PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Score_set: A CAPRI benchmark for scoring protein complexes

Marc F. Lensink[1]* and Shoshana J. Wodak[2,3,4]

[1] CNRS USR3078, University Lille North of France, Parc de la Haute Borne, F-59658 Villeneuve d'Ascq, France

[2] Structural Biology Program, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada

[3] Department of Biochemistry, University of Toronto, Ontario, Canada

[4] Department of Molecular Genetics, University of Toronto, Ontario, Canada

ABSTRACT

Critical Assessment of PRedicted Interactions (CAPRI) has proven to be a catalyst for the development of docking algorithms. An essential step in docking is the scoring of predicted binding modes in order to identify stable complexes. In 2005, CAPRI introduced the scoring experiment, where upon completion of a prediction round, a larger set of models predicted by different groups and comprising both correct and incorrect binding modes, is made available to all participants for testing new scoring functions independently from docking calculations. Here we present an expanded benchmark data set for testing scoring functions, which comprises the consolidated ensemble of predicted complexes made available in the CAPRI scoring experiment since its inception. This consolidated scoring benchmark contains predicted complexes for 15 published CAPRI targets. These targets were subjected to 23 CAPRI assessments, due to existence of multiple binding modes for some targets. The benchmark contains more than 19,000 protein complexes. About 10% of the complexes represent docking predictions of acceptable quality or better, the remainder represent incorrect solutions (decoys). The benchmark set contains models predicted by 47 different predictor groups including web servers, which use different docking and scoring procedures, and is arguably as diverse as one may expect, representing the state of the art in protein docking. The data set is publicly available at the following URL: http://cb.iri.univ-lille1.fr/Users/lensink/Score_set.

Proteins 2014; 82:3163–3169.

Key words: scoring function; protein-protein docking; protein-protein interaction; computational docking; benchmark set; Capri.

INTRODUCTION

Protein-protein interactions (PPI) are ubiquitous and essential to cellular function. PPI are the result of a physical interaction between two protein molecules and the detailed atomic-level prediction of such interaction is the topic of computational protein docking. The Critical Assessment of PRedicted Interactions (CAPRI) experiment has been instrumental in catalyzing the development of computational docking algorithms since its inception in 2001.[1,2] Modeled after Critical Assessment of protein Structure Prediction (CASP), CAPRI participants are invited to predict the three-dimensional structure of a protein complex prior to its publication. The predictions are assessed by an independent assessor team, from which the identity of the predictor groups is concealed. CAPRI counts 29 docking rounds so far, with a total of 65 targets, including diverse systems such as enzyme-inhibitor complexes, multi-domain single protein chains, and complexes of proteins with peptides, polysaccharides or nucleic acids.

Computational docking involves a number of independent steps.[3,4] These typically include rigid-body docking, filtering and ranking complexes based on empirical or knowledge-based force-fields, as well as atomic-level refinement. Various approaches to account for backbone or side-chain flexibility are also used, although not systematically.[5–7] At any of these steps,

Grant sponsors: The Canadian Institutes for Health Research (to S.J.W.) and French Agence Nationale de Recherche (to M.F.L.); Grant number: ANR-12-BSV5-0009-01.

Shoshana J. Wodak's current address is VIB Department of Structural Biology, VUB Building E, Pleinlaan 2, B-1050 Brussels, Belgium

*Correspondence to: Marc Lensink, Interdisciplinary Research Institute, USR3078 CNRS, University Lille1, Villeneuve d'Ascq, France.

E-mail: marc.lensink@iri.univ-lille1.fr

Received 26 June 2014; Revised 5 August 2014; Accepted 22 August 2014

Published online 1 September 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24678

biological information may be used to guide or filter candidate solutions. The final step always involves *scoring* the resulting models.[8] At this stage scoring aims not only at discriminating between non-native and native-like solutions, but also attempts to rank the native-like models in a manner that correlates with their relative quality and possibly also their binding affinity. CAPRI participants are allowed to submit a total of 10 predicted models per target and it has long been observed that the best-scoring models are not necessarily the ones of highest quality as judged by the CAPRI assessment criteria.

Recognizing the importance of meaningful scoring and ranking of complexes, many of the groups participating in CAPRI have devoted much attention to the development of scoring functions and some even make it their prime objective. Scoring functions use various criteria, usually formulated as a set of energy terms, to evaluate the binding interface of the predicted complex and to rank the models based on the consolidated value of these terms.[9] The energy terms may represent physics-based residue or atom pair potentials,[10,11] or a statistical simplification thereof.[12] However, in order to account for the complex interplay between interactions at the interface, multi-body, empirical or knowledge-based potentials[13–15] are increasingly used. Novel approaches use multi-body residue interactions to construct the ligand-receptor interaction networks and evaluate them.[16–19] Scoring functions were also shown to be effective in filtering models during various intermediate steps of the docking procedure.[20]

In order to test newly developed scoring functions, developers require good-quality benchmark sets, representing both specific and non-specific association modes between two proteins. Several benchmark sets of experimentally determined protein complexes, representing specific interactions or *true* examples, have been developed by members of the CAPRI community[21,22] and more recently these sets have been complemented with data on binding affinities.[23] These benchmarks have been widely used in the development of docking and scoring functions.[24,25] But so far developers needed to generate their own *false* examples (non-specific binding solutions), which they usually produce using their own or only a few other docking algorithms. This can be especially problematic when the same function is used to score intermediate and final docking solutions. Thus, although these private benchmarks are helpful, they are not diverse and generic enough, prompting the demand for larger and more realistic benchmark sets of docking models comprising both true and false examples.

Considering the importance of scoring functions in protein docking, the CAPRI experiment has been extended in 2005 to include a scoring challenge, designed to help developers test their scoring functions independently of the docking calculations.[26] Approximately 20 scorer groups participate on a regular basis in this challenge (in comparison to about 50 docking groups). For a given target, scorers are provided with a randomized and anonymized ensemble of models contributed on a voluntary basis by groups participating in the docking experiment, at the rate of 100 models per group. This typically yields a total of 1400 predicted models. Scorers are invited to submit the 10 highest-ranking models using the scoring method of their choice. The ranked models, as well as all contributed models are evaluated using the same criteria as those used for the CAPRI docking submissions for the corresponding Round and target.

The benchmark set described here represents a consolidation of contributed models for different targets that were made available on a restricted basis to groups that participated in the CAPRI scoring challenge so far. It currently includes atomic coordinates of predicted models for 15 CAPRI targets whose coordinates have been deposited in the PDB as well as the values for the various criteria used in CAPRI to assess the quality of each model. This enables developers to examine correlations between their score and the different established quality assessment criteria to help them identify shortcomings in their scoring functions and suggest further improvements.

METHODS

The Models

For any given target, as many as 40 different predictor groups may contribute models to the randomized set (in practice this number lies between 10 and 20). This introduces a sizable diversity of models that is desired, but also poses many technical problems. These include, but are not limited to, inconsistencies in the ordering of atoms, missing residues, different residue numbering, the addition (or not) of hydrogen atoms and the naming of these, the proper inclusion of TER records to indicate chain termination in the PDB file. While these problems may be relatively easily solved, more serious artifacts may arise from the docking of a homolog, used as template, instead of the modeled target structure, or the use of the wrong oligomer. These problems may hamper proper use of the benchmark, especially by groups new to the field. We have therefore re-processed all the entries to the benchmark set to produce models whose conventions and format are as standardized as possible.

Having defined the receptor and ligand chains in the target, we match every single chain in the predicted model (source) to those (target). A minimum sequence overlap of 70% in every chain of both entities (receptor and ligand) is required for a model to be retained. When a model presents too few chains and the length of one or several of these is significantly longer than the target chains, we attempt to split the chains in the model by performing global sequence alignments in order to define the sequence range that matches a target chain. Any

Table I
Illustrative Assessment Quantities for the First Six Models of T47

N^o	Clashes[a]	f_{nat}	L-rms	I-rms	S-rms	Classification
0	13/73	0.0179	35.1838	9.7769	10.7788	Incorrect
1	12/73	0.0000	38.4628	11.2163	12.3318	Incorrect
2	265/73	0.1071	19.0578	8.1879	9.1692	Clashes, incorrect
3	52/73	0.2321	10.7764	5.0532	5.7315	Incorrect
4	13/73	0.8393	1.4064	0.7006	1.8428	High
5	11/73	0.7500	2.8481	1.1723	2.1049	Medium

Additional quantities $f_{non\text{-}nat}$, p_{seqid}, and θ_L and d_L are included in the benchmark set, but not listed here. We refer to the text for a description of the quantities. Classification follows the criteria as shown in Table II.
[a]The clash threshold $n_{clash}^{threshold}$ for T47 was 73.

model matching the number of target chains with enough sequence overlap in any one of these chains is retained. Those that do not match the required number of receptor and ligand chains, or that do not show enough sequence overlap for any one of the target chains, are discarded. Also discarded are the surplus chains in the models, not present in the target. This may for instance be the case when a template comprising a dimer was used to build a target receptor, but only one of the monomers was found to bind the ligand.

The retained models are written to file after copying chain identifiers as they exist in the target structure. Residue atoms and names are written in a consistent order and convention as defined by the PDB format; hydrogen atoms are ignored. Proper chain termination records are written. The final set of models is written using the NMR multi-model format.

Model quality annotations

Each model in the benchmark is annotated with the CAPRI quality assessment criteria. These criteria are based on a rigorous comparison to the targets structures in each CAPRI Round as illustrated in Table I.[26–28] The criteria include the quantities f_{nat} and $f_{non\text{-}nat}$, representing respectively the fraction of receptor-ligand residue contacts in the target structure that is reproduced in the model and the fraction of contacts in the model that is not present in the target (over-predicted contacts). The global geometry of the model is assessed by calculating the root mean square displacement (rmsd) of the backbone atoms of the ligand (L-rms), and the misorientation angle θ_L and residual displacement d_L of the ligand geometric center after the receptor in the model and the target have been optimally superimposed.[29]

In addition to the fraction of native contacts f_{nat}, two other quantities are used to assess the quality of the predicted interface, namely, the rmsd of interface backbone (I-rms) and interface side-chain (S-rms) atoms, after superimposing the interface residues of the two complexes (see e.g. Refs. 30 and 31 for details). For the evaluation of residue-residue contacts a distance threshold of 5 Å between atoms is used. Interface residues are defined using a 10 Å distance threshold. In order to avoid artificially high f_{nat} values, models with too many atomic clashes (n_{clash}) are disqualified. A clash is defined as any atom-atom contact below 3 Å. Also disqualified are models that show insufficient sequence overlap with the target (p_{seqid}). Disqualified models are retained in the benchmark set, but annotated as such. However, a threshold $p_{seqid}^{threshold}$ of 70% was employed in the creation of the benchmark set, and models not meeting this criterion are discarded.

Lastly, the benchmark also provides the global ranking of each model following the CAPRI convention, into incorrect, acceptable, medium and high quality models (Table II and Refs. 28 and 30). Ignoring models that are rejected due to too many clashes or too little overlap in sequence, only f_{nat}, L-rms and I-rms are used to derive the overall CAPRI ranking of models.

RESULTS

A full list of the targets included in the benchmark set is provided in Table III. Only targets whose structures have been released to the PDB are included. Table III lists for every target its receptor and ligand entities. The convention in CAPRI is that docking is performed between only two molecular entities, each of which may consist of several polymer or oligomer chains. The entities are denoted respectively as *receptor* and *ligand*, with the receptor being defined as the larger entity. A global sequence alignment is used to define which residue chains in the target belong to the receptor and which are to be considered as the ligand. All the models in the benchmark are annotated with the values of the various CAPRI quality assessment criteria, and their global ranking, as described in Methods.

In total the benchmark set contains 19,013 predicted complexes for 15 published CAPRI targets. These complexes were contributed by a total of 47 groups, including both fully automatic web servers as well as expert human predictors. On average 13 groups contributed models for each target. About 11% of these models represent a prediction of acceptable quality or better, as witnessed from the assessment statistics summarized in

Table II
CAPRI Assessment Criteria

Rejected ("clashes" or "low_id")	$n_{clash} > n_{clash}^{threshold}$ OR $p_{seqid} < p_{seqid}^{threshold}$				
Incorrect	$f_{nat} < 0.1$	OR	L-rms > 10.0	AND	I-rms > 4.0
Acceptable	$f_{nat} \geq 0.3$	AND	L-rms > 5.0	AND	I-rms > 2.0
OR	$0.1 \leq f_{nat} < 0.3$	AND	(L-rms ≤ 10.0	OR	I-rms ≤ 4.0)
Medium	$f_{nat} \geq 0.5$	AND	L-rms > 1.0	AND	I-rms > 1.0
OR	$0.3 \leq f_{nat} < 0.5$	AND	(L-rms ≤ 5.0	OR	I-rms ≤ 2.0)
High	$f_{nat} \geq 0.5$	AND	L-rms ≤ 1.0	AND	I-rms ≤ 1.0

The criteria are applied from top to bottom.[27] See the text for a more detailed description.

Table IV. The complexes are written on a per-target basis to a multi-model PDB file. A basic description of every model (number of chains, number of residues in every chain, identification of receptor and ligand entities, sequence alignment to target chains) is given in two separate files. A third file lists the assessment results for every model. These include values for the quantities n_{clash} and p_{seqid}, f_{nat} and $f_{non\text{-}nat}$, L-rms, I-rms and S-rms, and θ_L and d_L, together with the global ranking of each model as per the CAPRI criteria outlined in Table II. For

Table III
Target Table

T29	Trm8/Trm82 tRNA guanine-N(7)-methyltransferase				
	Receptor	2VDU:D	TRM82_YEAST	2VDU:D	Bound
	Ligand	2VDU:F	TRM8_YEAST	2VDV:F	Unbound
T30	Rnd1-GTP bound to the RBD dimer				
	Receptor	2REX:B	RND1_HUMAN	2CLS:A	Unbound
	Ligand	2REX:A	PLXB1_HUMAN	2R2O:A	Unbound
T32	Protease savinase and the bifunctional inhibitor BASI				
	Receptor	3BX1:A	SUBS_BACLE	1SVN:A	Unbound
	Ligand	3BX1:C	IAAS_HORVU	1AVA:C	Unbound
T35,	BM22-1-GH10 modules of Clostridium Xyn10B xylanase				
T36	Receptor	2W5F:A	XYNY_CLOTM	1N82:A	Unbound
	Ligand	2W5F:A	XYNY_CLOTM	1DYO:A	Unbound
T37	Arf-effector complex				
	Receptor	2W83:A	ARF6_HUMAN	2A5D:A	Unbound
	Ligand	2W83:CD	JIP4_HUMAN	2ZTA:AB	Unbound
T38,	Centaurin-α1/FHA domain				
T39	Receptor	3FM8:C	ADAP1_HUMAN	3FEH:A	Unbound
	Ligand	3FM8:A	KI13B_HUMAN	2G1L:A	Unbound
T40	Bovine trypsin and protease inhibitor (API-A)				
	Receptor	3E8L:A	TRY1_BOVIN	1BTY:A	Unbound
	Ligand	3E8L:B/C	Q7M1P4_SAGSA	3E8L:B/C	Bound
T41	Colicin E9 DNase and Im2 immunity protein				
	Receptor	2WPT:B	CEA9_ECOLX	1FSJ:B	Unbound
	Ligand	2WPT:A	IMM2_ECOLX	2NO8:1	Unbound
T46	Methyl transferase Mtq2/Trm112				
	Receptor	3Q87:B	Q8SRR4_ENCCU	1P91:A	Unbound
	Ligand	3Q87:A	Q8SUP0_ENCCU	2J6A:A	Unbound
T47	Colicin E2 DNase and Im2 immunity protein				
	Receptor	3U43:B	CEA2_ECOLX	2WPT:B	Bound[a]
	Ligand	3U43:A	IMM2_ECOLX	2WPT:A	Bound[b]
T50	HB36.3 designed protein/flu hemagglutinin				
	Receptor	3R2X:AB	HEMA_I18A0	3GBN:AB	Unbound
	Ligand	3R2X:C	Designed	1U84:A	Unbound
T53	Designed Rep4/Rep2 α-repeat complex				
	Receptor	4JW2:B	Designed	3LTJ:A	Unbound
	Ligand	4JW2:A	Designed	3LTJ:A	Unbound
T54	Designed neocarzinostatin/Rep16 α-repeat complex				
	Receptor	4JW3:A	Designed	3LTJ:A	Unbound
	Ligand	4JW3:C	NCZS_STRML	2CBO:A	Unbound

For every target, PDB ID's and sequence identifiers for receptor and ligand entities are listed, followed by their (un)bound template PDB ID's. References for the targets are as follows: T29: Ref. 32; T30: Ref. 33; T32: Ref. 34; T35, T36: Ref. 35; T37: Ref. 36; T40: Ref. 37; T41: Ref. 38; T46: Ref. 39; T47: Ref. 40; T50: Ref. 41; T53, T54: Ref. 42; T38, T39: unpublished.
[a]Bound homolog.
[b]Bound to a homolog partner.

Table IV
Decoy Set Statistics

Target	Total	High	Medium	Acceptable	Incorrect
T29	2083	2	78	87	1916
T30	1343	0	0	2	1341
T32	599	0	3	12	584
T35	499	0	0	3	496
T36	309	0	0	0	309
T37	1500	11	46	42	1401
T38	899	0	0	0	899
T39	1400	0	3	1	1396
T40	2180	193	206	189	1592
T41	1200	2	120	249	829
T46	1699	0	0	24	1675
T47	1051	278	307	26	440
T50	1451	0	36	97	1318
T53	1400	0	17	113	1270
T54	1400	0	1	18	1381

Every decoy is classified following CAPRI criteria as either incorrect, or of acceptable, medium or high quality. The total number of models for every category is listed, with their total in column 2.

interested parties, the unprocessed multi-model PDB files, which comprise the presented benchmark plus all rejected models (408 in total), are also provided.

DISCUSSION

Scoring functions are employed in order to discriminate between native-like and non-native association modes, and ideally should also be able to rank native-like complexes in a manner that correlates with their binding affinity. The discriminatory role of scoring functions is particularly important during the intermediate stages of docking, where they are used to single out likely productive binding modes for further processing and refinement. Tight integration of scoring with docking and refinement was shown to lead to superior docking predictions.[20,43] In general however, the performance of available scoring function remains limited. Recently, a community-wide assessment of the ability of scoring function to estimate binding affinities was performed on a set of designed protein complexes. Although some promising results were obtained, the correlation between the estimated affinities based on scoring function and the experimental measures was in general poor, suggesting that there was much room for improvement even for the best performing functions.[44,45]

Developing improved function for scoring protein-protein complexes therefore remains a very important research goal. Achieving this goal requires the availability of adequate benchmark data sets for the development and systematic testing of new or improved scoring schemes. Existing protein docking benchmark sets[22,23] cannot fully serve this purpose since they comprise only experimental structures that serve as positive examples, which should score highly, but they lack negative examples, representing non-native association modes that should in principle score poorly. In a recent evaluation of scoring functions,[25] this limitation was overcome by using the SwarmDock server[46] to generate a limited number of putative binding modes using the complexes from the Docking Benchmark 4.0.[22] But the set of binding modes contained models produced by one docking server, and was not made publicly available.

The scoring benchmark described in this report comprises the ensemble of predicted protein complexes that were made available on a discretionary basis to participants of the CAPRI scoring challenge since 2005. This ensemble is made publicly available here for the first time. It includes both *positive* and *negative* examples, produced by automatic web servers and expert human predictors using a wide range of methods.

About 11% of the models in the benchmark are of acceptable quality or better, as evaluated by the CAPRI quality assessment criteria. This percentage is significantly higher than for the ensemble of complexes that usually needs to be scored in protein docking calculations. A single rigid-body docking trial can produce as many as tens of thousands of structures, only a small percentage of which represents native-like complexes. Thus, although the probability of singling out native-like models from the benchmark based on chance alone is only about 1 in 10, it is much higher than in docking calculations, as already noted in several CAPRI scoring assessments.[26–28] This not withstanding, the benchmark still represents a challenging test set for the scoring functions. The native-like models in the benchmark are not perfect and differ from the experimentally determined structures (the targets) in various ways, making it sometime more difficult to recognize them as positive examples.

The atomic models together with their assessed quality as per the CAPRI quality assessment criteria can be used in a number of ways. Computed scores for individual models can be correlated with values for any of the quality assessment criteria deemed meaningful. Distributions of score values as a function of one or more of quality criteria can be analyzed to derive useful trends. The positive and negative examples of the benchmark can be used to train classifiers capable of distinguishing between the two type of examples based on various features, using machine learning techniques.[47] The benchmark can also be used for the development and testing of methods for predicting interface residues, the group of residues on each protein that contribute to the interface of a native binding mode. Information on such residues can be obtained from the analysis of both native-like and non-native docking models, because it is not uncommon for a protein to have sticky surface patches with a higher than average propensity to form both specific and non-specific interactions. A recent analysis of submitted docking models for 20 CAPRI targets showed indeed that

almost 25% of the interfaces in submitted docking models ranked as incorrect had a good overlap with the native interface.[48]

The benchmark set is available at the following URL: http://cb.iri.univ-lille1.fr/Users/lensink/Score_set. It will be updated with models of new targets from the CAPRI experiment as soon as the corresponding structures are released, and should grow steadily. In the future we may also consider including a larger number of negative examples to mimic a yet more realistic environment for testing scoring function. This will not affect information on models for targets already included in the data set nor their quality assessment.

ACKNOWLEDGMENTS

The authors express gratitude to the ensemble of CAPRI predictor groups who have made their models available, and to the scorer groups who have made the scoring challenge a success. Additional thanks also go to Dima Kozakov and members of the CAPRI Management Committee, Alexandre Bonvin, Michael Sternberg, Sandor Vajda, Ilya Vakser, Sameer Velankar and Zhiping Weng, for critical reading of the article.

REFERENCES

1. Janin J, Henrick K, Moult J, ten Eyck L, Sternberg MJE, Vajda S, Vakser I, Wodak SJ. CAPRI: a Critical Assessment of PRedicted Interactions. Proteins 2003;52:2–9.
2. Janin J. Protein-protein docking tested in blind predictions: the CAPRI experiment. Mol Biosyst 2010;6:2351–2362.
3. Vakser IA, Kundrotas P. Predicting 3D structures of protein-protein complexes. Curr Pharm Biotechnol 2008;9:57–66.
4. Vajda S, Kozakov D. Convergence and combination of methods in protein-protein docking. Curr Opin Struct Biol 2009;19:164–170.
5. Lensink MF, Méndez R. Recognition-induced conformational changes in protein-protein docking. Curr Pharm Biotechnol 2008;9:77–86.
6. Moreira IS, Fernandez PA, Ramos MJ. Protein-protein docking dealing with the unknown. J Comput Chem 2010;31:317–342.
7. Zacharias M. Accounting for conformational changes during protein-protein docking. Curr Opin Struct Biol 2010;20:180–186.
8. Moal IH, Moretti R, Baker D, Fernández-Recio J. Scoring functions for protein-protein interactions. Curr Opin Struct Biol 2013;23:862–867.
9. Huang SY, Grinter SZ, Zou X. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. Phys Chem Chem Phys 2010;12:12899–12908.
10. Vreven T, Hwang H, Weng Z. Integrating atom-based and residue-based scoring functions for protein-protein docking. Protein Sci 2011;20:1576–1586.
11. Liu J, He X, Zhang JZ. Improving the scoring of protein-ligand binding affinity by including the effects of structural water and electronic polarization. J Chem Inf Model 2013;53:1306–1314.
12. Pons C, Talavera D, de la Cruz X, Orozco M, Fernández-Recio J. Scoring by intermolecular pairwise propensities of exposed residues (SIPPER): a new efficient potential for protein-protein docking. J Chem Inf Model 2011;51:370–377.
13. Khashan R, Zheng W, Tropsha A. Scoring protein interaction decoys using exposed residues (SPIDER): a novel multibody interaction scoring function based on frequent geometric patterns of interfacial residues. Proteins 2012;80:2207–2217.
14. Krüger DM, Ignacio Garzón J, Chacón P, Gohlke H. DrugScore(PPI) knowledge-based potentials used as scoring and objective function in protein-protein docking. PLoS One 2014;9:e89466.
15. Grinter SZ, Zou X. A Bayesian statistical approach of improving knowledge-based scoring functions for protein-ligand interactions. J Comput Chem 2014;35:932–943.
16. Chang S, Jiao X, Li CH, Gong XQ, Chen WZ, Wang CX. Amino acid network and its scoring application in protein-protein docking. Biophys Chem 2008;134:111–118.
17. Jiao X, Chang S. Scoring function based on weighted residue network. Int J Mol Sci 2011;12:8773–8776.
18. Andreani J, Faure G, Guerois R. InterEvScore: a novel coarse-grained interface scoring function using a multi-body statistical potential coupled to evolution. Bioinformatics 2013;29:1742–1749.
19. Masone D, Grosdidier S. Collective variable driven molecular dynamics to improve protein-protein docking scoring. Comput Biol Chem 2014;49:1–6.
20. Vajda S, Hall DR, Kozakov D. Sampling and scoring: a marriage made in heaven. Proteins 2013;81:1874–1884.
21. Liu S, Gao Y, Vakser IA. DOCKGROUND protein-protein docking decoy set. Bioinformatics 2008;24:2634–2635.
22. Hwang H, Vreven T, Janin J, Weng Z. Protein-protein docking benchmark version 4.0. Proteins 2010;78:3111–3114.
23. Kastritis 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;20:482–491.
24. Vreven T, Hwang H, Pierce BG, Weng Z. Prediction of protein-protein binding free energies. Protein Sci 2012;21:396–404.
25. Moal IH, Torchala M, Bates PA, Fernández-Recio J. The scoring of poses in protein-protein docking: current capabilities and future directions. BMC Bioinformatics 2013;14:286.
26. Lensink MF, Méndez R, Wodak SJ. Docking and scoring protein complexes: CAPRI, 3rd edition. Proteins 2007;69:704–718.
27. Lensink MF, Wodak SJ. Docking and scoring protein interactions: CAPRI 2009. Proteins 2010;78:3073–3084.
28. Lensink MF, Wodak SJ. Docking, scoring, and affinity prediction in CAPRI. Proteins 2013;82:2082–2095.
29. McLachlan AD. Rapid comparison of protein structures. Acta Cryst Sect A 1982;38:871–873.
30. Méndez R, Leplae R, De Maria L, Wodak SJ. Assessment of blind predictions of protein–protein interactions: current status of docking methods. Proteins 2003;52:51–67.
31. Méndez R, Leplae R, Lensink MF, Wodak SJ. Assessment of CAPRI predictions in rounds 3–5 shows progress in docking procedures. Proteins 2005;60:150–169.
32. Leulliot N, Chaillet M, Durand D, Ulryck N, Blondeau K, van Tilbeurgh H. Structure of the yeast tRNA m7G methylation complex. Structure 2008;16:52–61.
33. 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;282:37215–37224.
34. Micheelsen PO, Vevodova J, De Maria L, Ostergaard PR, Friis EP, Wilson K, Skjøt 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;380:681–690.
35. Najmudin S, Pinheiro BA, Prates JA, Gilbert HJ, Romao MJ, Fontes CM. Putting an N-terminal end to the Clostridium thermocellum xylanase Xyn10B story: crystal structure of the CBM22-1-GH10 modules complexed with xylohexaose. J Struct Biol 2010;172:353–362.
36. 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;28:2835–2845.

37. Bao R, Zhou CZ, Jiang C, Lin SX, Chi CW, Chen Y. The ternary structure of the double-headed arrowhead protease inhibitor API-A complexed with two trypsins reveals a novel reactive site conformation. J Biol Chem 2009;284:26676–26684.
38. Meenan NA, 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. Proc Natl Acad Sci USA 2010;107:10080–10085.
39. Liger D, Mora L, Lazar N, Figaro S, Henri J, Scrima N, Buckingham RH, van Tilbeurgh H, Heurgue-Hamard V, Graille M. Mechanism of activation of methyltransferases involved in translation by the Trm112'hub' protein. Nucleic Acids Res 2011;39:6249–6259.
40. Wojdyla JA, Fleishman SJ, Baker D, Kleanthous C. Structure of the ultra-high-affinity colicin E2 DNase–Im2 complex. J Mol Biol 2012;417:79–94.
41. Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch EM, Wilson IA, Baker D. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. Science 2011;332:816–821.
42. Guellouz A, Valerio-Lepiniec M, Urvoas A, Chevrel A, Graille M, Fourati-Kammoun Z, Desmadril M, van Tilbeurgh H, Minard P. Selection of specific protein binders for pre-defined targets from an optimized library of artificial helicoidal repeat proteins (alphaRep). PLoS One 2013;8:e71512.
43. Pierce B, Weng Z. A combination of rescoring and refinement significantly improves protein docking performance. Proteins 2008;72:270–279.
44. Fleishman SJ, Whitehead TA, Strauch EM, Corn JE, Qin S, Zhou HX, Mitchell JC, Demerdash ON, Takeda-Shitaka M, Terashi G, Moal IH, Li X, Bates PA, Zacharias M, Park H, Ko JS, Lee H, Seok C, Bourquard T, Bernauer J, Poupon A, Aze 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, Kastritis PL, Bonvin AM, 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-Rodriguez J, Kihara D, Stranges PB, Jacak R, Kuhlman B, Huang SY, Zou X, Wodak SJ, Janin J, Baker D. Community-wide assessment of protein-interface modeling suggests improvements to design methodology. J Mol Biol 2011;414:289–302.
45. Moretti R, Fleishman SJ, Agius R, Torchala M, Bates PA, Kastritis PL, Rodrigues JP, Trellet M, Bonvin AM, Cui M, Rooman M, Gillis D, Dehouck Y, Moal I, Romero-Durana M, Perez-Cano L, Pallara C, Jimenez B, Fernandez-Recio J, Flores S, Pacella M, Praneeth Kilambi K, Gray JJ, Popov P, Grudinin S, Esquivel-Rodriguez J, Kihara D, Zhao N, Korkin D, Zhu X, Demerdash ON, 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, Iwadate 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 X, Liu S, Xiao Y, Zacharias M, Qin S, Zhou HX, Huang SY, 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. Proteins 2013;81:1980–1987.
46. Torchala M, Moal IH, Chaleil RA, Fernández-Recio J, Bates PA. SwarmDock: a server for flexible protein-protein docking. Bioinformatics 2013;29:807–809.
47. Ravikant DV, Elber R. Energy design for protein-protein interactions. J Chem Phys 2011;135:065102.
48. Lensink MF, Wodak SJ. Blind predictions of protein interfaces by docking calculations in CAPRI. Proteins 2010;78:3085–3095.