PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Docking, scoring, and affinity prediction in CAPRI

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

[1] Interdisciplinary Research Institute, USR3078 CNRS, University Lille North of France, Parc de la Haute Borne, 50 avenue de Halley, F-59658 Villeneuve d'Ascq cedex, France

[2] Structure and Function Program, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada

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

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

ABSTRACT

We present the fifth evaluation of docking and related scoring methods used in the community-wide experiment on the Critical Assessment of Predicted Interactions (CAPRI). The evaluation examined predictions submitted for a total of 15 targets in eight CAPRI rounds held during the years 2010–2012. The targets represented one the most diverse set tackled by the CAPRI community so far. They included only 10 "classical" docking and scoring problems. In one of the classical targets, the new challenge was to predict the position of water molecules in the protein–protein interface. The remaining five targets represented other new challenges that involved estimating the relative binding affinity and the effect of point mutations on the stability of designed and natural protein–protein complexes. Although the 10 classical CAPRI targets included two difficult multicomponent systems, and a protein–oligosaccharide complex with which CAPRI participants had little experience, this evaluation indicates that the performance of docking and scoring methods has remained quite robust. More remarkably, we find that automatic docking servers exhibit a significantly improved performance, with some servers now performing on par with predictions done by humans. The performance of CAPRI participants in the new challenges, briefly reviewed here, was mediocre overall, but some groups did relatively well and their approaches suggested ways of improving methods for designing binders and for estimating the free energies of protein assemblies, which should impact the field of protein modeling and design as a whole.

Proteins 2013; 81:2082–2095.

Key words: protein interactions; docking; blind predictions; CAPRI; interface water molecules; affinity.

INTRODUCTION

Characterizing protein interactions is a crucial step in gaining an understanding of how cells function, as most cellular processes are carried out by physically interacting proteins.[1] Recent technological advances have fostered a rapid growth in protein–protein interaction (PPI) data derived from many small-scale studies, as well as from genome-scale interrogations in several model organisms.[2–6] These data are commonly represented as networks of linked protein components, which describe the interaction landscape of specific cellular processes, particular cell types, or entire organisms. Although this representation has provided some biological insights, it remains very abstract and incomplete. It maps simple connections between proteins and offers no information on the stoichiometry of the interactions, or on their temporal or spatial distributions. Some of this information can be inferred from known structures of protein assemblies and from data on related proteins stored in public databases[7] and efforts have been undertaken over the years to better exploit these data.[8,9] But progress has been slow, because protein assemblies remain poorly represented in the Protein Data Bank (PDB).[10,11]

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: French Agence Nationale de Recherche; Grant number: ANR-12-BSV5-0009-01 (to M.F.L.); Grant sponsors: Canada Institute of Health Research, Sickkids Foundation, Ontario Research Fund, PrioNet Network of Excellence, Canada.

*Correspondence to: Marc F. Lensink, Interdisciplinary Research Institute, USR3078 CNRS, University Lille North of France, Parc de la Haute Borne, 50 avenue de Halley, F-59658 Villeneuve d'Ascq cedex, France. E-mail: marc.lensink@iri.univ-lille1.fr (or) Shoshana J. Wodak, Structure and Function Program, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. E-mail: shoshana.wodak@gmail.com or shoshana@sickkids.ca

Received 16 July 2013; Revised 12 September 2013; Accepted 13 September 2013
Published online 1 October 2013 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24428

Large-scale projects to determine the atomic structures of such assemblies such as the 3D repertoire and SPINE2 (http://www.spine2.eu/SPINE2/) have so far been hampered by the difficulties in obtaining suitable crystals and diffraction data on a large enough scale. On the other hand, efforts to fill the repertoire of individual protein 3D structures have been very successful. Given a newly sequenced protein, the odds are usually high that its 3D structure can be built by homology modeling techniques using known protein structures as templates.[9,12] Structures from this rich repertoire are also increasingly used as templates or scaffolds in protein design projects that have useful medical applications.[13,14] Furthermore, in efforts to build models of larger protein assemblies, information on individual protein structures is being integrated with various other types of data, including low-resolution maps obtained by cryo-electron microscopy, using the so-called hybrid modeling techniques.[15]

Computational approaches play a key role in all these applications. Of particular importance are methods for deriving accurate structural models of multiprotein assemblies, starting from the atomic coordinates of the individual components, the so-called "docking" algorithms, and the associated energetic criteria for singling out stable binding modes.[4,16–18]

The community-wide initiative on the Critical Assessment of Predicted Interactions (CAPRI), established in 2001, has played a central role in fostering the development of better docking algorithms and closely monitoring their performance. In a given Round of CAPRI, individual groups that develop docking procedures are provided with information on the components of a complex, the target, which they use to predict the three-dimensional structure of the interacting proteins. Depending on the target this information may consist of the atomic coordinates of the unbound form of the components, the amino acid sequence of the latter, or the known structures of templates to be used in homology-based modeling. In addition to submitting 10 models for each target, predictors are also invited to upload a set of 100 models. Once all the submissions are completed, the uploaded models are shuffled and made available to all groups as part of the CAPRI scoring experiment. The "scorer" groups are invited to evaluate the ensemble of uploaded models using the scoring function of their choice, and submit their own 10 best ranking ones.

At the end of each Round, all the submitted models are assessed by comparing them with the target: the structure of the complex determined by experimental methods (X-ray diffraction or NMR), which their authors deposit with the CAPRI management before publication. All the predictions are made blindly, with no knowledge of the correct answer. Furthermore, the identity of the predictors is concealed from the independent team that evaluates the predicted models.

Since the inception of CAPRI, 19 prediction Rounds were completed with a total of 42 targets. The results were presented at four Assessment Meetings, held in 2002, 2004, 2007, and 2009, and described in the literature.[19–22]

Here, we present the prediction results for Rounds 20–27 held during the years 2010–2012. These results were discussed at the fifth CAPRI evaluation meeting held April 2013 in Utrecht, The Netherlands. The assessed Rounds comprised a total of 15 targets. Of these only 10 targets represented "classical" docking and scoring challenges. The remaining five involved the estimation of the relative binding affinity of protein–protein complexes, and for one of the classical targets, the new challenge was to predict the position of water molecules in the protein–protein interface. Furthermore, the five affinity prediction targets and three of the docking and scoring targets were complexes with proteins designed by humans. Two other docking targets were challenging multicomponent systems, and one was a protein–polysaccharide complex, completing one of the most diverse set of challenges that CAPRI predictors took up so far.

We present a detailed analysis of the results obtained for the 10 docking targets based on a total of 14,605 evaluated models. Detailed accounts of the water prediction results and for the five affinity prediction targets have been reported in three independent collective reports[13,23,24] and will only be briefly summarized here.

THE TARGETS

A total of 15 targets were evaluated in Rounds 20–27 (See Janin, this issue, for a detailed description). An additional 16th target (T52) was cancelled after a simple Google search had revealed an image of the complex. The novelty in the 15 targets resided in that more than half (eight targets) comprised complexes involving proteins designed by humans rather than evolved by natural selection. Five of these targets (T43–45 and T55–56) represented affinity prediction challenges, an experiment new to CAPRI, initiated by the group of David Baker. The remaining three targets (T50, T53, and T54) were submitted for the docking and scoring experiments. In T53, both components of the complex were designed helical repeat proteins, whereas in T50 and T54 a designed helical protein was bound to a natural target protein. T54 was a difficult target, likely due to the fact, of which predictors were not made aware, that the complex formed as a hetero-tetramer comprising two copies of each component that also interacted with one another.

The seven natural complexes (T46–49, T51, T57, and T58) were quite diverse. Targets T46, T47, and T58 were two-component complexes of the type typically encountered in CAPRI. T46, a complex of the methyltransferase MTq2 with Trm12, an activator protein from *E. cuniculi*, was a difficult case. Predictors had to build models for

the components using rather distant homologs as structural templates and little biochemical information was available on regions of either protein that were involved in binding. On the other hand, T47, a complex between the DNase domain of colicin E2 and the Im2 immunity protein against this colicin, was a very easy target. CAPRI predictors were familiar with this system having previously had as target (T41) the complex of Im2 with the DNase domain of colicin E9. This time however, the challenge was to predict not the structure of the complex, but the position of water molecules at the protein–protein interface, known to play an important role in stabilizing the interface in this family of complexes. T58 was a rather classical enzyme-inhibitor complex between the g-type lysozyme SalG from salmon and the PliG lysozyme inhibitor from *E. coli*, for which unbound structures were available. In addition, predictors were provided with small-angle X-ray scattering (SAXS) data for the complex.

Targets T48/49 and T51 were more challenging multicomponent assemblies. T48/49 was based on the same complex between the T4moH di-iron toluene 4-monooxygenase/hydroxylase with Rieske-type ferrodoxin T4moC. The T4moH forms a hetero-hexamer comprising two trimers $(\alpha,\beta,\gamma)_2$ and the complex with the ferredoxin is part of a larger multicomponent electron transfer system. The difficulty of this system was that the ferredoxin binds to a different subunit on each T4moH trimer, and one of the binding sites competes with that of an effector protein, which was present in one of known T4moH hexamer structures that predictors were invited to use in the docking calculations.

T51 comprised the first four structural domains and a fifth unstructured domain of xylanase Cthe_2193 (Najmudin et al., in preparation). The xylanase is an 885-residue single-chain protein comprising a total of six structural domains. Unbound structures were available for Domains 1 and 2 in tandem, and for Domain 4, whereas Domain 3 had to be built by homology. The challenge was to predict all pair-wise interfaces as well as the structure of the entire assembly.

The last of the natural assemblies was T57, a complex between a fragment of heparin composed of six sugar units, and the bacterial surface protein Bt4661. This target was offered along with the new structure of the free protein. This was the second time after T33, a protein/RNA complex,[25] that CAPRI predictors have been challenged with docking a large flexible nonprotein molecule onto a protein. But dealing with sugar molecules was a first for CAPRI predictors and their docking procedures. To help predictors, the authors of the target structure also provided predictors with an atomic model of the heparin moiety, built with the sugar rings in an arbitrary conformation.

Lastly, as in previous CAPRI rounds, predictors (and scorers) also exploited, whenever possible, biochemical data or information on sequence conservation in related proteins, to identify protein regions involved in the interaction. This information was then used to bias the docking and scoring calculations or to filter solutions. This was usually not the case for most web-servers, which perform the predictions completely automatically.

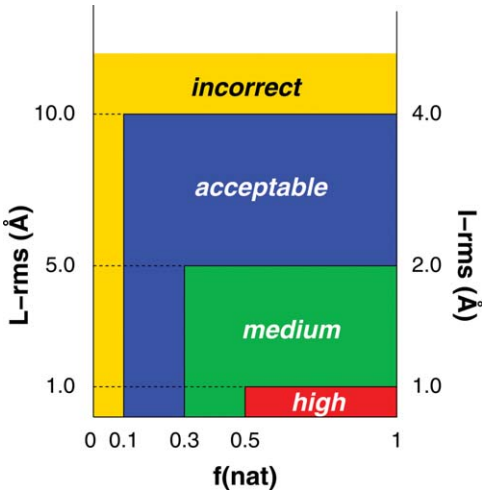

Figure 1

CAPRI criteria for assessing the quality of docking models. Relationship between the CAPRI assessment criteria *f*(nat), L-rms, and I-rms, and the classification into high (***), medium (**), acceptable (*) or incorrect docking models. The quantity *f*(nat) represents the fraction of native ligand-receptor contacts that is recalled in the docking model. The so-called L-rms is the root mean square deviation (RMSD), calculated over the ligand backbone atoms after fitting the receptor entities onto each other. The related I-rms uses the backbone atoms of the ligand and receptor interface residues for both the fit and the calculation. The figure is a simplification and for the application of the criteria we refer to previous CAPRI evaluations.[21,22]

Evaluation protocol for the docking and scoring experiments

The criteria for evaluating the quality of the predicted complexes are summarized in simplified form in Figure 1. For the detailed application of the criteria we refer to previous CAPRI evaluations.[21,22] Submitted models containing a number of interatomic clashes (atoms closer than 3 Å) exceeding a threshold were not evaluated, as such models may retrieve a large number of native interactions simply due to the interpenetration of the subunits. The threshold for the number of accepted clashed was defined as previously described.[21,22]

Evaluation of the interactions with heparin in the BT4661-heparin complex (T57) was nontrivial as we had to deal with inconsistent atom and residue labeling and ordering. For the sake of simplicity we decided to treat the heparin moiety as a single residue and include 60 of the 105 heavy atoms of the molecule in the calculation of ligand and interface RMSDs.

Table I
Number of Predictor, Uploader and Scorer Groups, and Number of Models Submitted by Every Category of Groups for the Individual Targets of CAPRI Rounds 22–27

		Number of groups			Number of models		
Round	Target	Predictors	Uploaders	Scorers	Predictors	Uploaders	Scorers
21	43/44/45			Affinity prediction			
22	46	40	17	16	397	1700	160
23	47	29	12	14	277	1051	128
	48	32	14	—	308	1346	—
	49	33	14	13	308	1350	110
24	50	40	15	17	357	1251	130
	51	46	13	13	305	1000	120
25	52			Cancelled			
26	53	42	15	13	384	1400	130
	54	41	16	13	377	1400	130
	55/56			Affinity prediction			
27	57	31	—	—	256	—	—
	58	23	—	—	230	—	—

Predictors and scorers were allowed to submit at most 10 models. Uploaders could each contribute 100 models for the Scorers to select from. Targets 43–45 and 55/56 refer to the affinity prediction experiments and target 52 was cancelled. For T48, there was no scorer submission, but uploaded models were pooled together with T49. For T57 and T58, there was only a predictor round. The absence of scorer and uploader submissions is indicated by "—."

RESULTS AND DISCUSSION

Prediction results for the docking and scoring experiments

This section describes the results for the classical docking and scoring experiments. It is divided into three parts. The first part describes the prediction performance for the 11 individual targets for which docking and scoring experiments were conducted. In the second part, we present an overview of the results across predictor groups and targets, and in the third part we discuss the performance achieved by the automatics servers.

Table I lists the number of predictor, uploader and scorer groups submitting models for each target, as well as the total number of models submitted for individual targets by each category of groups. The number of models of acceptable quality or higher collectively submitted for the 11 evaluated targets by predictors, scorers, and uploaders are listed in Table II. The best models obtained for each target and their quality measures are summarized in Table III, and the results for individual participants for each target are provided in the Supporting Information (Tables S1–S10). A pictorial illustration of the models collectively predicted for each target is presented in Figure 2.

Table II
Summary of Prediction Performance in CAPRI Rounds 22–27

	RMS	RMS	High			Medium			Acceptable		
	(ligand)	(receptor)	P	U	S	P	U	S	P	U	S
T46	1.34	3.19	0	0	0	0	0	0	13	24	15
T47	0.28	0.48	95	278	63	108	305	38	13	20	0
T48.3	0.35	0.91	0	2	—	11	12	—	39	76	—
T48.6	0.35	0.91	0	0	—	4	6	—	17	43	—
T49.3	0.50	0.83	0	6	0	1	10	1	35	87	19
T49.6	0.50	0.83	0	0	0	1	3	1	11	45	4
T50	0.68	0.78	0	0	0	17	19	14	35	37	21
T51.1	0.35	2.02	0	0	0	0	1	0	3	27	9
T51.2	2.02	0.44	0	0	0	0	0	0	1	2	1
T51.3	2.37	0.44	0	0	0	0	0	0	0	0	0
T53	0.24	0.53	1	0	0	11	9	12	31	92	24
T54	0.31	0.49	0	0	0	0	1	0	6	18	0
T57	0.78	0.23	0	—	—	5	—	—	26	—	—
T58	0.31	0.31	0	—	—	15	—	—	18	—	—

The number of submitted models of acceptable, medium or high accuracy for targets 46–58. Data listed separately for Predictor (P), Uploader (U), and Scorer (S) submissions. Root mean square deviation (RMS, in Å) between bound and unbound structures for ligand and receptor entities are listed in Columns 2 and 3, respectively. For T48 and T49 models submitted were assessed against different versions of the targets, with respectively, the trimer and the hexamer form of the T4moH moiety of the target being referred to as T48.3/T49.3 and T48.6/T49.6. For T51, a single chain xylanase comprising 5 structural domains (GH5, CBM6, CBM13, Fn3, and CBM62) different pairs of components of the assembly were assessed separately. These were CBM13/Fn3 (T51.1), CBM13/CBM6-GH5 (T51.2), and Fn3-CBM13/CBM6-GH5 (T51.3). For T48, there was no scorer submission, but uploaded models were pooled together with T49. For T57 and T58 there was only a predictor round. The absence of scorer or uploader submissions is indicated as "—."

Table III
The Best Quality Measures for Docking Models and for the Reconstituted Model from the Unbound Components, for Individual Targets and the Various Versions of T48/49 and T51

	Best values of submitted models						Reconstituted model		
Target	f_{nat}	Type	L-rms	Type	I-rms	Type	f_{nat}	L-rms	I-rms
T46	0.565	P	6.757	U	3.215	S	0.290	36.965	3.974
T47	0.929	U	0.638	U	0.443	S	0.714	0.649	0.386
T48.3	0.857	U	2.455	U	0.905	U	0.771	0.472	0.556
T48.6	0.625	U	3.600	P	11.526	P	0.708	0.840	0.584
T49.3	0.800	U	2.060	U	0.806	U	0.857	0.524	0.381
T49.6	0.646	U	1.565	U	1.047	U	0.812	0.606	0.613
T50	0.857	P	2.202	U	1.060	U	0.612	0.746	0.362
T51.1	0.483	U	2.625	U	1.954	U	0.200	19.972	0.379
T51.2	0.279	U	10.496	P	3.294	P	0.000	11.800	6.275
T51.3	0.279	U	16.749	U	4.097	P	0.528	1.939	0.200
T53	0.769	U	2.741	S	0.985	P	0.788	0.172	0.351
T54	0.520	U	4.518	U	2.420	U	0.780	0.227	0.331
T57	1.000	P	3.216	P	1.369	P	0.647	3.683	1.678
T58	0.613	P	3.602	P	1.522	P	0.775	0.347	0.285

For each individual measure f_{nat}, L-rms and I-rms, the source group category of the model producing this best value is listed in the column "Type" (P for Predictor, U for Uploader, and S for Scorer). The very high L-rms values for T46 and T51.1/T51.2 are due to the low-sequence identity between target and template structures used in homology modeling (in T51 this was the case for the CBM13 domain), which resulted in a structural misalignment. For T48, there was no scorer submission, but uploaded models were pooled together with T49. For T57 and T58, there was only a predictor round. For T48 and T49 models submitted were assessed against different versions of the targets, with, respectively, the trimer and the hexamer form of the T4moH moiety of the target being referred to as T48.3/T49.3 and T48.6/T49.6. For T51, a single chain xylanase comprising five structural domains (GH5, CBM6, CBM13, Fn3, and CBM62) different pairs of components of the assembly were assessed separately. These were CBM13/Fn3 (T51.1), CBM13/CBM6-GH5 (T51.2), and Fn3-CBM13/CBM6-GH5 (T51.3). For T48, there was no scorer submission, but uploaded models were pooled together with T49. For T57 and T58 there was only a predictor round.

For the two multicomponent assemblies (T48/49 and T51), several separate assessments were carried out. For T48 and T49 docking and scoring predictions were assessed for two different trimer and hexamer geometries, respectively. For T51, predictions for individual pair-wise domain–domain interfaces were assessed separately. In total this represented 14 different target interfaces for which predictions were evaluated. The values of all the quality measures computed for all the submitted models for individual targets and participants can be found on the CAPRI web site (http://www.ebi.ac.uk/msd-srv/capri).

Results for individual targets

Round 22: T46

T46 was the complex of the methyltransferase MTq2 with Trm12, an activator protein from *E. cuniculi* (PDB ID: 3Q87).[26] Both components had to be built by homology using rather distantly related templates (~12 and 20% sequence identify for MTq2 and Trm12, respectively). This yielded many inaccurate models for the unbound structures (3.2 Å rms for MTq2). In addition, there was also little biochemical information available on regions involved in binding, of either protein.

Prediction results were overall poor. Only one predictor group (Bonvin) and the HADDOCK server by the same group submitted a total of 13 acceptable models for this target. Twenty-four acceptable models were among the uploaded structures, with Bonvin and Haddock contributing 11 acceptable models each, and Bates contributing two acceptable models. The successful performance by the server and group of Bonvin is noteworthy. It was due to the use in their data-driven docking method of interaction restraints derived from interdomain contacts in a distant structural homolog (PDB ID: 1P91).[31] This homolog was identified by structural alignments, illustrating recent claims that information on interfaces in distant homologs can successfully inform docking procedures.[32]

The scorer performance was somewhat better, in that eight groups identified at least one acceptable model, and the groups of Kihara and Weng identifying four and three of the models, respectively (see Supporting Information Tables S1 and S11B).

Round 23: T47–49

T47 was the complex between the DNase domain of colicin E2 and the cognate Im2 immunity protein (PDB ID: 3U43),[27] a system with which CAPRI participant were well familiar. This target involved carrying out docking predictions, which we discuss here, as well as the prediction of interface water positions in the docking models, which are reported in Lensink *et al.*[24] and summarized here in a separate section.

As expected, T47 was an easy docking problem, with predictors submitting as many as 95 high quality, 108 medium quality, and 61 models of acceptable quality. Of the 25 groups (including three servers) submitting models, none submitted only acceptable docking solutions.

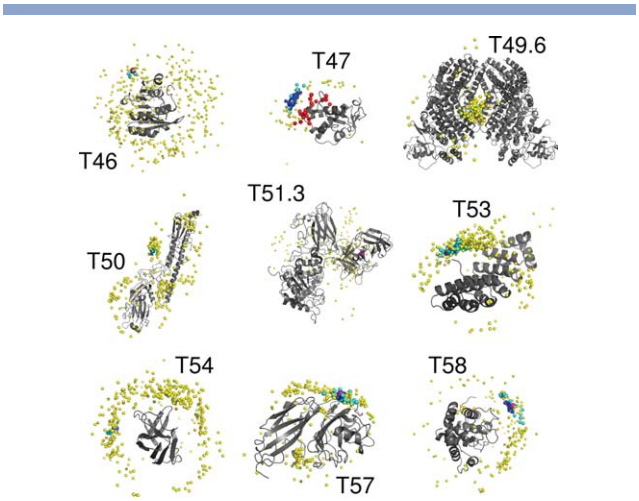

Figure 2

Pictorial view of docking prediction results for targets T46–T58. For each target complex (labeled by the target number), the geometric center of the ligand is represented by a large purple sphere, and the receptor, defined as the larger binding partner, is shown in cartoon representation. For the ligand, the positions of the geometric centers of the predictor-submitted models are plotted relative to each receptor molecule. Color coding of the geometric centers is as follows: high quality: green; medium quality: cyan; acceptable: blue; incorrect: yellow. The yellow spheres for the incorrect models are given a smaller diameter. Finally, the interface water molecules in T47 are displayed as red spheres. The shown targets are: T46: Mtq2/Trm112 methyltransferase, PDB ID 3Q87 (Ref. 26); T47: Colicin E2 DNase/Im2, PDB ID 3U43 (Ref. 27); T48/49: T4moC/T4moH, PDB ID 4I57 (Acheson et al., in preparation; Ref. 28; T50: HB36.3 design/flu HA, PDB ID 3R2X (Ref. 13); T51: Xylanase Xyn10B (Najmudin et al., in preparation); T53: Rep4/Rep2 designs, PDB ID 4JW2 (Ref. 29; T54: Neocarzinostatin/Rep16, PDB ID 4JW3 (Ref. 29); T57: BT4661 protein/heparin, PDB ID 4AK2 (Lowe et al., in preparation); T58: PliG/SalG, PDB ID 4G9S (Ref. 30).

Four groups, those of Weng, Zou, Seok, and Vajda/Kozakov, submitted as many as 10 high quality models each, and two of those (Weng and Zou) submitted very accurate models with L-rms $\leq$ 1 Å. Among the 1051 models uploaded by 12 groups, 278, 305, and 20 were of high, medium and acceptable quality respectively. Scorers likewise performed well, with 63 high quality models submitted by 12 different groups and 38 medium quality models submitted by 10 groups, of a total of 128 scored models.

T48 and T49 were based on the T4moC/T4moH mono-oxygenase complex. In T49, predictions were initiated from the unbound components of both proteins, where the unbound T4moH was an unpublished structure provided by the crystallographers (PDB ID: 4I57) (Acheson et al., in preparation). For T48, the known hexameric structure of T4moH bound to an effector protein (PDB ID: 3DHH)[28] was used instead. For both targets, predictors were given the choice to submit models with either the trimeric or hexameric structure for the T4moH moiety, which were independently assessed (we refer to T48.3/T49.3 for the trimer and T48.6/T49.6 for the hexamer variant of the receptor).

For T48, 1654 models were submitted by predictors using the trimer version of T4moH. Of those, 11 were of medium quality and 39 were acceptable models. Docking calculations that used the hexamer version yielded a similar performance, with 4 medium, and 17 acceptable quality models out of a total of 1004 submitted models.

The performance was overall lower for T49. Here also, fewer acceptable quality models were submitted for the hexamer than for the trimer (see Supporting Information Table S4). Interestingly however, uploaded models submitted for T48 and T49 contained several high quality models (two and six for the trimer complexes of T48 and T49, respectively), as well as a substantial number of acceptable models (around 80 for the trimer versions and about half as many for the hexamer version). For T49, some of these models were successfully identified by scorers. The best performing groups for this target were those of Shen and Zou, followed by Wolfson and Weng.

Round 24: T50, T51

T50 was the complex of the Spanish influenza virus hemagglutinin (HA) with HB36.3, a protein *de novo* designed by the Baker group to bind a conserved epitope on the HA surface, and subsequently optimized using *in vitro* selection.[13] The X-ray structure of the complex, offered here as the target, fully validated the designed model. The docking calculations were initiated using the known structure of HA and a model of HB36.3 built from a bacterial protein of unknown function that was used as scaffold for the design.

Forty groups submitted a total of 357 docking models, among which 16 were of medium quality and 33 of acceptable quality. No high quality models were submitted. The 1251 uploaded models included 19 medium and 37 acceptable quality models only. Of those, scorers were able to identify 14 medium quality and 16 acceptable quality models. The best prediction performance for this target was achieved by the groups of Fernandez-Recio (four medium and five acceptable quality models), followed by the groups of Vajda/Kozakov and Bonvin, and two servers (DOCK-PIE and CLUSPRO) which submitted two medium quality models each. The best scorer performance was achieved by the groups of Wang and Xiao, who identified six and four medium quality models, respectively, and by Fernandez-Recio, Gray, Zou, and Korkin, who identified one medium quality model each. It is furthermore noteworthy that of the 13 participating scorer groups, 12 were able to identify a model of acceptable quality or better.

T51 was a new problem for CAPRI predictors, as it involved docking the different structural domains (in their unbound forms) in this single chain xylanase to one another, while respecting the constraints imposed by chain connectivity. Models for the different pair-wise domain–domain interactions were evaluated

independently. Overall very poor results were obtained for this target. For the CBM13-Fn3 interface, only two groups (Bonvin and Nakamura) produced acceptable models. Many more acceptable models were among the uploaded set by three groups (Bonvin, Nakamura, Mitchell) and one server (SURFIT), and Bonvin uploaded the only medium quality model. Only nine acceptable models were identified by four scorer groups (Wang, Weng, Zou, and Gray).

Round 26: T53, T54

T53 was the complex of two designed alpha repeats Rep2 and Rep4 (PDB ID: 4JW2),[29] and was initiated from the unbound structure of Rep4 (PDB ID: 3LTJ, Ref. 33). For Rep2 only the sequence was supplied, and a model could be built by homology using Rep4 as a high quality template.

Overall quite good results were obtained for this target. Of the total of 384 submitted models, one (by the Vajda/Kozakov group) was of high quality. Eleven medium quality models were submitted by nine groups including two servers (CLUSPRO, LZERD), whereas 11 additional groups submitted at least one acceptable model each, including yet again two servers (SWARMDOCK, SURFIT).

The best performing groups for this target were Vajda/Kozakov, Bonvin and Fernandez-Recio. Of the 1400 uploaded models, 9 were of medium quality, and 92 were acceptable models. Of the 13 groups submitting scoring predictions, nine submitted at least one medium quality model, and two additional groups submitted acceptable models.

T54 was another complex involving the association of an alpha repeat protein Rep16, that had to be modeled by homology, with neocarzinostatin (PDB ID: 4JW3).[29] As already mentioned, this turned out to be a difficult target, presumably in part because the complex formed a hetero-tetramer, in which identical subunits also interacted with one another, a fact of which participants had not been made aware.

Quite poor prediction results were obtained for this target, with only four groups producing six acceptable models, of a total of 377 submissions. The 1400 uploaded models included only one medium quality model (originating from the SWARMDOCK server), and 18 acceptable models. But the 130 models submitted for the scoring exercise were all wrong.

Round 27: T57, T58

T57 of this round was the complex of the BT4661 protein with a six-sugar unit heparin molecule (PDB ID: 4AK2) (Lowe et al., in preparation). Of the 256 submitted models for this target, only five were of medium quality (submitted by the groups of Vajda/Kozakov, Bonvin, and Zou, as well as the HADDOCK server), and 26 were acceptable models, representing a rather low-level performance overall. The best medium quality model was the one by Zou (f_{nat} = 0.71, L-rms = 3.2 Å).

The performance was better for T58, the lysozyme-inhibitor complex (PDB ID: 4G9S).[30] Of the 230 models submitted for this target 15 were of medium quality and 18 were acceptable models. The medium quality predicted complexes were submitted by the groups of Bates, Gray, Fernandez-Recio, Grudinin, and the SWARMDOCK server. The latter featured the best performance (six of medium quality and one acceptable model). The group of Bates submitted the model with the lowest L-rms (3.7 Å). SAXS data was also provided for this target. But a short survey with predictors indicated that they made little use of these data or didn't find the data helpful in improving their predictions. Most of the predictor groups attempted to use the SAXS data in the final selection of models. Only Seok adopted a procedure where the SAXS data were used in the initial selection of decoys, and found it to be helpful. It is indeed unlikely that the low-resolution of SAXS data would help in discriminating between acceptable and almost acceptable high-resolution docking models, but they may prove to be a useful in guiding the initial search for candidate docking solutions with the correct quaternary geometry, which is moreover a computationally expensive step in docking algorithms.

Results across groups and targets

Results of the docking predictions for all 10 assessed targets in CAPRI Rounds 22–27, obtained by all groups that submitted models for at least one target are summarized in Supporting Information Table S11.

As in previous summaries,[20–22] listed results represent only the best prediction obtained by each group for each target. Thus, if a group submitted two acceptable predictions and one high-accuracy prediction for a given target, only the high accuracy result is listed. For a full account of the results obtained by each group the reader is referred to the CAPRI web site (http://www.ebi.ac.uk/msd-srv/capri).

In total, 64 groups including 12 web-servers submitted predictions for at least 1 of the 10 assessed targets. Of these groups, 38 (including eight servers) submitted a model ranked acceptable or higher for at least one target, a slightly lower fraction than in Rounds 13–19.[22] But as in the last evaluation, many of the successful groups/servers in Supporting Information Table S11 are different from those submitting correct predictions in the previous assessment.[22] Of the above-mentioned 38 groups, twelve (including four docking servers) have not participated in the previous CAPRI rounds, reflecting the dynamic turnover of the docking community.

Once again, we find that that success rate crucially depends on the target, and those evaluated here were quite

diverse, with several posing new challenges to the CAPRI community. Among the more challenging targets was T46, the MTq2–Trm112 complex, where the models of the independent components had to be derived using very distantly related templates, and were therefore very inaccurate. There also was no available biochemical information on the interacting regions that data driven procedure such as that of HADDOCK or others could use.

The other two difficult targets were T51 and T54. T51 was a new problem for CAPRI participants, which required finding stable interfaces between covalently linked structural domains, a problem which should in principle be easier than docking of noncovalent subunits, but apparently was not, maybe because two of the six domains of the xylanase protein (one of which was disordered) were not considered in the predictions. The difficulty encountered in T54, the neocarzinostatin-Rep16 complex, came as somewhat of a surprise. It may have been caused by the fact that the two-component complex forms a hetero tetramer, with two copies of each protein, a fact not taken into account by the predictors. T48/49, based on the T4moH/T4moC complex was likewise not straightforward given the asymmetric binding mode of the T4moC ferredoxin molecule to the hexamer.

With the exception of T57, the BT4661-heparin complex, where participants had for the first time to dock a flexible sugar molecule, other targets were much easier problems which CAPRI participants and their software have encountered before. However, with one exception all the high quality models were predicted for the easy target T47. One other high quality model was predicted for T53 by the group of Vajda/Kozakov.

The success rate of individual groups can also be gauged from the ranking presented in Supporting Information Table S11. But care must be taken to account for the fact that some predictor groups and servers only submitted predictions for a subset of the targets.

This time, a total of 11 groups and one docking server submitted acceptable or better models for 5 of the 10 evaluated docking targets. Among those, the group of Bonvin submitted correct models for nine targets, including one high and three medium quality models. The groups of Bates and Vakser submitted eight and seven correct models, respectively. Bates also submitted two medium quality models (for T47, T58) and Vakser 1 high-quality model (for T47). Five other groups (Vajda/Kozakov, Fernandez-Recio, Shen, Zou, and Zacharias) and the CLUSPRO server submitted six correct models, including two to three medium quality ones, and a high-quality model for T47. Two additional groups (Eisenstein, Grudinin) submitted five correct models, and one to two medium quality predictions in addition to one high-quality model each, for T47.

Another 21 groups (including six docking servers) submitted correct predictions for one to four targets, including at least one medium quality model or better, and an additional five predictor groups and the SURFIT server, submitted a correct model for one target. The remaining groups, including four servers, submitted no correct models at all. But no conclusions can be drawn from this performance given that most of these groups/servers submitted predictions for a few targets only.

Progress in scoring docking models

Results of the scoring experiment are summarized in the Supporting Information Table S11B. Figure 3 compares the percentage of correct models submitted for individual targets by scorers, uploaders and predictors.

In total, 31 groups including two automatic servers (SURFIT and HADDOCK) submitted 10 models for at least one of the seven targets for which scoring experiments were run. Eighteen of these groups (but neither of the servers) successfully identified a correct model for at least one of the targets. All groups failed to identify a correct model for T54 the Rep16-neocarzinostatin complex for which very few correct docking solutions were among the uploaded structures [Fig. 3(a)].

Three groups (Zou, Bates, and Weng) submitted models of acceptable quality or better for the remaining six targets. The group of Gray submitted a correct model for five of the targets, followed by Fernanez-Recio, Wang, and Bonvin, with correct models for four of the scored targets. Interestingly, three groups (Elber, Wolfson, and Pal), failed to identify even one correct model for T47, the very easy target for which many correct models were uploaded [Fig. 3(a)].

As in previous evaluations, we find that scoring submissions are enriched with correct models in comparison with the docking submissions, except for the very easy target T47 [Fig. 3(a)]. This enrichment is partly due to the fact that the uploaded ensembles are much smaller (fewer than 2000 models) and are themselves highly enriched with correct solutions in comparison to the vast space of mostly incorrect solutions, that docking calculations typically deal with.

Lastly we noticed that the models submitted by some scorer groups tend to be of better quality than the original models in the uploaded set from which they were derived, indicating that some model refinement procedure has been applied. For example, we see that the seven models submitted by Wolfson (for only two targets) are systematically of better quality [Fig. 3(b)], whereas Zou improves on 75% of the models (for five targets) and Gray improves on ~50% of the models in his scoring submissions. But the same groups—except for Wolfson's—may also submit some poorer quality models than those in the uploaded set, suggesting that for whatever reason, the refinement procedure that is being applied is not always successful, with an outcome that is likely target-dependent. For Zou, these lower quality models represent <10%, whereas for Gray they are just under 50% [Fig.

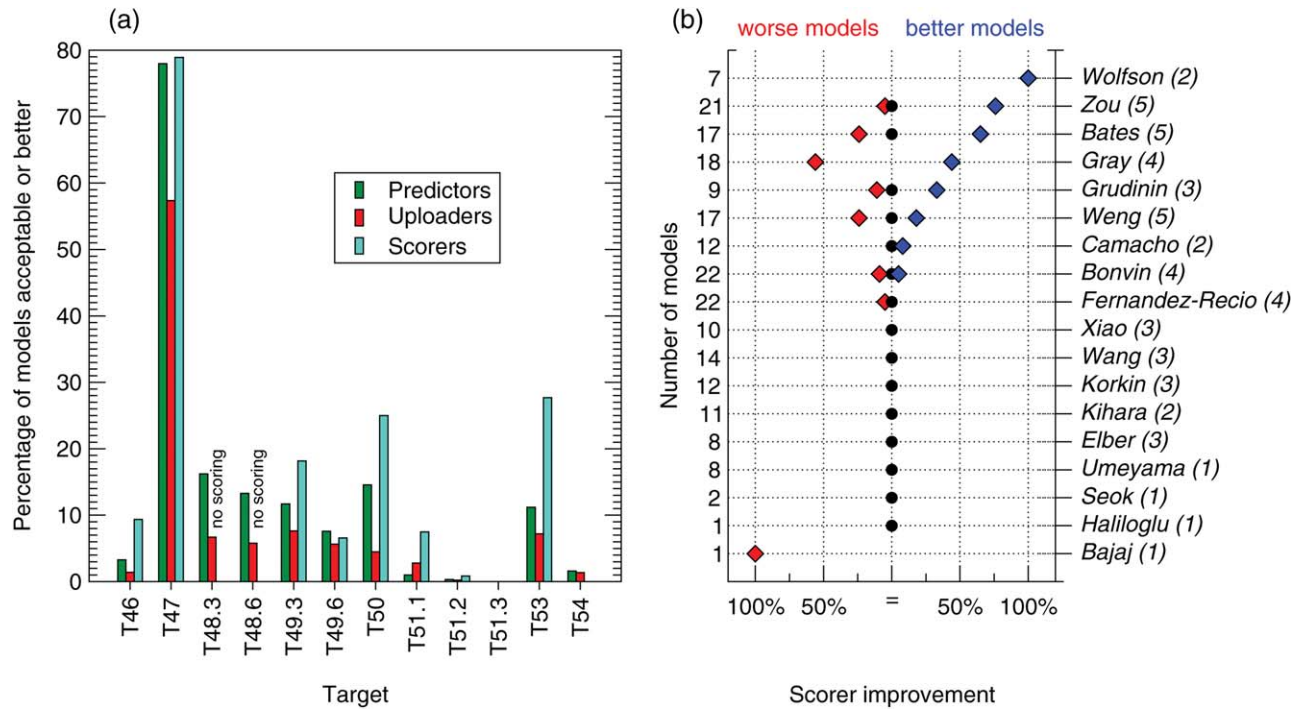

Figure 3

Relative success rates in submitted models, and model quality improvements by scorers. (**a**) Fractions (%) of predicted models, of acceptable quality or better, in CAPRI docking (predictor) and scoring experiments for targets T46–T54. In the scoring experiment, the "Scorer" bars represent the results of models submitted by the scorer groups, whereas the "Uploader" bars represent the fractions of acceptable predictions in the sets of 100 structures uploaded in the scoring experiment. The uploaded models of T48 were pooled together with those of T49 for the scorers to select from. There was no scoring round for targets T57 and T58. (**b**) Illustrates the fraction of models submitted by individual group in the scoring experiment that had an improved quality over the uploaded version, as judged by a decrease in I-rms value. The scorer groups are listed on the vertical axis (right), with in parentheses the number of targets that they submitted models for. The total number of models submitted by each group is listed along the vertical axis (left). The fraction of submitted models (%), which were, respectively, better (blue diamonds) or worse (red diamonds) than the uploaded version from which they were derived is plotted along the horizontal axis. Black dots mean that the model quality was unchanged.

3(b)], and a total of eight scorer groups just re-rank uploaded models without modifying them.

Performance of docking servers

Automatic docking servers play a very important role for the scientific community at large; as they provide nonexperts with access to docking procedures that they can apply to the problem of their choice. Servers operate completely automatically. Their allowed turnaround time in CAPRI is therefore much shorter (1–3 days) than for the docking and scoring predictions, precluding as a result any manual intervention in selecting the final 10 models. In addition, since servers need to behave robustly, they usually implement older but more extensively tested versions of the docking and scoring methods. It is therefore not unexpected for servers to perform less well than human dockers and scorers.

Here, as in the last evaluation, a total of 12 servers took part in the docking experiments. Of these seven are established servers and five (SWARMDOCK, DOCK-PIE, LZERD, SURFIT, and TACOS) are new servers evaluated here for the first time.

It was quite remarkable to see that in Rounds 20–27, some servers performed on par with predictors and scorers. The best performing server CLUSPRO (group of Vajda/Kozakov) (Table IV) is among the seven best performing predictor groups (Supporting Information Table S11A), with correct models submitted for 6 of the 10 evaluated targets. This excellent performance is followed closely by that of HADDOCK (Bonvin), which like CLUSPRO submitted models for all 10 targets and was successful for 4 of them.

The CLUSPRO server maintained its position as top-performing docking server compared to the previous assessment,[22] followed closely by HADDOCK, which remained at Rank 2 (Table IV). SWARMDOCK (Bates) and DOCK-PIE (Elber), which did not participate prior to Round 20, also deserve praise. In particular SWARMDOCK (Bates) submitted models for only four targets (T53–T58), but for each of these, including T54 the very difficult target, this server submitted at least one correct

Table IV
Prediction Performance of Web Servers

	T46	T47	T48	T49	T50	T51	T53	T54	T57	T58	
CLUSPRO	0	**	**	*	**	0	**	0	*	0	6/4**
HADDOCK	*	***	0	*	0	0	0	0	**	0	4/1***/1**
SWARMDOCK	—	—	—	—	—	—	*	*	*	**	4/1**
DOCK-PIE	0	—	*	*	**	—	0	0	—	—	3/1**
HEXSERVER	0	***	—	—	—	—	—	—	—	*	2/1***
FIBERDOCK	—	***	0	0	0	—	—	—	—	—	1***
LZERD	—	—	—	—	—	—	**	0	—	—	1**
SURFIT	0	—	0	0	0	0	*	0	—	—	1
GRAMM-X	0	—	0	0	0	—	0	0	—	—	0
TACOS	—	—	—	—	—	—	0	0	—	—	0
PATCHDOCK	0	—	—	—	—	—	—	—	—	—	0
SKE-DOCK	0	—	—	—	—	—	—	—	—	—	0

Performance is indicated by the best quality of the submitted models (*** for high quality, ** for medium quality and * for acceptable quality) and then summed over the targets. A zero entry specifies that no acceptable model was submitted, whereas a dash (—) indicates no participation for that target.

model. On the other hand servers like FIBERDOCK and GRAMM-X performed less well than previously. We thus see that the performance of docking servers has reached a level of reliability that positions them well to become a powerful research tool in structural biology.

Blind prediction of interface water positions (T47)

Interfacial water molecules are known to play a critical role in both the stability and specificity of colicin DNase-Immunity protein complexes[34,35] Groups submitting standard docking predictions for T47 the DNase domain of colicin E2 and IM2 immunity protein (CAPRI target 47)[27] were therefore invited to predict the positions of water molecules involved in the interface of the complex. The predictions had to be carried out on the basis of docked models of the target complex produced by each predictor group, which as already mentioned, were in general of high quality although clearly not identical to the experimental structure.

Twenty groups rose up to the challenge, submitting water predictions for a total 195 models of the Colicin E2 DNase—IM2 complex. The CAPRI assessment team evaluated these predictions by measuring the fraction of water-mediated protein contacts in the target that was recalled in the submitted models. The results reported in Lensink *et al.*[24] show that of the 176 high or medium quality docking models, which represent a very good performance of the docking calculations, only 78 (44%) had a water-mediated contacts recall fraction above 0.3, and a mere 11 (6%) had a recall fraction above 0.5. Many of the predicted contacts represented false positives—contacts that were not observed in the target, and the actual interface water positions were in general predicted to an accuracy level no better than ~1.5 Å. Interestingly however, the three most highly conserved interface water positions, which are part of the protein–protein interface hotspots[34,35] were quite well predicted, and so was one of the water positions that is believed to stabilize the loop that confers specificity in these complexes.

Overall the best interface water prediction performance was achieved by groups that also produced high quality docking models, indicating that accurate modeling of the protein portion is a determining factor. The use of established molecular mechanics force fields that incorporate electrostatics and solvation effects also seemed to confer an advantage.

Predicting the affinity of protein–protein complexes (T43–45, T55–56)

Using their recently developed procedure for the *de novo* design of binders, David Baker and his team produced two proteins that bind Spanish influenza hemagglutinin (HA).[13] Subsequent *in vitro* evolution yielded variants that bound HA with low, nanomolar dissociation constants (K_d), one of which inhibited HA function. In general however, most of the proteins predicted to bind HA by Baker's Rosetta software were found not to bind the antigen when tested experimentally. Applying their procedure to other target proteins the Baker group found a similar low yield of binders. This prompted D. Baker and S. Fleishman to turn to the CAPRI docking community for help in finding out if force fields developed by docking experts would do better than Rosetta in predicting whether two proteins will form a stable complex or not.

CAPRI Rounds 20 and 21 were set up to address this question. The targets in these rounds (T43–45) were three sets of atomic models representing a total of 87 complexes designed by Rosetta to bind: (a) Spanish influenza HA (62% of the models), (b) the acyl-carrier protein 2 from *M. tuberculosis* (25% of the models) and the Fc region of human IgG1 antibodies (13% of the models). Participating groups were asked to rank these

complexes relative to a reference set of 120 natural complexes from a benchmark set for testing docking procedures.[36] The natural complexes had known K_d values spanning a wide range,[37] whereas Baker's experimental procedure was able to detect binding at a K_d of 10^{-6} M. Only one of the designs was found to bind in Baker's experiment, but after the predictions had already been completed. So an accurate prediction had to rank the majority of the natural complexes as more stable than all but one of the designed models. A total of 28 CAPRI groups rose to the challenge, using a wide range of force fields and scoring functions. The predictions were evaluated by S. Fleishman of the Baker laboratory, and the results were described in a separate paper published collectively.[38]

Overall, the CAPRI community was only marginally successful in discriminating between the natural complexes and the designed ones, and none of the groups managed to identify the successful design. However, analyzing the approaches of the best performing groups allowed identifying specific energy contributions, in particular electrostatic and solvation terms that were not adequately accounted for in the Rosetta force field. This analysis also suggested that structural elements of the originally designed interfaces were not sufficiently well embedded in the structure of the monomers. Modifying the Rosetta design procedure to account for these shortcomings was shown to improve its performance.[38] The experiment was hence successful in helping improve the Rosetta protein design protocol, despite the mediocre prediction performance.

Two other targets from the Baker group (T55–56) were offered for another affinity prediction experiment. These targets were variants of two other proteins, HB36.4 and HB80.3, designed to bind the Spanish influenza HA.[13] The variants were generated using a deep mutational scanning approach aimed at further optimizing the HA binding affinity of these proteins.[39] In this approach, each residue of the HB36.4 and HB80.3 sequences was mutated in turn to all 20 amino acids yielding a total of 1537 single mutants. A DNA library encoding these mutants was engineered and the corresponding proteins were expressed on the surface of yeast cells[14] and subjected to weak selection for HA binding. Deep sequencing techniques[40] were then used to quantify the relative abundance—or enrichment—of a mutant at a given position before and after the selection step, and the enrichment value was taken as an estimate of the contribution of the mutation to the HA binding affinity.

CAPRI participants were given the coordinates of the HB36/HA and HB80/HA X-ray structures and invited to: (1) produce: a rank ordered list of the variants' propensity to bind HA on a scale of 0–1, and (2) rank each mutation as beneficial, neutral, or deleterious to the association with HA.

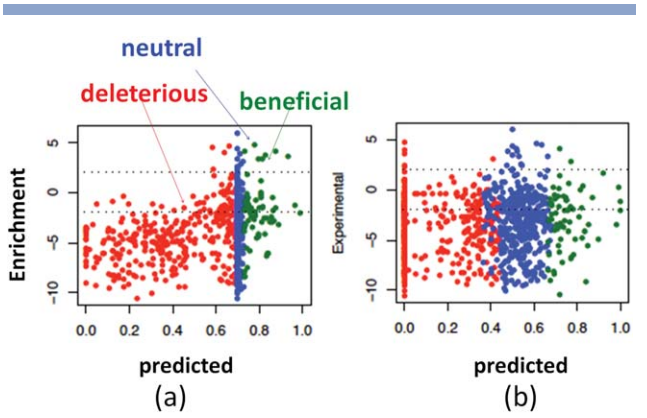

Figure 4

Representative plots of the participant predictions (on an arbitrary 0–1 scale) for the prediction of the effect of mutations in targets T55/56 versus the experimental enrichment value.[23] Participant predictions of beneficial/neutral/deleterious are plotted as green/blue/red dots. The experimental cutoffs of −2 and 2 for the same are represented by dotted lines. (**a**) Plot of all submitted predictions for HB80 for one of the top performing groups. (**b**) Plot of all submitted predictions for HB80 for an average performing group. Reproduced from Ref. 23, with permission from Protein (Wiley).

After these predictions were submitted, a second round of predictions was launched. This time all groups were supplied with the experimental enrichment values for about nine variants at each position of the two proteins, selected at random (unblinded data), and invited to use this information to optimize their prediction methods and then submit updated predictions for the remaining mutants (T56). Twenty-two groups took up the first challenge, and 14 groups submitted updated predictions. R. Moretti of the Baker laboratory evaluated all the submissions.[23]

The evaluation showed that there was considerable variability among groups in distinguishing between beneficial and deleterious mutations. But in general groups were more successful in predicting deleterious mutations than beneficial ones (Fig. 4). The best procedures were those that took into account the effect of the mutation on the stability of the monomer, and which considered packing metrics, Lennard–Jones type potentials, electrostatic and solvation terms. More extensive sampling of side chain conformations and incorporation of some backbone flexibility also conferred an advantage. In the second round, groups using machine-learning techniques[41] performed best. The top performers included information from position- or site-specific models derived from the unblinded portion of the data.

Overall this community-wide experiment showed that a number of current methods for predicting the effects of point mutations on protein interactions produce quite reasonable results, and the analysis of the best approaches suggested, once more, ways of improving on the current Rosetta procedure.

Participants likewise suggested ways in which affinity prediction challenges such as T55 and T56 should be improved in the future, for example, by providing quantitative guidelines for defining the neutral or deleterious mutation categories. Concerns were also raised that enrichment values derived from yeast display and selection followed by deep sequencing may not adequately reflect binding affinities, and therefore future experiments should explicitly benchmark these values against biophysical affinity measures such as those provided in a recent benchmark.[37]

CONCLUDING REMARKS

This fifth edition of the CAPRI assessment dealt with one of the most diverse set of targets and challenges that the CAPRI community has taken on. Of the 15 targets offered in Rounds 20–27, only 10 targets (based on nine complexes) represented classical docking and scoring problems. Two of these targets were difficult multicomponent systems, one with asymmetric binding modes (T48/49), and the other was composed of multiple structural domains of the same chain (T51). In these cases, docking algorithms either lacked the tools or were not properly set up to tackle the problem. Another distinctive feature has been that three of the docking targets (T50, T53, and T54) involved proteins designed by humans. But this was of little consequence on the prediction performance. T57, the protein–oligosaccharide complex was likewise rather well predicted even though CAPRI predictors had little experience with modeling and docking flexible sugar moieties.

Overall, our detailed analysis of these results indicates that the performance of docking algorithms has remained quite robust in face of the changing landscape of targets. Particularly noteworthy is the improved performance of automatic docking servers, which for some servers is now on par with that of human predictions, a very significant achievement by the CAPRI community, which will foster wider use of docking algorithms by the scientific community at large.

This assessment furthermore reports significant forays of the CAPRI community into new areas: that of predicting the positions of water molecules at the interface of a protein–protein complex (the main challenge of T47), and the affinity estimation experiments set up and evaluated by the Baker group, for T43–45 and T55/56. The results of these new forays, published elsewhere[23,24,38] were only summarized here to provide a fuller overview of where CAPRI might be headed in the future.

To meet these new challenges the CAPRI community had to work hard to modify their software, often under stiff time constraints. With very few exceptions[42] docking algorithms were not designed to handle water molecules, and were ill-equipped for estimating relative binding affinities. In particular, scoring functions routinely used in docking calculations are designed to single out models corresponding to a native (stable) complex from a very large set of non-native decoys of the same system. But they cannot be used to predict if a given pair of proteins will interact of not, as this requires to compare the binding energies of different systems and to this end, must take into account both the bound and unbound states of these systems.

The overall mediocre performance by the CAPRI community in singling out native from non-native binders in T43–45, or in ranking the effects of single point mutations on the outcome of the deep mutational scanning experiments (T55–56), thus came as no surprise. Nevertheless the "crowd-sourcing" to CAPRI performed in these experiments has proven extremely useful in suggesting improvements to the Rosetta force field and design protocol, used by the Baker group.

For the interface water predictions in T47, is was also encouraging to see that the two to three water positions known to play an important role in stabilizing the DNase colicin E2–IM2 complex were in general much better predicted than the remaining positions.

Lastly, we note that crystallography groups have significantly contributed to foster progress in docking methods by proposing unusual targets such as the T57 (protein–oligosaccharide complex) or the water prediction challenge of T47. Likewise the affinity prediction experiments initiated by the Baker group have propelled CAPRI participants into the completely new area of estimating the free energy of an interaction rather than just its atomic details, positioning them favorably to start tackling important questions on the biological function and assembly process of protein complexes.

ACKNOWLEDGMENTS

The authors thank all CAPRI groups to the structural biologists who provided the targets of Rounds 20–27: M. Graille and P. Minard (Université Paris-Sud, Orsay, France), J.A. Wojdyla and C. Kleanthous (University of Oxford, UK), B. Fox (University of Wisconsin, Madison, Wiconsin), S. Najmudin (Technical University, Lisbon, Portugal), A. Basle and R. Lewis (Newcastle University, UK), S. Leysen and S. Strelkov (Catholic University Leuven, Belgium), and D. Baker, S. Fleishman, T. Whitehead, R. Moretti, and their colleagues at University of Washington (Seattle, WA), for initiating and providing targets for the affinity prediction experiment. They also thank the CAPRI management team and all predictor groups for stimulating discussion, valuable input, and cooperation.

REFERENCES

1. Alberts B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. Cell 1998;92:291–294.

2. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B, Edelmann A, Heurtier MA, Hoffman V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick JM, Kuster B, Bork P, Russell RB, Superti-Furga G. Proteome survey reveals modularity of the yeast cell machinery. Nature 2006;440:631–636.
3. Guruharsha KG, Rual JF, Zhai B, Mintseris J, Vaidya P, Vaidya N, Beekman C, Wong C, Rhee DY, Cenaj O, McKillip E, Shah S, Stapleton M, Wan KH, Yu C, Parsa B, Carlson JW, Chen X, Kapadia B, VijayRaghavan K, Gygi SP, Celniker SE, Obar RA, Artavanis-Tsakonas S. A protein complex network of *Drosophila melanogaster*. Cell 2011;147:690–703.
4. Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, Li Z, Wang PI, Boutz DR, Fong V, Phanse S, Babu M, Craig SA, Hu P, Wan C, Vlasblom J, Dar VU, Bezginov A, Clark GW, Wu GC, Wodak SJ, Tillier ER, Paccanaro A, Marcotte EM, Emili A. A census of human soluble protein complexes. Cell 2012;150:1068–1081.
5. Hu P, Janga SC, Babu M, Diaz-Mejia JJ, Butland G, Yang W, Pogoutse O, Guo X, Phanse S, Wong P, Chandran S, Christopoulos C, Nazarians-Armavil A, Nasseri NK, Musso G, Ali M, Nazemof N, Eroukova V, Golshani A, Paccanaro A, Greenblatt JF, Moreno-Hagelsieb G, Emili A. Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. PLoS Biol 2009; 7:e96.
6. Babu M, Vlasblom J, Pu S, Guo X, Graham C, Bean BD, Burston HE, Vizeacoumar FJ, Snider J, Phanse S, Fong V, Tam YY, Davey M, Hnatshak O, Bajaj N, Chandran S, Punna T, Christopolous C, Wong V, Yu A, Zhong G, Li J, Stagljar I, Conibear E, Wodak SJ, Emili A, Greenblatt JF. Interaction landscape of membrane-protein complexes in *Saccharomyces cerevisiae*. Nature 2012;489:585–589.
7. Kim PM, Lu LJ, Xia Y, Gerstein MB. Relating three-dimensional structures to protein networks provides evolutionary insights. Science 2006;314:1938–1941.
8. Aloy P, Russell RB. Interrogating protein interaction networks through structural biology. Proc Natl Acad Sci USA 2002;99:5896–5901.
9. Kundrotas PJ, Zhu Z, Vakser IA. GWIDD: a comprehensive resource for genome-wide structural modeling of protein-protein interactions. Hum Genomics 2012;6:7.
10. Berman HM, Coimbatore Narayanan B, Di Costanzo L, Dutta S, Ghosh S, Hudson BP, Lawson CL, Peisach E, Prlic A, Rose PW, Shao C, Yang H, Young J, Zardecki C. Trendspotting in the Protein Data Bank. FEBS Lett 2013;587:1036–1045.
11. Dutta S, Berman HM. Large macromolecular complexes in the Protein Data Bank: a status report. Structure 2005;13:381–388.
12. Pieper U, Webb BM, Barkan DT, Schneidman-Duhovny D, Schlessinger A, Braberg H, Yang Z, Meng EC, Pettersen EF, Huang CC, Datta RS, Sampathkumar P, Madhusudhan MS, Sjolander K, Ferrin TE, Burley SK, Sali A. ModBase, a database of annotated comparative protein structure models, and associated resources. Nucleic Acids Res 2011;39:D465–D474.
13. 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.
14. Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, Myers CA, Kamisetty H, Blair P, Wilson IA, Baker D. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. Nat Biotechnol 2012;30:543–548.
15. Ward AB, Sali A, Wilson IA. Biochemistry. Integrative structural biology. Science 2013;339:913–915.
16. Ritchie DW. Recent progress and future directions in protein-protein docking. Curr Protein Pept Sci 2008;9:1–15.
17. Vajda S, Kozakov D. Convergence and combination of methods in protein-protein docking. Curr Opin Struct Biol 2009;19:164–170.
18. Wodak SJ, Janin J. Structural basis of macromolecular recognition. Adv Protein Chem 2002;61:9–73.
19. Mendez 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.
20. Mendez 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.
21. Lensink MF, Mendez R, Wodak SJ. Docking and scoring protein complexes: CAPRI 3rd edition. Proteins 2007;69:704–718.
22. Lensink MF, Wodak SJ. Docking and scoring protein interactions: CAPRI 2009. Proteins 2010;78:3073–3084.
23. 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, DOI: 10.1002/prot.24356.
24. Lensink MF, Moal IH, Bates PA, Kastritis PL, Melquiond AS, Karaca E, Schmitz C, Van Dijk M, Bonvin AM, Eisenstein M, Jimenez-Garcia B, Grosdidier S, Solernou A, Perez-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, 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 HX, Huang SY, Zou X, Wojdyla JA, Kleanthous C, Wodak SJ. Blind predictions of interfacial water positions in CAPRI. Proteins, in press.
25. Janin J. The targets of CAPRI Rounds 13-19. Proteins 2010;78: 3067–3072.
26. 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.
27. 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.
28. Bailey LJ, McCoy JG, Phillips GN, Jr, Fox BG. Structural consequences of effector protein complex formation in a diiron hydroxylase. Proc Natl Acad Sci USA 2008;105:19194–19198.
29. Guellouz A, Valerio-Lepiniec M, Urvoas A, Chevrel A, Graille M, Fourati-Kammoun Z, Desmadril M, Vam Tilburgh H, Minard P. Selection of specific protein binders for pre-defined targets from an optimized library of artifical helicoidal repeat proteins (alphaRep). PLOS One 2013;8:e71512.
30. Leysen S, Vanderkelen L, Weeks SD, Michiels CW, Strelkov SV. Structural basis of bacterial defense against g-type lysozyme-based innate immunity. Cell Mol Life Sci 2013;70:1113–1122.
31. Das K, Acton T, Chiang Y, Shih L, Arnold E, Montelione GT. Crystal structure of RlmAI: implications for understanding the 23S rRNA G745/G748-methylation at the macrolide antibiotic-binding site. Proc Natl Acad Sci USA 2004;101:4041–4046.
32. Kundrotas PJ, Zhu Z, Janin J, Vakser IA. Templates are available to model nearly all complexes of structurally characterized proteins. Proc Natl Acad Sci USA 2012;109:9438–9441.

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-helicoidal repeat proteins (alphaRep) based on thermostable HEAT-like repeats. J Mol Biol 2010;404:307–327.
34. Kuhlmann 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;301:1163–1178.
35. 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.
36. Hwang H, Vreven T, Janin J, Weng Z. Protein-protein docking benchmark version 4.0. Proteins 2010;78:3111–3114.
37. Kastritis PL, Moal IH, Hwang H, Weng Z, Bates PA, Bonvin AM, Janin J. A structure-based benchmark for protein-protein binding affinity. Protein Sci 2011;20:482–491.
38. 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.
39. Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, Fields S. High-resolution mapping of protein sequence-function relationships. Nat Methods 2010;7:741–746.
40. Metzker ML. Sequencing technologies—the next generation. Nat Rev Genet 2010;11:31–46.
41. Murphy KP. Machine learning: a probabilistic perspective (adaptive computation and machine learning series). MIT press, Cambridge, MA; 2012.
42. van Dijk AD, Bonvin AM. Solvated docking: introducing water into the modelling of biomolecular complexes. Bioinformatics 2006;22: 2340–2347.