Published in final edited form as:

Proteins. 2013 December ; 81(12): 2175–2182. doi:10.1002/prot.24432.

Performance of ZDOCK in CAPRI rounds 20–26

Thom Vreven, Brian G. Pierce, Howook Hwang, and Zhiping Weng*

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

Abstract

We report the performance of our approaches for protein-protein docking and interface analysis in CAPRI rounds 20–26. At the core of our pipeline was the ZDOCK program for rigid-body protein-protein docking. We then reranked the ZDOCK predictions using the ZRANK or IRAD scoring functions, pruned and analyzed energy landscapes using clustering, and analyzed the docking results using our interface prediction approach RCF. When possible, we used biological information from the literature to apply constraints to the search space during or after the ZDOCK runs. For approximately half of the standard docking challenges we made at least one prediction that was acceptable or better. For the scoring challenges we made acceptable or better predictions for all but one target. This indicates that our scoring functions are generally able to select the correct binding mode.

Keywords

Protein-protein docking; Protein-protein interaction; Complexes; ZRANK; IRAD

Introduction

The Critical Assessment of PRedicted Interactions (CAPRI) experiment has been an important driving force for the development of methods for the prediction of protein–protein structures.¹ Our laboratory has participated since the earliest rounds of CAPRI, using the ZDOCK algorithm and refinement scoring functions.^{2–5} In this paper we report our results from rounds 20 to 26, which were held between 2010 and 2012 (Table I).

The ZDOCK program for protein-protein docking was introduced approximately 10 years ago.^{6, 7} ZDOCK uses Fast Fourier Transform (FFT) methods to perform an exhaustive search for potential binding modes of two component proteins. The proteins are kept rigid, and the six-dimensional (6D) conformation space is separated into 3 translational degrees of freedom (1.2 Å sampling) and three rotational degrees of freedom (6 or 15 degree sampling). For each angle combination only the best scoring translation is kept, resulting in 3,600 or 54,000 predictions for 15 degree or 6 degree angular sampling, respectively. Various versions of the ZDOCK algorithm were released, and the latest uses the IFACE statistical potential that was developed in our lab as well.⁸ In addition to ZDOCK we

*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.

developed a series of scoring functions for reranking docking results,^{9–11} protein design,¹² and protein-protein binding free energy prediction.¹³

Since our most recent report on the performance of ZDOCK in the CAPRI experiment,⁵ we have published several algorithmic developments related to protein-protein docking. First, we released a new version of the ZDOCK program, ZDOCK 3.0.2, which uses a new 3D convolution library.¹⁴ Instead of using a 3D FFT, we now use a series of 1D FFTs, which allows the calculation of empty grid points to be avoided. In combination with additional performance improvements, ZDOCK 3.0.2 represents a 5-fold speedup over our previous version ZDOCK 3.0, with exactly the same accuracy.

We also developed a hybrid-resolution approach that led to further speedup.¹⁵ We first perform a 15 degree angular sampling docking run. The top 400 (about 10%) predictions are selected, and we carry out a 6 degree angular sampling run only for those angle combinations that are within 10 degree of any of the 400 predictions. This approach reduces the search space dramatically, requiring only 10% of the full 6 degree sampling. Because the angular distances can be precomputed and do not add to the computational cost, we obtain a 6-fold speedup compared with the standard approach. The hit rates and success rates of the full hybrid-resolution and standard approaches are, however, virtually indistinguishable. ZDOCK 3.0.2 and the hybrid-resolution can be applied simultaneously, resulting in a 30-fold speedup over the standard ZDOCK 3.0 algorithm. We also tested the use of angular distance for clustering docking solutions and the exploration of energy landscapes, and found that results are the same as obtained using RMSD distances, but with significant speedup.¹³

The use of FFT in docking requires the scoring functions to be expressed in the form of correlations, which are therefore more limiting than the scoring functions of other molecular modeling methods. This is partially resolved by re-ranking the rigid-body docking solutions using more sophisticated scoring functions. Recently we developed the IRAD function,¹⁰ with the novel aspect that it combined atom-based and residue-based potentials. The residue terms allow for some degree of softness in the potential and reduce the noise in the function as fewer interactions are computed. We showed that the success rate was typically improved with about 10% over the ZDOCK rigid-body docking results.

Finally, we developed the Residue Contact Frequency (RCF) method.¹⁶ RCF uses a collection of ZDOCK docking results to predict interfacial residues, which can subsequently be used to guide experiments or, as we applied in CAPRI, select probable complex structures out of the large sets of predictions. Interface prediction by docking methods have been presented by others as well.^{17–19} In contrast with docking-based algorithms, interfaces can also be predicted based on the properties of the monomers alone,²⁰ and both docking-based and monomer-based interface prediction approaches have been used in CAPRI.^{21, 22} We found that the performance of docking-based and monomer-based approaches is very similar. A novel aspect of our work was the combination of RCF with monomer-based interface prediction methods through a support vector machine, which significantly improved the overall performance.

Methods

Here we outline our general strategy for generating CAPRI predictions. Any deviations from this strategy for specific targets are described in the corresponding sections in the Results. At the core of our approach is the ZDOCK program for rigid-body docking with the IFACE statistical potential. We perform a search in the 6D rotational and translational space, which can either be exhaustive, or knowledge-based, e.g, we may block some residues from being at the binding interface. We use 6 degree angular sampling, which results in 54,000 solutions for each docking run. Using a clustering/pruning approach we remove redundancies from the collection of results,^{5, 13} and also identify regions with a high density of predictions that are therefore likely to contain the correct solution. We generally choose the prediction with the highest ZDOCK score to represent a cluster.

From the reduced list we use the ZDOCK score and prediction density to make a selection of structures to inspect manually. In the manual step we usually are guided by RCF to select the final models for submission. The 10 selected models were then refined using Rosetta or other molecular modeling algorithms to remove possible clashes between the component proteins.

One of the main efforts in our lab is the development of scoring functions, of which several were used for the targets discussed in this report. Our original docking algorithm ZDOCK was based on shape complementarity⁶ and later expanded by adding electrostatic interactions.⁷ These functions can be considered physics-based, but over the years it became apparent that accurate predictions can also be made using functions that are knowledge-based and use experimentally observed propensities. Physics-based and knowledge-based approaches can be combined—we developed the knowledge-based IFACE potential and combined it with electrostatics and shape complementarity in our docking algorithm.⁸ Fast docking algorithms such as ZDOCK, however, are restricted in the form and complexity of the scoring functions that can be used. To alleviate this problem we use a two-stage procedure in which ZDOCK predictions are re-scored using a more detailed scoring function ZRANK.⁹ Both ZDOCK and ZRANK combine physics-based and knowledge-based terms and are atom-based. The past decade showed the promise of course-grain methods that represent amino acids by only a few interaction sites. Just like we combined physics-based and knowledge-based potentials, we also combined atom-based and course-grain terms, resulting in the IRAD scoring function for re-ranking.¹⁰ Using ZRANK and IRAD as basis, we also developed several scoring functions for specific purposes other than re-ranking rigid-body docking results. For example, we developed the ZRANK2 potential to score predictions refined by Rosetta,¹¹ ZAPP for the prediction of binding free energies,¹³ and ZAFFI for the effect of point mutations on binding free energies.¹²

Results and Discussion

Targets 43 and 44

Round 20, with targets 43 and 44, was a scoring-only challenge on protein-protein complexes designed by David Baker and co-workers. Target 43 consisted of 20 models (decoys) of designed complexes that were shown not to bind and a crystal structure of a

complex that did bind, with the goal of identifying the crystal complex. Target 44 consisted of 21 designed complexes and the goal was to predict which of the complex(es) could bind.

We have developed a series of scoring functions for various purposes. For example ZRANK⁹ and IRAD¹⁰ were designed for ranking rigid body docking solutions while allowing for minor clashes between the binding partners, and ZAFFI¹² was designed for predicting the effect of point mutations on binding free energies. We also developed ZRANK 2.0,¹¹ specifically for ranking complex structures that were structurally refined using the Rosetta²³ program, and generally do not contain clashes. Because it is likely that Rosetta was used in the design process and we could observe that the structures were free of clashes, we used our scoring function for refined structures in this round.

Using ZRANK 2.0, we ranked the binding complexes the 8th and the 6th (both out of 21 structures) for targets 43 and 44, respectively. This was reasonably successful, especially considering that the ZRANK scoring function was developed to rank complexes that consisted of the same proteins in different docked orientations, while the designed complexes involved a variety of proteins. Interestingly, in a ‘postmortem’ analysis, we found that a scoring function consisting of only long-range electrostatic interactions ranked the binding complex second and first for target 43 and 44, respectively. This raises the question whether exercises like these truly determine the quality of functions for binding prediction, or merely identify the functions that best remedy the deficiencies of the original design process.

Target 45

Target 45 was a continuation of the scoring round 20 involving targets 43 and 44, and the findings have been published separately as a community paper.²⁴ David Baker and co-workers provided 87 designed protein complexes, and the participants were asked to rank their binding affinities along with 120 complexes from the protein-protein docking benchmark version 3.²⁵ This round was held prior to the development of our energy function ZAPP¹³ that was specifically optimized for binding free energies, thus we used our ZRANK scoring function for reranking docking solutions. Performance was assessed by the area under the ROC curves that discriminate native complexes from designed (non-binding) complexes, and we ranked 18th out of 28 participating groups. Some properties of the designed and benchmark sets, however, differed significantly. For example, the interface sizes of the designed complexes were smaller than typically observed in the benchmark set. This may be related to the design process. When we (in retrospect) applied various components of the ZRANK function that correlated well with interface size, we achieved much better performance. Such findings again raise the question whether good performance in this exercise indicated a generally applicable function, or a function that happens to complement the design process well.

Target 46

Target 46 involved the complex of the methyltransferase Mtb2 and activator Trm112. The Mtb2-Trm112 complex methylates a glutamine residue of the eukaryotic eRF1 transcription factor in *Cuniculi*.²⁶ This target entailed homology modeling of both proteins prior to

docking (homology-homology docking). We built Mtg2 and Trm112 based on the given sequence alignments and PDB templates 1T43 and 2J6A, respectively, using the Accelrys Insight II software package. After homology modeling we prepicked side chains using RosettaDock.²³ Available biological information in the literature was limited for this target, and we only blocked the C-terminus and N-terminus residues of each protein during docking. After clustering and RCF analysis, we selected the 10 most populated clusters that agreed with RCF results. The 10 selected predictions were then refined using CHARMM.²⁷

Our submission did not include acceptable or better predictions. The crystal structure of the target has been released since this CAPRI round (PDB entry 3Q87),²⁶ and we compared our homology modeled structures with the experimentally obtained bound structures to gauge the accuracy of our docking input structures. We found RMSDs between the modeled and experimental structures of 5.4 Å and 6.1 Å for Trm112 and Mtg2, respectively. Because ZDOCK is a rigid-body docking algorithm and lacks structural refinement it performs best with high quality input structures, and the relatively high structural differences between our homology modeled structures and the experimentally determined structures, possibly contributed to us not making correct predictions for this target.

Target 47

Target 47 entailed the prediction of the complex between the colicin E2 DNase domain with the IM2 immunity protein, including bound interface water molecules, and is thus the first CAPRI challenge for predicting interface water molecules. Our methodology and results for this target are described in more detail elsewhere,²⁸ thus we provide a summary here.

To model the colicin E2/Im2 complex, we performed homology modeling using PHYRE²⁹ and 3D-Jigsaw³⁰ to generate two models of the colicin E2 domain, which were individually fitted to the E9/Im2 complex (PDB entry 2WPT)³¹ and refined using Rosetta and ZRANK as previously described¹¹ to generate 300 predictions per homology model. These 600 refined predictions were pooled and selected based on their ZRANK score for submission. Prior to submission, we added water molecules to the modeled interface by selecting waters from the homologous E9/Im2 interface that were responsible for water-mediated hydrogen bonding between the proteins and would not clash with atoms in the modeled interface (distance ≤ 2.5 Å), resulting in 1 to 6 modeled waters per prediction.

All 10 of our models were evaluated as High quality, including one with the lowest Lrmsd among all groups, and our top-ranked model had lowest interface side chain RMSD (below 1.5 Å) among the submitted predictions of all predictors. The predicted water molecules, as expected based on our simplistic modeling scheme, were rated as “fair” for 6 out of 10 models (the remaining 4 had water assessments below this level), which given the fidelity of our interface side chain predictions indicates that water molecules were not necessary to accurately produce high resolution models of this complex.

Targets 48 and 49

For both target 48 and 49 we predicted the complex formed by T4moC and diiron-hydroxylase toluene 4-monooxygenase. T4moC (Toluene-4-monooxygenase system protein C) is a Riske-type ferredoxin and an electron donor. Diiron-hydroxylase is a multidomain

(alpha, beta and gamma subunits) enzyme that catalyzes the NADH and O₂ dependent hydroxylation of toluene. For both targets, unbound crystal structures of the binding partners were used, and the two targets differed in the source of unbound form of diiron-hydroxylase and the presence of the T4moD effector protein: For target 48, a previously published structure (PDB entry 3DHH) of diiron-hydroxylase with the T4moD effector protein was used,³² as suggested by the CAPRI organizers, whereas an unpublished crystal structure of diiron-hydroxylase without the T4moD effector protein was provided for target 49. We also were informed by the CAPRI organizers that diiron-hydroxylase was a hetero-hexamer, and for target 48 could be created from symmetric units of the ABC chains of 3DHH. We used the PyMol “symmetry mates” function to create the hexamer, and retained the T4moD effector protein. The unpublished crystal structure used for target 49 was already in the hetero-hexameric form. For T4moC the CAPRI organizers suggested the use of a previously published crystal structure (PDB entry 1VM9),³³ for both target 48 and 49.

Literature information indicated that W69 of T4moC is conserved between functionally similar proteins and therefore likely plays a role in the interaction.³⁴ However, RCF analysis did not predict W69 to be in the interface. RCF ranked W69 26 out of 82 surface residues, and the residues with high ranks were clustered on the face of the protein opposite of W69. Still we relied on the literature information and selected only models that had W69 in the interface, and used RCF results only for the diiron-hydroxylase.

None of our predictions for target 48 were evaluated as acceptable or better predictions. For target 49, our best model was evaluated as acceptable, with an interface RMSD of 3.59 Å. We achieved different results for the two targets, despite the high similarity of the diiron-hydroxylase models we used (RMSD = 0.35 Å). We therefore speculate that the presence of the T4moD effector protein affected the docking results. Until the target structure is released, however, we can not fully understand the results.

Target 50

Target 50 featured a de novo protein interaction between a small designed protein (HB36.3) and influenza hemagglutinin. For the initial-stage docking, we used MODELLER³⁵ to generate a homology model using the structure 1U84 (a *Geobacillus stearothermophilus* protein; unpublished, Kim et al., Midwest Center for Structural Genomics) and the alignment provided by the CAPRI organizers. We used the top-scoring model based on MODELLER's DOPE score, and removed the terminal amino acids, which were not present in the template structure, as they would likely be flexible and possibly clash during the rigid-body docking procedure. The side chains and backbone were relaxed using Rosetta.³⁶ To generate an alternative model, we used Rosetta's “prepack” protocol to relax the side chains only. The first and last residues of these models were blocked to avoid predictions with the truncated termini buried in the interface. We docked the homology models to the unbound crystal structure of hemagglutinin (PDB entry 3GBN, chains A and B)³⁷. To avoid docking predictions to the homo-trimeric interface of hemagglutinin, we blocked all hemagglutinin residues having at least two contacts (within 5.0 Å) with the other hemagglutinin subunits in the biological unit. The docking results were clustered, and twelve cluster representatives were selected that partially or fully overlapped with the CR6261 antibody interface.³⁷ These

twelve models were refined using RosettaDock and ZRANK as described previously,¹¹ pooled with the original unrefined models and ranked using ZRANK for selection of submitted models.

Based on CAPRI evaluation, we submitted three models of acceptable quality. In Figure 1 we show the top model (based on L_rmsd), which was the refined model from the top-ranked ZDOCK prediction (using the ligand generated using prepacking rather than minimization), highlighting the power of ZDOCK's scoring function to select near-native poses from global docking.

Target 51

Target 51 involved Xylanase Cthe_2193, which consists of six connected modules: GH5, CBM6, CBM13, Fn3, CBM62, and dockerin. The target structure was missing the dockerin module, and CBM62 was not included in the assessment as it was mobile in the target X-ray structure. The structure of the GH5-CBM6 pair was solved independently by the contributors and provided, and the unbound structure of the Fn3 module was previously published (PDB code 3MPC)³⁸. We homology modeled the structure of CBM13 using MODELLER³⁵. Although the CAPRI organizers recommended the lectin-like xylan binding domain from *Streptomyces lividans* xylanase 10A (PDB code 1KNL)³⁹ as the template for CBM13, we used a template (PDB entry 2VSE)⁴⁰ that we considered more reliable as it aligned to a greater portion of the CBM13 sequence. The challenge was therefore reduced to finding the binding modes of three fragments: GH5-CBM6, CBM13, and Fn3.

The docking approaches we have developed do not include methods that are explicitly designed for the prediction of multimeric complexes. We therefore used a stepwise approach. We first docked GH5-CBM6 with CBM13. We selected two fragment structures that seemed reasonable according to our general approach. Furthermore, the distances between the C-terminus of GH5-CBM6 and the N-terminus of CBM13 were 9 and 15 Å for the two solutions, which were possible given the length of the linker.

We then docked Fn3 to the two different GH5-CBM6-CBM13 fragments (picking 5 models for each of the two GH5-CBM6-CBM13 fragments), again applying distance constraints to match the length of the linker between the CBM13 module and the Fn3 module. The interfaces between the modules were assessed independently, so that even a partially correct model would qualify as a success according to the CAPRI criteria, yet none of our predictions turned out to be correct. There are several possible causes. For example, it appeared to be a difficult target, with only a few participants making correct predictions. Also the quality of the homology model we used was questionable. In addition, our stepwise docking approach is not as sophisticated as approaches dedicated for multimeric docking.^{41, 42}

Target 53

Target 53 involved the complex of the alpha repeat protein rep2 that was designed to bind a second alpha repeat protein rep4 (PDB entry 3LTJ).⁴³ We obtained the structure of rep2 by homology modeling using Rosetta³⁶ with 3LTJ⁴³ as the template. Thus the structures of

rep2 and rep4 are very similar, and only differ in the number of repeats. Comparing the sequences of the repeats in 3LTJ and rep2, it was clear that 12 rep2 residues were subject to mutation in the design process. These residues were all located on the same face of rep2, and we assumed that a large proportion of these residues would be in the interface. Furthermore, RCF results strongly suggested that rep2 was bound to the concave face of rep4. Therefore we selected only models that involved the concave face of rep4, and had at least 10 out of 12 of the designed rep2 residues within 5 Å of rep4.

Target 53 was a relatively easy challenge, with about 20 groups submitting predictions that were acceptable or better. We submitted one medium quality model and two acceptable models. Our best model contained all 12 variable rep2 residues in the interface, indicating that our assumption about the design process was correct.

Target 54

Target 54 entailed modeling a de novo designed interaction between a variant of neocarzinostatin and a designed alpha repeat protein (rep16). We first used Rosetta to generate a homology model of rep16 using an existing unbound structure (PDB entry 3LTJ), by modeling and packing 13 point mutations on the template structure with a fixed backbone⁴⁴. ZDOCK was then used to dock this model to the unbound neocarzinostatin structure (PDB entry 2CBO)⁴⁵, and results were clustered, followed by selection of models based on the existence of designed rep16 residues in the predicted binding interface. This yielded 20 models, which were refined using RosettaDock and ZRANK.¹¹ Of these 20 models, 10 refined structures were selected for submission based on ZRANK score and the extent of putative ligand occlusion on neocarzinostatin, while avoiding submission of redundant models. As with the majority of other groups (37 out of 41), all of our 10 submitted models were incorrect.

Targets 55 and 56

Targets 55 and 56 featured a scoring challenge wherein each team provided predictions of binding for point mutations of two proteins designed to interact with influenza hemagglutinin. These predictions were evaluated against high throughput experimental data that provided a readout from systematic mutagenesis of a large number of positions of both designed proteins,⁴⁶ resulting in a sizable and diverse set of mutations to model and predict (1007 mutants for target 55, and 855 mutants for target 56). As our methods and results are presented in detail elsewhere,⁴⁷ we briefly summarize our strategy and performance here.

To predict binding affinities for these targets, we made an adaptation to the ZAFFI scoring algorithm that we previously developed to predict high affinity mutants of the A6 TCR.¹¹ Rather than utilize a separate electrostatics-based filter function as with our original ZAFFI implementation, we integrated hydrogen bonding and Coulombic electrostatics directly into the affinity scoring function. We also used the DFIRE2 scoring function⁴⁸ to identify mutations in non-surface residues for which point substitutions would potentially destabilize folding of the designed protein.

For the fully blind predictions for these two targets, we had considerable agreement with experimental data; the Kendall tau rank order correlation between our predictions and the

experimental measurements ranked the 4th out of 22 groups for target 55, and first out of 20 groups for target 56. Based on our post-prediction analysis of scoring function terms and experimentally measured data for these targets,⁴⁷ shape complementarity (the attractive component of van der Waals) played the largest role in our success, with solvation, pair potential, and electrostatics terms also providing some contributions (differing in extents for the two targets).

Scoring

Each complex structure prediction challenge was followed by a challenge in which a large set of models from various predictor groups were scored (and refined, if desired) using the participants' algorithms of choice. For these challenges, we used the ZRANK⁹ and later IRAD¹⁰ functions developed for reranking docking predictions. Of the 7 scoring challenges we participated in, we achieved a prediction that was at least acceptable for all targets except target 54 (Table I). Since no other group made a correct prediction for this target, we consider our scoring performance to be very good and consistently accurate. In recent years we have indeed focused on the development of accurate scoring functions. These CAPRI results indicate directions for our development to follow, and which of our strengths can particularly benefit the protein-protein interaction field.

Conclusions

In this paper we report the performance of ZDOCK and our other programs for CAPRI rounds 20 to 26. For docking challenges, we made predictions that were acceptable or better for about half of the targets (Table I).

In CAPRI each docking challenge is followed by a scoring challenge. In the current rounds the emphasis on scoring was increased by the introduction of scoring-only targets, where binding affinities or binding versus non-binding were predicted. This is a direction of interest to the protein modeling field, and these targets will likely stimulate further algorithmic improvements. Indeed, we were involved in the development of a protein-protein affinity benchmark,⁴⁹ and developed the ZAPP function for predicting protein-protein binding free energies.¹³ For the scoring challenges that followed the prediction targets, we selected models that were acceptable or better for all but one target. This shows that the scoring has matured much.

Of the six targets that required docking (not counting target 47 as the prediction could be made using homology modeling alone, and counting targets 48 and 49 as a single target), one involved only homology modeled component proteins, one involved only unbound crystal structures, and the remaining four included a homology modeled protein and an unbound crystal structure. It is interesting to see that we made a correct prediction for the unbound-unbound target, no correct prediction for the homology-homology target, and made correct predictions for half of the unbound-homology targets. Although we have to be careful to generalize the results of such a small set of targets, these results suggest that our current rigid-body docking approach is more effective for predicting the complexes of unbound crystal structures than homology models.⁵

We do notice that in the scoring challenges, we made acceptable predictions for most of the targets even though we did not make predictions of high or medium quality (Table I). That indicates that our scoring functions are able to determine whether a prediction is in the vicinity of the native structure, but less able to recover the high-resolution features. A likely reason is that ZRANK and IRAD are optimized for use with rigid-body docking algorithms, thus some softness is incorporated in their terms to account for conformational changes upon complex formation. In the future, we may further develop scoring functions and algorithms dedicated to structural refinement.

Acknowledgments

This work was funded by the National Institutes of Health grant R01 GM084884 awarded to ZW. We thank the experimentalists who supplied the targets and the CAPRI committee for organizing these challenges and for evaluating submitted 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. Chen R, Weng Z. A novel shape complementarity scoring function for protein-protein docking. *Proteins*. 2003; 51(3):397–408. [PubMed: 12696051]
7. Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. *Proteins*. 2003; 52(1):80–87. [PubMed: 12784371]
8. 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]
9. Pierce B, Weng Z. ZRANK: reranking protein docking predictions with an optimized energy function. *Proteins*. 2007; 67(4):1078–1086. [PubMed: 17373710]
10. 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]
11. Pierce B, Weng Z. A combination of rescoring and refinement significantly improves protein docking performance. *Proteins*. 2008; 72(1):270–279. [PubMed: 18214977]
12. Haidar JN, Pierce B, Yu Y, Tong W, Li M, Weng Z. Structure-based design of a T-cell receptor leads to nearly 100-fold improvement in binding affinity for pepMHC. *Proteins*. 2009; 74(4):948–960. [PubMed: 18767161]
13. Vreven T, Hwang H, Pierce BG, Weng Z. Prediction of protein-protein binding free energies. *Protein Sci*. 2012; 21(3):396–404. [PubMed: 22238219]
14. Pierce BG, Hourai Y, Weng Z. Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS ONE*. 2011; 6(9):e24657. [PubMed: 21949741]
15. Vreven T, Hwang H, Weng Z. Exploring angular distance in protein-protein docking algorithms. *PLoS ONE*. 2013; 8(2):e56645. [PubMed: 23437194]
16. Hwang, H.; Vreven, T.; Weng, Z. Binding Interface Prediction by Combining Protein-protein Docking Results. Submitted

17. De Vries SJ, Bonvin AMJJ. CPORT: a consensus interface predictor and its performance in prediction-driven docking with HADDOCK. *PLoS ONE*. 2011; 6(3):e17695. [PubMed: 21464987]
18. Fernandez-Recio J, Totrov M, Abagyan R. Identification of protein-protein interaction sites from docking energy landscapes. *J Mol Biol*. 2004; 335(3):843–865. [PubMed: 14687579]
19. Oliva R, Vangone A, Cavallo L. Ranking multiple docking solutions based on the conservation of inter-residue contacts. *Proteins*. 2013
20. Zhou H-X, Shan Y. Prediction of protein interaction sites from sequence profile and residue neighbor list. *Proteins*. 2001; 44(3):336–343. [PubMed: 11455607]
21. Qin S, Zhou H-X. A holistic approach to protein docking. *Proteins*. 2007; 69(4):743–749. [PubMed: 17803232]
22. van Dijk ADJ, de Vries SJ, Dominguez C, Chen H, Zhou H-X, Bonvin AMJJ. Data-driven docking: HADDOCK's adventures in CAPRI. *Proteins*. 2005; 60(2):232–238. [PubMed: 15981252]
23. Gray J, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl C, 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]
24. Fleishman SJ, Whitehead TA, Strauch E-M, Corn JE, Qin S, Zhou H-X, Mitchell JC, Demerdash ONA, Takeda-Shitaka M, Terashi G, Moal IH, Li X, Bates PA, Zacharias M, Park H, Ko J-S, Lee H, Seok C, Bourquard T, Bernauer J, Poupon A, Azé J, Soner S, Ovali SK, Ozbek P, Tal NB, Haliloglu T, Hwang H, Vreven T, Pierce BG, Weng Z, Perez-Cano L, Pons C, Fernandez-Recio J, Jiang F, Yang F, Gong X, Cao L, Xu X, Liu B, Wang P, Li C, Wang C, Robert CH, Guharoy M, Liu S, Huang Y, Li L, Guo D, Chen Y, Xiao Y, London N, Itzhaki Z, Schueler-Furman O, Inbar Y, Potapov V, Cohen M, Schreiber G, Tsuchiya Y, Kanamori E, Standley DM, Nakamura H, Kinoshita K, Driggers CM, Hall RG, Morgan JL, Hsu VL, Zhan J, Yang Y, Zhou Y, Kastitis PL, Bonvin AMJJ, Zhang W, Camacho CJ, Kilambi KP, Sircar A, Gray JJ, Ohue M, Uchikoga N, Matsuzaki Y, Ishida T, Akiyama Y, Khashan R, Bush S, Fouches D, Tropsha A, Esquivel-Rodríguez J, Kihara D, Stranges PB, Jacak R, Kuhlman B, Huang S-Y, Zou X, Wodak SJ, Janin J, Baker D. Community-wide assessment of protein-interface modeling suggests improvements to design methodology. *J Mol Biol*. 2011; 414(2):289–302. [PubMed: 22001016]
25. Hwang H, Pierce B, Mintseris J, Janin J, Weng Z. Protein-protein docking benchmark version 3.0. *Proteins*. 2008; 73(3):705–709. [PubMed: 18491384]
26. Liger D, Mora L, Lazar N, Figaro S, Henri J, Scrima N, Buckingham RH, van Tilbeurgh H, Heurgué-Hamard V, Graille M. Mechanism of activation of methyltransferases involved in translation by the Trm112 “hub” protein. *Nucleic Acids Res*. 2011; 39(14):6249–6259. [PubMed: 21478168]
27. Brooks BR, Brooks CL, Mackerell AD, 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, Hodoseck M, Im W, Kucsera 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]
28. Lensink MF, Moal IH, Bates PA, Kastitis PL, Melquiond ASJ, Karaca E, Schmitz C, van Dijk M, Bonvin AMJJ, Eisenstein M, Jimenez-Garcia B, Grosdidier S, Solernou A, Prez-Cano L, Pallara C, Fernandez-Recio J, Xu J, Muthu P, Kilambi KP, Gray JJ, Grudinin S, Derevyanko G, Mitchell JC, Wieting J, Kanamori E, Tsuchiya Y, Murakami Y, Sarmiento J, Standley DM, Shirota M, Kinoshita K, Nakamura H, Chavent M, Ritchie DW, Park H, Ko J, Lee H, Seok C, Shen Y, Vajda S, Kundrotas PJ, Vakser IA, Pierce BG, Hwang H, Vreven T, Weng Z, Buch I, Farkash E, Wolfson HJ, Zacharias M, Zhou H-X, Huang S-Y, Zou X, Wojdyla J, Kleanthous C, Wodak SJ. Blind prediction of interfacial water positions in CAPRI. In preparation.
29. Kelley LA, Sternberg MJE. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc*. 2009; 4(3):363–371. [PubMed: 19247286]
30. Bates PA, Kelley LA, MacCallum RM, Sternberg MJ. Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins*. 2001; (Suppl 5):39–46. [PubMed: 11835480]

31. Meenan NAG, Sharma A, Fleishman SJ, Macdonald CJ, 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. *P Natl Acad Sci Usa*. 2010; 107(22):10080–10085.
32. Bailey LJ, McCoy JG, Phillips GN, Fox BG. Structural consequences of effector protein complex formation in a diiron hydroxylase. *P Natl Acad Sci Usa*. 2008; 105(49):19194–19198.
33. Moe LA, Bingman CA, Wesenberg GE, Phillips GN, Fox BG. Structure of T4moC, the Rieske-type ferredoxin component of toluene 4-monooxygenase. *Acta Crystallogr D Biol Crystallogr*. 2006; 62(Pt 5):476–482. [PubMed: 16627939]
34. Elsen NL, Moe LA, McMartin LA, Fox BG. Redox and functional analysis of the Rieske ferredoxin component of the toluene 4-monooxygenase. *Biochemistry-Us*. 2007; 46(4):976–986.
35. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*. 1993; 234(3):779–815. [PubMed: 8254673]
36. 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, Popov3 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]
37. Ekiert DC, Bhabha G, Elsliger M-A, Friesen RHE, Jongeneelen M, Throsby M, Goudsmit J, Wilson IA. Antibody recognition of a highly conserved influenza virus epitope. *Science*. 2009; 324(5924):246–251. [PubMed: 19251591]
38. Alahuhta M, Xu Q, Brunecky R, Adney WS, Ding S-Y, Himmel ME, Lunin VV. Structure of a fibronectin type III-like module from *Clostridium thermocellum*. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2010; 66(Pt 8):878–880.
39. Notenboom V, Boraston AB, Williams SJ, Kilburn DG, Rose DR. High-resolution crystal structures of the lectin-like xylan binding domain from *Streptomyces lividans* xylanase 10A with bound substrates reveal a novel mode of xylan binding. *Biochemistry-Us*. 2002; 41(13):4246–4254.
40. Treiber N, Reinert DJ, Carpusca I, Aktories K, Schulz GE. Structure and mode of action of a mosquitocidal holotoxin. *J Mol Biol*. 2008; 381(1):150–159. [PubMed: 18586267]
41. Inbar Y, Benyamini H, Nussinov R, Wolfson H. Prediction of multimolecular assemblies by multiple docking. *J Mol Biol*. 2005; 349(2):435–447. [PubMed: 15890207]
42. Karaca E, Melquiond ASJ, De Vries SJ, Kastiris PL, Bonvin AMJJ. Building macromolecular assemblies by information-driven docking: introducing the HADDOCK multibody docking server. *Mol Cell Proteomics*. 2010; 9(8):1784–1794. [PubMed: 20305088]
43. 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]
44. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein-protein complexes. *P Natl Acad Sci Usa*. 2002; 99(22):14116–14121.
45. Drevelle A, Graille M, Heyd B, Sorel I, Ulryck N, Pecorari F, Desmadril M, van Tilbeurgh H, Minard P. Structures of in vitro evolved binding sites on neocarzinostatin scaffold reveal unanticipated evolutionary pathways. *J Mol Biol*. 2006; 358(2):455–471. [PubMed: 16529771]
46. Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch E-M, Wilson IA, Baker D. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science*. 2011; 332(6031):816–821. [PubMed: 21566186]
47. Moretti R, Fleishman SJ, Agius R, Torchala M, Bates P, Kastiris PL, Rodrigues JP, Trellet M, Bonvin AMJJ, Cui M, Rooman M, Gilles D, Dehouck Y, Moal IH, Romero-Durana M, Perez-Cano L, Pallara C, Jimenez B, Fernandez-Recio J, Flores S, Pacella M, Kilambi KP, Gray JJ, Popov P, Grudin S, Esquivel-Rodríguez J, Kihara D, Zhao N, Korkin D, Zhu X, Demerdash ONA, Mitchell JC, Kanamori E, Tsuchiya Y, Nakamura H, Lee H, Park H, Seok C, Sarmiento J, Liang S, Teraguchi S, Standley DM, Shimoyama H, Terashi G, Takeda-Shitaka M, Iwade M, Umeyama H, Beglov D, Hall DR, Kozakov D, Vajda S, Pierce BG, Hwang H, Vreven T, Weng Z,

Huang Y, Li H, Yang X, Ji Z, Liu S, Xiao Y, Zacharias M, Qin S, Zhou H-X, Huang S-Y, Zou X, Velankar S, Janin J, Wodak SJ, Baker D. Community-wide evaluation of methods for predicting the effect of mutations on protein-protein interactions. Submitted.

48. Yang Y, Zhou Y. Ab initio folding of terminal segments with secondary structures reveals the fine difference between two closely related all-atom statistical energy functions. *Protein Sci.* 2008; 17(7):1212–1219. [PubMed: 18469178]
49. Kastiris PL, Moal IH, Hwang H, Weng Z, Bates PA, Bonvin AMJJ, Janin J. A structure-based benchmark for protein-protein binding affinity. *Protein Sci.* 2011; 20(3):482–491. [PubMed: 21213247]

Figure 1.

Submitted model for target 50 (influenza hemagglutinin/HB36.3 designed protein). HB36.3 from the crystal structure is shown in magenta, and our modeled HB36.3 is slate. The hemagglutinin of our model (not shown) was superposed to hemagglutinin from the crystal structure (3R2X), shown in green (HA1) and cyan (HA2), while the other units from the trimeric biological assembly are shown in gray. This model had a 2.14 Å interface backbone RMSD from the crystal structure and was rated “acceptable”.

Table I

Summary of the CAPRI rounds and our performance. Rounds 20 and 21 (targets 43–45) and targets 55 and 56 were scoring-only and are not listed. Round 25 was cancelled.

Round	Target	Type	Prediction	Scoring
22	46	Homology/homology	0	3
23	47	Homology ¹	10/10***	9/3**/6***
	48	Unbound/unbound	0	n/a ³
	49	Unbound/unbound	1	1
24	50	Unbound/homology	3	1
	51	Unbound/homology ²	0	3 ⁴
26	53	Unbound/homology	2/1**	2/1**
	54	Unbound/homology	0	0

¹ A template for the entire complex was available and docking was not needed.

² Docking three linked domains, of which two were unbound and one was homology modeled.

³ Scoring combined with target 49.

⁴ The three possible interfaces were assessed separately; our acceptable models were for CBM13-Fn3.