after filtering, we identified four that closely resembled the solved complex structure and would have resulted in correct predictions. Therefore we conclude that our docking strategy for this target was sound, but we failed to select the correct predictions for final submission.

Targets 96 and 97

Targets 96 and 97 were complexes of the beta-barrel protein eGFP with the engineered alpha-repeat proteins binder eGFP A and binder eGFP C, respectively. The unbound structure was available for eGFP (PDB ID 1JBZ³²). Unbound structures or templates with the correct number of repeats were not available for binder eGFP A (eight repeats) or binder eGFP C (five repeats), but a structure with six repeats was available (PDB ID 3LTJ³³). We therefore used the superposition feature of PyMOL (www.pymol.org) to extend this structure for binder eGFP A, and shorten it for binder eGFP C. The resulting structures were then used as templates for Modeller to perform homology modeling.

RCF analysis for both targets predicted interfaces primarily at the concave side of the alpha-repeat proteins. We noticed that the residues on the concave side of the two proteins were variable among the repeats, whereas residues at the convex side were identical for each repeat. We predicted that the concave side was engineered for binding eGFP. RCF analysis of eGFP showed that predicted interfaces were localized on one half of the barrel. The complex structures have since been released (PDB IDs 4XL5 and 4XVP for Targets 96 and 97, respectively), and agree with our RCF results.

For Target 96 we selected ZDOCK predictions that agree with the RCF analysis, with eGFP binding to the concave side of the alpha-repeat protein and the axis of the beta barrel in parallel with the direction of the alpha helices. The convex surface of the beta barrel and the concave side of the alpha repeat protein were highly complementary and produced a large interface. The bound structure, however, shows that the axis of the eGFP barrel is perpendicular to the direction of the alpha helices, and we had only *incorrect* predictions.

For Target 97 we followed the same RCF guided approach. The experimental structure shows that the barrel is close to parallel with the alpha helices, and indeed we had two *medium accuracy* predictions (Figure 2). In fact, one of the correct predictions corresponded to the top ranked ZDOCK prediction, and the other to the prediction with the second highest clustering density.

Targets 98–101

The deubiquitinating enzyme UCH-L5 can be inhibited and activated by regulatory proteins INO80G and RPN13, respectively. Targets 98–101 were UCH-L5 with (targets 99 and 100) or without (targets 98 and 101) covalently attached ubiquitin-propargyl, bound to a fragment of RPN13 (targets 98 and 99) or two different constructs of INO80G (targets 100 and 101).

All four bound structures have been released after the round was held (4UEM, 4UEL, 4UF6, 4UF5,³⁴ for targets 98–101, respectively), and show that RPN13 clamps the C-terminal end of UCH-L5 (Figure 3). The presence of ubiquitin-propargyl did not affect the RPN13–UCH-L5 binding mode, although the C-terminal helices of UCH-L5 were somewhat shifted with respect to the N-terminal domain. Also the two constructs of INO80G displayed identical binding modes, which also resembled the binding mode of RPN13, but for Target 101 caused larger shifts of the C-terminal helices (Figure 3).

Our strategies for targets 98 and 99, RPN13-UCH-L5 binding, were based on the observation that RPN13 bound to the KEKE motif at the C-terminal end of UCH-L5.³⁵ Through ZDOCK's blocking feature we restricted binding to the C-terminus, and the bound structures showed that this assumption was correct. The unbound structures for RPN13 and UCH-L5 were available (PDB IDs 2KQZ³⁶ and 3IHR³⁷, respectively), although the C-terminus of UCH-L5 needed to be extended to include the presumed binding site for RPN13. We modeled the missing fragment using the QUARK server³⁸. For targets 99 and 100 the ubiquitin-propargyl was attached to UCH-L5 using ubiquitin bound to another deubiquitinase UCH37 (PDB ID 4I6N³⁹) as a template. Comparison with the bound structures showed that this approach yielded the correct position. The constructs of INO80G were homology modeled using RPN13 as the template, following the work by Sanchez-Pulido et al.⁴⁰ suggested by CAPRI.

All our predictions were evaluated as *incorrect*, which could be traced to the large differences between the bound and unbound structures. For UCH-L5, the differences between the bound structure and the structure we used as unbound were primarily in the region that bound RPN13 and INO80G. The C-terminal helices were connected through two hinges that were correctly positioned in the unbound structure, but the orientations of the helices were rather different between bound and unbound structures. More importantly, the unbound form of RPN13 was much more compact than the bound form, and did not allow the observed clamp style binding to UCH-L5 (Figure 3B). Clearly this is a limitation of our rigid body algorithm. Finally, for Targets 100 and 101, our homology modeled INO80G constructs did not have the correct fold, in addition to being too compact to allow the clamp style binding.

Target 103

This target was the complex of FAT10, a ubiquitin-like modifier, with its conjugating enzyme UBE2Z. We homology modeled UBE2Z using 3CEG⁴¹ as the template. After the CAPRI round the unbound structure of UBE2Z became available (PDB ID 5A4P⁴²), and 3CEG indeed had the correct fold. FAT10 consisted of two domains, with 42% and 30% sequence identities with ubiquitin, and we assumed both domains had this fold. We

homology modeled FAT10 using di-ubiquitin as a template (PDB ID 3U30). We suspected that the linker between the two domains might be flexible, but since ZDOCK is a rigid body algorithm we decided to restrict the orientation of the domains to the di-ubiquitin template. The ZDOCK predictions were filtered using biological information. Aichem et al. reported that the activating site of UBE2Z was Cys188, which bound the carboxy terminus of FAT10 and transferred FAT10 to Lys323 through self-FAT10ylation.^{43, 44} We submitted three predictions in which Cys188 of UBE2Z and Cys160 (the last non-flexible residue along the chain) of FAT10 were within 15 Å, three predictions in which Lys323 of UBE2Z and Cys160 of FAT10 were within 15 Å, and three predictions without distance constraint. In addition, we included one complex generated by superposing UBE2Z and FAT10 onto the complex structure of a homologous E2 ligase and ubiquitin (PDB ID 3UGB⁴⁵).

Despite the available biological information, all our predictions were deemed *incorrect*. Because the complex structure has not yet been released we cannot assess where our strategy failed. Perhaps our rigid-body assumption for FAT10 was not correct, and it will be instructive to see whether we correctly predicted the interface of at least one of the domains.

Targets 104 and 105

Targets 104 and 105 were the complexes of immunity proteins with the pyocin AP41 DNase and the Pyocin S2 Nuclease Domain, respectively. These were easy targets as homologous complexes were available with sequence identities of 46% or better. We used Modeller to build homology models of the component proteins (templates with PDB IDs 1EMV⁴⁶ and 3U43⁴⁷ for targets 104 and 105, respectively), and PyMOL (www.pymol.org) to superimpose these onto the template complex structures. The complex structures are now available (PDB IDs 4UHP⁴⁸ and 4QKO for targets 104 and 105, respectively), and show the same folds and binding modes as the homologous complex structures we used. Indeed, we submitted *high accuracy* predictions for Target 104, and *medium accuracy* predictions for Target 105. An additional challenge was to predict side chain conformations and water positions. For side chain conformations we used Modeller to generate initial structures and then Rosetta for side-chain packing. For water positions we used Rosetta solvation and crystal water from homologous structures (PDB IDs 1EMV⁴⁶ and 3U43⁴⁷). We retained only those crystallographic waters that were within 6 Å of an oxygen or nitrogen atom and associated with a target amino acid residue that was identical in the template structure. According to the assessment by the CAPRI management team, we obtained $f^{wmc}(nat)^{49}$ values of 0.308 (*good*) and 0.588 (*excellent*) for targets 104 and 105, respectively, when using the crystallographic waters of the homologous structures. Using the Rosetta determined water molecules we obtained much lower $f^{wmc}(nat)$ values, which seems to indicate that Rosetta's algorithm is not focused on optimizing bridging water.

Scoring

We submitted predictions for ten scoring targets (no scoring challenges were held for the protein-peptide complexes and Target 95). The four targets of Round 32 had no uploaded structures with at least *acceptable* quality. Out of the remaining six targets, we selected correct structures for four (Table 1). These included the two targets of Round 34 that we homology modeled correctly using template structures of the complex. For the targets that

included a scoring challenge, we either submitted correct predictions in both prediction and scoring challenges, or in neither challenges. One might ask whether we simply selected our own uploaded structures because our docking and ranking performance were aligned, but surprisingly none of our correct scoring predictions was uploaded by ourselves.

Conclusions

Our performance was similar to what we achieved historically if we exclude the very difficult targets 98–101 which had no *acceptable* predictions by any of the participating groups. It is clear that large differences between bound and unbound structures are still a challenge for protein-protein docking algorithms. It is interesting to see that none of the complexes of these rounds would qualify for the protein-protein docking benchmark that we maintain,⁵⁰ as we consider only complexes that have unbound structures for both components available and the entire interface region resolved. This supports the challenging nature of the CAPRI targets.

For all but two targets we used experimental data found in the literature to aid the generation of our predictions. This type of information ranged from highly specific data such as homologous complex structures (targets 104 and 105), to less exact data such as the assumption that the target peptide binds at the same protein site as other peptides (targets 60–64 and 67). For all but one target we could confirm that the experimental data we used agreed with either the released complex structure or our *acceptable* prediction (with the exception of Target 103, for which we neither had a correct prediction nor a released complex structure). However, the classification of the peptides from targets 60–64 based on experimental data in the literature led us to focus our docking effort on the minor NLS binding site, which precluded submission of models of peptides bound to the major NLS binding site. Additionally, as the experimental data were usually of low resolution, despite being correct, the biological information did not always guide the computational algorithms to make a correct prediction. The observation that experimental information was available for nearly all targets and did not conflict with the complex structures shows the importance of effective integration of complex structure prediction algorithms with experimental data,^{51–59} as well as the automated retrieval of such data from the literature as pioneered by Badal et al.⁶⁰

Acknowledgments

This work was funded by the National Institutes of Health grant R01 GM116960 awarded to ZW. We thank the experimentalists for supplying the targets, and the CAPRI committee for organizing these challenges and evaluating the predictions.

References

1. Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJE, Vajda S, Vakser I, Wodak SJ. Critical Assessment of PRedicted Interactions. 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. Hwang H, Vreven T, Pierce BG, Hung J-H, Weng Z. Performance of ZDOCK and ZRANK in CAPRI rounds 13–19. *Proteins*. 2010; 78(15):3104–3110. [PubMed: 20936681]
6. Vreven T, Pierce BG, Hwang H, Weng Z. Performance of ZDOCK in CAPRI rounds 20–26. *Proteins*. 2013; 81:2175–2182. [PubMed: 24123140]
7. Chen R, Weng Z. A novel shape complementarity scoring function for protein-protein docking. *Proteins*. 2003; 51(3):397–408. [PubMed: 12696051]
8. Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. *Proteins*. 2003; 52(1):80–87. [PubMed: 12784371]
9. 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]
10. Pierce B, Weng Z. ZRANK: reranking protein docking predictions with an optimized energy function. *Proteins*. 2007; 67(4):1078–1086. [PubMed: 17373710]
11. Vreven T, Hwang H, Weng Z. Integrating atom-based and residue-based scoring functions for protein-protein docking. *Protein Sci*. 2011; 20(9):1576–1586. [PubMed: 21739500]
12. Pierce B, Weng Z. A combination of rescoring and refinement significantly improves protein docking performance. *Proteins*. 2008; 72(1):270–279. [PubMed: 18214977]
13. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*. 1993; 234(3):779–815. [PubMed: 8254673]
14. Hwang H, Vreven T, Weng Z. Binding interface prediction by combining protein-protein docking results. *Proteins*. 2014; 82(1):57–66.
15. Vreven T, Hwang H, Pierce BG, Weng Z. Prediction of protein-protein binding free energies. *Protein Sci*. 2012; 21(3):396–404. [PubMed: 22238219]
16. Vreven T, Hwang H, Weng Z. Exploring angular distance in protein-protein docking algorithms. *PLoS ONE*. 2013; 8(2):e56645. [PubMed: 23437194]
17. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban Y-EA, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popovi Z, Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D, Bradley P. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Meth Enzymol*. 2011; 487:545–574. [PubMed: 21187238]
18. Fromm SA, Truffault V, Kamenz J, Braun JE, Hoffmann NA, Izaurrealde E, Sprangers R. The structural basis of Edc3- and Scd6-mediated activation of the Dcp1:Dcp2 mRNA decapping complex. *EMBO J*. 2012; 31(2):279–290. [PubMed: 22085934]
19. Kosugi S, Hasebe M, Matsumura N, Takashima H, Miyamoto-Sato E, Tomita M, Yanagawa H. Six classes of nuclear localization signals specific to different binding grooves of importin alpha. *J Biol Chem*. 2009; 284(1):478–485. [PubMed: 19001369]
20. Matsuura Y, Stewart M. Nup50/Npap60 function in nuclear protein import complex disassembly and importin recycling. *EMBO J*. 2005; 24(21):3681–3689. [PubMed: 16222336]
21. Marfori M, Lonhienne TG, Forwood JK, Kobe B. Structural basis of high-affinity nuclear localization signal interactions with importin- α . *Traffic*. 2012; 13(4):532–548. [PubMed: 22248489]
22. Raveh B, London N, Schueler-Furman O. Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins*. 2010; 78(9):2029–2040. [PubMed: 20455260]
23. Katayanagi K, Miyagawa M, Matsushima M, Ishikawa M, Kanaya S, Nakamura H, Ikehara M, Matsuzaki T, Morikawa K. Structural details of ribonuclease H from *Escherichia coli* as refined to an atomic resolution. *J Mol Biol*. 1992; 223(4):1029–1052. [PubMed: 1311386]
24. Bhattacharyya B, George NP, Thurmes TM, Zhou R, Jani N, Wessel SR, Sandler SJ, Ha T, Keck JL. Structural mechanisms of PriA-mediated DNA replication restart. *Proceedings of the National Academy of Sciences*. 2014; 111(4):1373–1378.
25. Kozlov AG, Jezewska MJ, Bujalowski W, Lohman TM. Binding specificity of *Escherichia coli* single-stranded DNA binding protein for the chi subunit of DNA pol III holoenzyme and PriA helicase. *Biochemistry-U.S.* 2010; 49(17):3555–3566.

26. Cadman CJ, McGlynn P. PriA helicase and SSB interact physically and functionally. *Nucleic Acids Research*. 2004; 32(21):6378–6387. [PubMed: 15576682]
27. Bobby R, Medini K, Neudecker P, Lee TV, Brimble MA, McDonald FJ, Lott JS, Dingley AJ. Structure and dynamics of human Nedd4-1 WW3 in complex with the α ENaC PY motif. *Biochim Biophys Acta*. 2013; 1834(8):1632–1641. [PubMed: 23665454]
28. Bentley ML, Corn JE, Dong KC, Phung Q, Cheung TK, Cochran AG. Recognition of UbcH5c and the nucleosome by the Bmi1/Ring1b ubiquitin ligase complex. *EMBO J*. 2011; 30(16):3285–3297. [PubMed: 21772249]
29. Vasudevan D, Chua EYD, Davey CA. Crystal structures of nucleosome core particles containing the “601” strong positioning sequence. *J Mol Biol*. 2010; 403(1):1–10. [PubMed: 20800598]
30. 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]
31. McGinty RK, Henrici RC, Tan S. Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *NATURE*. 2014; 514(7524):591–596. [PubMed: 25355358]
32. Hanson GT, McAnaney TB, Park ES, Rendell MEP, Yarbrough DK, Chu S, Xi L, Boxer SG, Montrose MH, Remington SJ. Green fluorescent protein variants as ratiometric dual emission pH sensors. 1. Structural characterization and preliminary application. *Biochemistry-Us*. 2002; 41(52):15477–15488.
33. Urvoas A, Guellouz A, Valerio-Lepiniec M, Graille M, Durand D, Desravines DC, van Tilbeurgh H, Desmadril M, Minard P. Design, production and molecular structure of a new family of artificial alpha-helical repeat proteins (α Rep) based on thermostable HEAT-like repeats. *J Mol Biol*. 2010; 404(2):307–327. [PubMed: 20887736]
34. Sahtoe DD, van Dijk WJ, Oualid El F, Ekkebus R, Ovaa H, Sixma TK. Mechanism of UCH-L5 activation and inhibition by DEUBAD domains in RPN13 and INO80G. *Mol Cell*. 2015; 57(5):887–900. [PubMed: 25702870]
35. Hamazaki J, Iemura S-I, Natsume T, Yashiroda H, Tanaka K, Murata S. A novel proteasome interacting protein recruits the deubiquitinating enzyme UCH37 to 26S proteasomes. *EMBO J*. 2006; 25(19):4524–4536. [PubMed: 16990800]
36. Chen X, Lee B-H, Finley D, Walters KJ. Structure of proteasome ubiquitin receptor hRpn13 and its activation by the scaffolding protein hRpn2. *Mol Cell*. 2010; 38(3):404–415. [PubMed: 20471946]
37. Burgie SE, Bingman CA, Soni AB, Phillips GN. Structural characterization of human Uch37. *Proteins*. 2012; 80(2):649–654. [PubMed: 21953935]
38. Xu D, Zhang Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins*. 2012; 80(7):1715–1735. [PubMed: 22411565]
39. Morrow ME, Kim M-I, Ronau JA, Sheedlo MJ, White RR, Chaney J, Paul LN, Lill MA, Artavanis-Tsakonas K, Das C. Stabilization of an unusual salt bridge in ubiquitin by the extra C-terminal domain of the proteasome-associated deubiquitinase UCH37 as a mechanism of its exo specificity. *Biochemistry-Us*. 2013; 52(20):3564–3578.
40. Sanchez-Pulido L, Kong L, Ponting CP. A common ancestry for BAP1 and Uch37 regulators. *Bioinformatics*. 2012; 28(15):1953–1956. [PubMed: 22645167]
41. Sheng Y, Hong JH, Doherty R, Srikumar T, Shloush J, Avvakumov GV, Walker JR, Xue S, Neculai D, Wan JW, Kim SK, Arrowsmith CH, Raught B, Dhe-Paganon S. A human ubiquitin conjugating enzyme (E2)-HECT E3 ligase structure-function screen. *Mol Cell Proteomics*. 2012; 11(8):329–341. [PubMed: 22496338]
42. Schelpe J, Monté D, Dewitte F, Sixma TK, Rucktooa P. Structure of UBE2Z Enzyme Provides Functional Insight into Specificity in the FAT10 Protein Conjugation Machinery. *J Biol Chem*. 2016; 291(2):630–639. [PubMed: 26555268]
43. Aiche A, Pelzer C, Lukasiak S, Kalveram B, Sheppard PW, Rani N, Schmidtke G, Groettrup M. USE1 is a bispecific conjugating enzyme for ubiquitin and FAT10, which FAT10ylates itself in cis. *Nat Commun*. 2010; 1:13. [PubMed: 20975683]
44. Aiche A, Catone N, Groettrup M. Investigations into the auto-FAT10ylation of the bispecific E2 conjugating enzyme UBA6-specific E2 enzyme 1. *FEBS J*. 2014; 281(7):1848–1859. [PubMed: 24528925]

45. Page RC, Pruneda JN, Amick J, Klevit RE, Misra S. Structural Insights into the Conformation and Oligomerization of E2 Ubiquitin Conjugates. *Biochemistry-Us*. 2012; 51(20):4175–4187.
46. Kühlmann UC, Pommer AJ, Moore GR, James R, Kleanthous 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]
47. Wojdyla JA, Fleishman SJ, Baker D, Kleanthous C. Structure of the ultra-high-affinity colicin E2 DNase-Im2 complex. *J Mol Biol*. 2012; 417(1–2):79–94. [PubMed: 22306467]
48. Joshi A, Grinter R, Josts I, Chen S, Wojdyla JA, Lowe ED, Kaminska R, Sharp C, McCaughey L, Roszak AW, Cogdell RJ, Byron O, Walker D, Kleanthous C. Structures of the Ultra-High-Affinity Protein-Protein Complexes of Pyocins S2 and AP41 and Their Cognate Immunity Proteins from *Pseudomonas aeruginosa*. *J Mol Biol*. 2015; 427(17):2852–2866. [PubMed: 26215615]
49. Lensink MF, Moal IH, Bates PA, Kastiris PL, Melquiond ASJ, Karaca E, Schmitz C, van Dijk M, Bonvin AMJJ, Eisenstein M, Jiménez-García B, Grosdidier S, Solernou A, Perez-Cano L, Pallara C, Fernandez-Recio J, Xu J, Muthu P, Praneeth Kilambi K, Gray JJ, Grudinin S, Derevyanko G, Mitchell JC, Wieting J, Kanamori E, Tsuchiya Y, Murakami Y, Sarmiento J, Standley DM, Shiota M, Kinoshita K, Nakamura H, Chavent M, Ritchie DW, Park H, Ko J, Lee H, Seok C, Shen Y, Kozakov D, Vajda S, Kundrotas PJ, Vakser IA, Pierce BG, Hwang H, Vreven T, Weng Z, Buch I, Farkash E, Wolfson HJ, Zacharias M, Qin S, Zhou H-X, Huang S-Y, Zou X, Wojdyla JA, Kleanthous C, Wodak SJ. Blind prediction of interfacial water positions in CAPRI. *Proteins*. 2014; 82(4):620–632. [PubMed: 24155158]
50. Vreven T, Moal IH, Vangone A, Pierce BG, Kastiris PL, Torchala M, Chaleil R, Jiménez-García B, Bates PA, Fernandez-Recio J, Bonvin AMJJ, Weng Z. Updates to the Integrated Protein-Protein Interaction Benchmarks: Docking Benchmark Version 5 and Affinity Benchmark Version 2. *J Mol Biol*. 2015; 427(19):3031–3041. [PubMed: 26231283]
51. Li L, Huang Y, Xiao Y. How to use not-always-reliable binding site information in protein-protein docking prediction. *PLoS ONE*. 2013; 8(10):e75936. [PubMed: 24124522]
52. van Ingen H, Bonvin AMJJ. Information-driven modeling of large macromolecular assemblies using NMR data. *J Magn Reson*. 2014; 241:103–114. [PubMed: 24656083]
53. Karaca E, Bonvin AMJJ. On the usefulness of ion-mobility mass spectrometry and SAXS data in scoring docking decoys. *Acta Crystallogr D Biol Crystallogr*. 2013; 69(Pt 5):683–694. [PubMed: 23633577]
54. Esquivel-Rodríguez J, Kihara D. Fitting multimeric protein complexes into electron microscopy maps using 3D Zernike descriptors. *J Phys Chem B*. 2012; 116(23):6854–6861. [PubMed: 22417139]
55. Schneidman-Duhovny D, Rossi A, Avila-Sakar A, Kim SJ, Velázquez-Muriel J, Strop P, Liang H, Krukenberg KA, Liao M, Kim HM, Sobhanifar S, Dötsch V, Rajpal A, Pons J, Agard DA, Cheng Y, Sali A. A method for integrative structure determination of protein-protein complexes. *Bioinformatics*. 2012; 28(24):3282–3289. [PubMed: 23093611]
56. Schmitz C, Bonvin AMJJ. Protein-protein HADDOCK using exclusively pseudocontact shifts. *J Biomol NMR*. 2011; 50(3):263–266. [PubMed: 21626213]
57. Pons C, D'Abramo M, Svergun DI, Orozco M, Bernadó P, Fernandez-Recio J. Structural characterization of protein-protein complexes by integrating computational docking with small-angle scattering data. *J Mol Biol*. 2010; 403(2):217–230. [PubMed: 20804770]
58. Lasker K, Sali A, Wolfson HJ. Determining macromolecular assembly structures by molecular docking and fitting into an electron density map. *Proteins*. 2010; 78(15):3205–3211. [PubMed: 20827723]
59. Ritchie DW, Kozakov D, Vajda S. Accelerating and focusing protein-protein docking correlations using multi-dimensional rotational FFT generating functions. *Bioinformatics*. 2008; 24(17):1865–1873. [PubMed: 18591193]
60. Badal VD, Kundrotas PJ, Vakser IA. Text Mining for Protein Docking. *Plos Comput Biol*. 2015; 11(12):e1004630. [PubMed: 26650466]

Figure 1.

Our best submitted model for Target 67 (NEDD4 WW3 domain/ARRDC3 peptide complex). This model was rated medium accuracy, showing strong agreement between the modeled peptide structure (blue) and the peptide in the crystal structure of the complex (cyan), particularly for the central PPSY residues (side chains shown as sticks). NEDD4 WW3 domains are colored tan and yellow for model and crystal structure, respectively. This and all subsequent figures were produced using PyMOL (www.pymol.org).

Figure 2.
Top ranked by ZDOCK, medium accuracy prediction for Target 97. The five residues with the largest RCF values in each protein are in red.

Figure 3.

A: Bound structures of Targets 98–101, with the UCH-L5 N-terminal domains (red) aligned, and its C-terminal helices of the four structures in different shades of blue (darkest shade for Target 99, lightest shade for Target 101, and the two shades in between for the nearly overlapping targets 98 and 100). For clarity, ubiquitin-propargyl is not shown, bound RPN13 (magenta) is shown only for Target 98 (not for Target 99), and the INO80G constructs are not shown. **B:** Bound (magenta) and unbound (pink) forms of RPN13. Note that the unbound structure is more compact than the bound structure.

Table 1

Summary of the CAPRI rounds and our performance.

Round	Target	Type	Modeling type	Prediction ¹	Scoring ¹
28	59	Protein/protein	Homology/unbound	1	1
	60 ²	Protein/peptide	Unbound/peptide	10/9**	Not held
	61 ²	Protein/peptide	Unbound/peptide	10**	Not held
	62 ²	Protein/peptide	Unbound/peptide	10**	Not held
	63 ²	Protein/peptide	Unbound/peptide	9	Not held
29	64 ²	Protein/peptide	Unbound/peptide	10/9**	Not held
	65	Protein/peptide	Unbound/peptide		Not held
	66	Protein/peptide	Unbound/peptide		Not held
	67 ³	Protein/peptide	Unbound/peptide	10/1**	Not held
	95	Protein-DNA/Protein	Unbound/unbound		Not held
31	96	Protein/protein	Homology/unbound		
	97	Protein/protein	Homology/unbound	2**	1/1**
	98	Protein/protein	Unbound ⁴ /unbound		
32	99	Protein/protein	Unbound ⁴ /unbound		
	100	Protein/protein	Unbound ⁴ /homology		
	101	Protein/protein	Unbound ⁴ /homology		
33	103	Protein/protein	Homology/homology		
34	104	Protein/protein	Complex template	6***	10/7****/3**
	105	Protein/protein	Complex template	6**	10/1****/9**

¹The first number indicates the number of predictions of *acceptable accuracy* or better, with double asterisk the number of predictions of *medium accuracy*, and with triple asterisk the number of predictions of *high accuracy*

²Assessment for minor peptide binding site, 6 peptide residues

³Assessed using PPSY residues

⁴Part of structure required building using a template