PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Docking and scoring protein interactions: CAPRI 2009

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

[1] Genome and Network Bioinformatics, CP 263, BC6, Université Libre de Bruxelles, Blvd du Triomphe, 1050 Bruxelles, Belgium
[2] Department of Structure and Function of Biological Membranes, Université Libre de Bruxelles, Brussels, Belgium
[3] Structural Biology Program, Hospital for Sick Children, 555 University Av. Toronto, Ontario M5G 1X8 Canada
[4] Departments of Biochemistry and of Molecular Genetics, University of Toronto, Toronto Ontario, Canada

ABSTRACT

Protein docking algorithms are assessed by evaluating blind predictions performed during 2007–2009 in Rounds 13–19 of the community-wide experiment on critical assessment of predicted interactions (CAPRI). We evaluated the ability of these algorithms to sample docking poses and to single out specific association modes in 14 targets, representing 11 distinct protein complexes. These complexes play important biological roles in RNA maturation, G-protein signal processing, and enzyme inhibition and function. One target involved protein–RNA interactions not previously considered in CAPRI, several others were hetero-oligomers, or featured multiple interfaces between the same protein pair. For most targets, predictions started from the experimentally determined structures of the free (unbound) components, or from models built from known structures of related or similar proteins. To succeed they therefore needed to account for conformational changes and model inaccuracies. In total, 64 groups and 12 web-servers submitted docking predictions of which 4420 were evaluated. Overall our assessment reveals that 67% of the groups, more than ever before, produced acceptable models or better for at least one target, with many groups submitting multiple high- and medium-accuracy models for two to six targets. Forty-one groups including four web-servers participated in the scoring experiment with 1296 evaluated models. Scoring predictions also show signs of progress evidenced from the large proportion of correct models submitted. But singling out the best models remains a challenge, which also adversely affects the ability to correctly rank docking models. With the increased interest in translating abstract protein interaction networks into realistic models of protein assemblies, the growing CAPRI community is actively developing more efficient and reliable docking and scoring methods for everyone to use.

Proteins 2010; 78:3073–3084.

Key words: protein–protein interaction; docking; CAPRI; macromolecular; predictions; conformational change; scoring; structural genomics.

INTRODUCTION

Substantial efforts are being devoted to the experimental characterization of protein–protein interactions and multiprotein complexes both on the genome scale,[1–4] and in focused experimental studies.[5] Results of these efforts are commonly represented as networks of linked protein components describing the interaction landscapes of cellular processes, specialized cells, or entire organisms. Although this representation might be biologically relevant, it is too abstract to be useful on its own. It lacks crucial information on whether a protein link actually corresponds to a physical contact, whether links form simultaneously, or on the number of protein copies involved (stoichiometry).

However, some of this information can be inferred from known structures of protein assemblies and data on homologs stored in public databases,[6] and work to better exploit these data is in progress.[7] But protein assemblies are still poorly represented in the Protein Data Bank (PDB).[8,9] Although large-scale projects to determine structures of multicomponent complexes (e.g., SPINE2: http://www.spine2.eu/; 3D repertoire: http://www.3drepertoire.org/), have been undertaken, the technologies involved are complex and remain low throughput. In contrast, the repertoire of individual protein 3D structures is increasingly well covered. Given a new protein sequence, chances are that its three-dimensional structure can be built using comparative modeling techniques,[10] as suggested by several large-scale modeling studies.[11,12]

In this context, computational procedures capable of deriving accurate structural models of multiprotein assem-

Additional Supporting Information may be found in the online version of this article.
The authors state no conflict of interest.
Grant sponsors: Canada Institute of Health Research, Sickkids Foundation, Ontario Research Fund, PrioNet Network of Excellence, Canada; Grant sponsor: Région Wallonne of Belgium; Grant number: DGTRE 515993
*Correspondence to: Shoshana J. Wodak, Structural Biology Program, Hospital for Sick Children, 555 University Av. Toronto, Ontario M5G 1X8, Canada. E-mail: shoshana@sickkids.ca.
Received 4 May 2010; Revised 11 June 2010; Accepted 15 June 2010
Published online 14 July 2010 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.22818

blies, starting from the atomic coordinates of the individual components, the so-called "docking" methods, can play an important role in bridging the gap.

Fostering the development of improved docking algorithms and closely monitoring their performance are therefore important tasks that the critical assessment of predicted interactions (CAPRI) has been pursuing since its inception in 2001.

In each Round of CAPRI, individual groups that develop docking procedures are provided with the atomic coordinates of the components of a complex, which they then use to predict the three-dimensional structure of the interacting proteins. In addition to submitting 10 models for each protein complex, predictor groups are also invited to upload a set of 100 models. Immediately after all the predictions are submitted, the uploaded models are shuffled and made available to all groups as part of the CAPRI scoring experiment. The "scorer" groups are invited to rerank all the uploaded models using their preferred scoring function and submit their own 10 best-ranking ones.[13]

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 is deposited with the CAPRI management before publication. All the predictions are made blindly, with no knowledge of the correct answer, and the evaluation is carried out by an independent team, from whom the identity of the predictors is concealed.

Since CAPRI's inception, 12 prediction Rounds were completed with a total of 28 targets. The results were presented at three Assessment Meetings, held in 2002, 2004, and 2007, and described in the literature.[13–16]

Here, we describe the prediction results for Rounds 13–19, which were completed between June 2007 and November 2009. These results were presented at the fourth CAPRI evaluation meeting held in Barcelona in December 2009. The seven prediction Rounds had a total of 14 targets, representing 10 X-ray structures and one solution structure. The targets comprised 10 heterocomplexes, one homo-oligomer and, for the first time, a protein–RNA complex.[17] All of the targets except two involved predictions with unbound components, meaning that predictors were provided with the atomic coordinates of the free components, taken from the PDB, or these had to be modeled by them on the basis of a known structure of a related or similar protein. As in previous CAPRI rounds, no limits were set on the source or type of additional information (homologous proteins, biochemical data on interacting residues) that could be used to bias docking calculations.

We evaluated models submitted in the docking experiment, models uploaded by predictors as well as the reranked models in the scoring experiment. Overall 18,734 models for 14 targets were evaluated. The number of models evaluated in each category, overall and per target, is listed in Table I. These models were contributed by a total of 54 different groups and 12 automatic webservers, 5 more than in 2007.

Table I
Target Participation Statistics

		Number of groups			Number of models				
Round	Target	P	U	S	P	U	S	Receptor	Ligand
13	29	39	22	15	381	2192	145	Trm82	Trm8
14	30	35	14	14	350	1346	131	Rnd1	RBD
	31	Cancelled							
15	32	36	6	12	354	599	120	Savinase	BASI
	33	28	4	7	272	376	70	RNA	Protein
	34	28	3	4	277	210	38		
	35	34	5	11	339	499	101	CBM22	GH10
	36	32	4	8	319	309	80		
16	37	39	17	11	390	1700	110	Arf6	LZ2-JIP4
17	38	40	10	10	399	999	100	Centaurin-α_1	FHA-KIF13B
	39	37	14	14	366	1400	130		
18	40	38	22	15	368	2180	141	Trypsin	API-A
19	41	33	12	13	327	1208	130	E9 DNase	Im2
	42	28			278			homodimer	

Number of predictor (P), uploader (U), and scorer (S) groups, and number of submitted of models submitted by each category of groups for individual targets of CAPRI Rounds 13–19. Predictors and Scorers were allowed to submit at most 10 models for each target, whereas Uploaders could upload 100 models each. The Receptor and Ligand definitions are also given for each target.

The evaluation was performed using a set of previously established criteria, largely adopted by the CAPRI community. Results obtained for models submitted by predictors (docking), uploaders and scorers are discussed for individual targets, as well as across participant groups and across targets. This allows us to present a global overview of the current level of performance of protein docking methods and highlight the problem areas and challenges that docking continues to face.

THE TARGETS

The fourteen evaluated targets can be classified into three broad categories, according to their functional roles[17]: Complexes involved in RNA maturation, those consisting of G-proteins and their associated partners, and complexes involving enzymes. The level of difficulty of the different targets was variable but in general quite challenging. For most of the targets either unbound or homology-built models were used in the predictions. As a result, in four targets (T30, T32, T35/36, and T38/39) models used in the calculations differed from the bound structures in excess of 2.0 Å rms (root mean square deviation) as measured for their backbone atoms (see Table III for details). Furthermore, targets T33/34 in Round 15 involved, for the first time in CAPRI, a protein–RNA complex. In T33, both the protein and RNA components had to be homology modeled, whereas in T34 the structure of the bound RNA moiety was pro-

Table II
CAPRI Assessment Criteria

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 following quantities were computed for each target: (i) all the residue-residue contacts between the Receptor (R) and the Ligand (L), and (ii) the residues contributing to the interface of each of the components of the complex. Interface residues were defined on the basis of their contribution to the interface area, as described previously.[15,16] For each predicted model, the following quantities were computed: the fractions f_{nat} of native and $f_{non\text{-}nat}$ of non-native contacts in the predicted interface; the root mean square displacement (rmsd) of the backbone atoms of the ligand (L-rms), the misorientation angle θ_L and the residual displacement d_L of the ligand center of mass after the receptor in the model and experimental structures were optimally superimposed.[49] In addition, we computed I-rms, the rmsd of the backbone atoms of all interface residues after they have been optimally superimposed. Here the interface residues were defined less stringently, on the basis of residue-residue contacts.[13,15,16] As previously described,[15,16] models exhibiting a number of close atomic contacts (clashes) exceeding by at least two standard deviations the average number of such clashes in all the models submitted for a given target were not evaluated. It should be noted that in the protocol for classifying predicted model into the 4 categories ("Incorrect," "Acceptable," "Medium," and "High"), the listed inequalities were applied from top to bottom, that is, starting with those defining incorrect predictions.

vided. Interestingly, in a number of targets (T30, T32, T37, T38/39) one component formed a homodimer, which associates symmetrically with two identical copies of the second protein. Two targets (T40, T42) had one component simultaneously bound to two copies of the second protein forming two distinct association modes. Predicted models were evaluated against both modes.

Targets T35/36 were a somewhat controversial case. They consisted of a single polypeptide chain (Xyn10B xylanase) with two covalently linked globular domains (GH10 and CBM22), whose mode of association had to be predicted. However, the short peptide linking the domains is disordered in the target crystal structure,[18] and inspection of the experimental structure suggests that the interacting domains may actually belong to two different molecules. It is therefore possible that the two domains do not form stable intramolecular interactions in solution, but this needs to be confirmed experimentally.

Lastly, as in previous CAPRI rounds, predictors (and scorers) commonly exploit biochemical data or information on sequence conservation in related proteins to identify protein regions involved in the interaction. This information is then used to bias the docking and scoring calculations or to filter solutions. This may not be the case for some web-servers, which perform the predictions completely automatically.

THE EVALUATION PROTOCOL

The criteria used to evaluate the quality of the predicted complexes, summarized in Table II, are exactly the same as in previous CAPRI evaluations. The reader is referred to previous CAPRI reports[13,15,16] for a detailed description of these criteria and the corresponding thresholds used in classifying predictions as being of "high," "medium," and "acceptable" accuracy, denoted respectively as "***," "**," and "*" in the summary Tables.

Submitted models containing a number of interatomic clashes (atoms closer than 3 Å) exceeding a certain threshold were not evaluated, as such models may retrieve a large number of native interactions simply because of the interpenetration of the corresponding structures. The threshold used in this and previous evaluations is defined as $C = (N_{clashes}) + 2\sigma$, where the quantity in the brackets is the average number of clashes in all the models submitted for a target, and σ is the standard deviation of this number. C typically ranges between 60–140 for most CAPRI targets, a value quite lenient in comparison to the number of clashes observed in the target structures themselves (usually fewer than 20). However, the exact threshold differs for each target. This has the advantage that predictors do not know in advance what the acceptable number of clashes would be and therefore tend to minimize clashes as much as possible. Alternative ways of defining C have been suggested and will be tested in future CAPRI Rounds.

PREDICTION RESULTS AND DISCUSSION

This section is divided into two main parts. The first part describes the prediction performance for individual targets in different Rounds by the docking and scoring experiments. In the second part, we present an overview

Table III
Summary of Target Prediction Performance in CAPRI Rounds 13–19

			***			**			*		
	L-rms (Å)	R-rms (Å)	P	U	S	P	U	S	P	U	S
T29	1.7	**B**	0	2	1	9	78	13	8	87	13
T30	1.7	2.3	0	0	0	0	0	0	2	2	0
T32	0.3	2.1	15	0	0	13	3	0	6	12	2
T33	2.0	2.6	0	0	0	0	0	0	0	0	0
T34	2.0	**B**	0	0	0	25	13	4	40	165	26
T35	2.9	2.9	0	0	0	0	0	0	1	2	1
T36	2.9	**B**	0	0	0	0	0	0	1	0	0
T37	0.6	0.4	1	8	5	7	34	13	13	34	11
T38	3.2	1.9	0	0	0	0	0	0	0	0	0
T39	3.2	**B**	1	0	0	2	3	0	0	1	0
T40	**B**	0.4	79	176	39	54	163	40	31	149	13
T41	2.0	1.5	24	2	2	58	99	16	67	198	51
T42	1.5	1.5	9			5			6		

B = Bound.
Number of submitted models of acceptable (*), medium (**), or high (***) accuracy, for targets 29–42. Data listed separately for Predictor (P), Uploader (U), and Scorer (S) groups. Ligand and Receptor rmsd (L-rms and R-rms, respectively) calculated between unbound (or homolog) and bound states, with B indicating that a bound target component was used in the docking calculations.

Table IV
Quality of Prediction of the Best Submitted Model

		Best submitted model			Reconstituted model		
	Type	f_{nat}	L-rms (Å)	I-rms (Å)	f_{nat}	L-rms (Å)	I-rms (Å)
T29	**S**	0.71	2.52	0.99	0.78	1.08	0.59
T30	**P**	0.43	7.76	2.52	0.75	0.67	0.54
T32	**P**	0.75	2.08	0.52	0.69	0.54	0.30
T33	**P**	0.10	20.15	15.09	0.08	5.24	11.10
T34	**P**	0.43	1.68	1.51	0.63	1.04	0.69
T35	**U**	0.19	8.76	3.23	0.37	2.57	1.85
T36	**P**	0.37	7.37	3.65	0.37	2.57	1.85
T37	**S**	0.94	1.15	0.78	0.71	0.21	0.26
T38	**P**	0.02	14.68	6.69	0.61	2.21	1.07
T39	**P**	0.82	1.51	0.86	0.61	2.21	1.07
T40	**P**	0.89	0.44	0.31	0.93	0.09	0.26
T41	**P**	0.97	0.74	0.54	0.76	1.15	0.86
T42	**P**	0.66	1.29	0.73	0.51	1.51	0.43

Quality criteria for the best submitted model, and of the reconstituted model from the unbound components, in terms of f_{nat}, L-rms, and I-rms, for individual targets. Also indicated is the source of the best submitted model, that is, predictor (P), uploader (U), or scorer (S).

of the results across predictor groups and targets. In addition we contrast the performance of the docking and scoring experiments, and discuss the results obtained by the automatic servers.

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).

Prediction Results for Individual Targets

The number of models of acceptable quality or higher, collectively submitted for the 13 evaluated targets (T31 is provisionally cancelled) by predictors, scorers, and uploaders, are listed in Table III. The best models obtained for each target and their quality measures are summarized in Table IV, 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 selected targets is presented in Figure 1.

Round 13: T29

T29 is the Trm8/Trm82 complex that carries out guanine methylation in the tRNAs of the yeast *S. cerevisiae*.[19] The catalytic component of the complex Trm8 had to be built by homology using as template the known structure of one of the bacterial methyl transferases in the PDB, whereas Trm82, the noncatalytic component, was a novel structure given to the predictors in the bound form.

Overall the results for this target where quite good, considering that the homology-built model of Trm8 was somewhat inaccurate (1.7 Å backbone rms relative to the bound form). Predictors submitted a total of nine medium quality and eight acceptable models (Table III). Six of the medium accuracy models were from the 10 Eyck group, whereas Bonvin submitted 1 medium accuracy

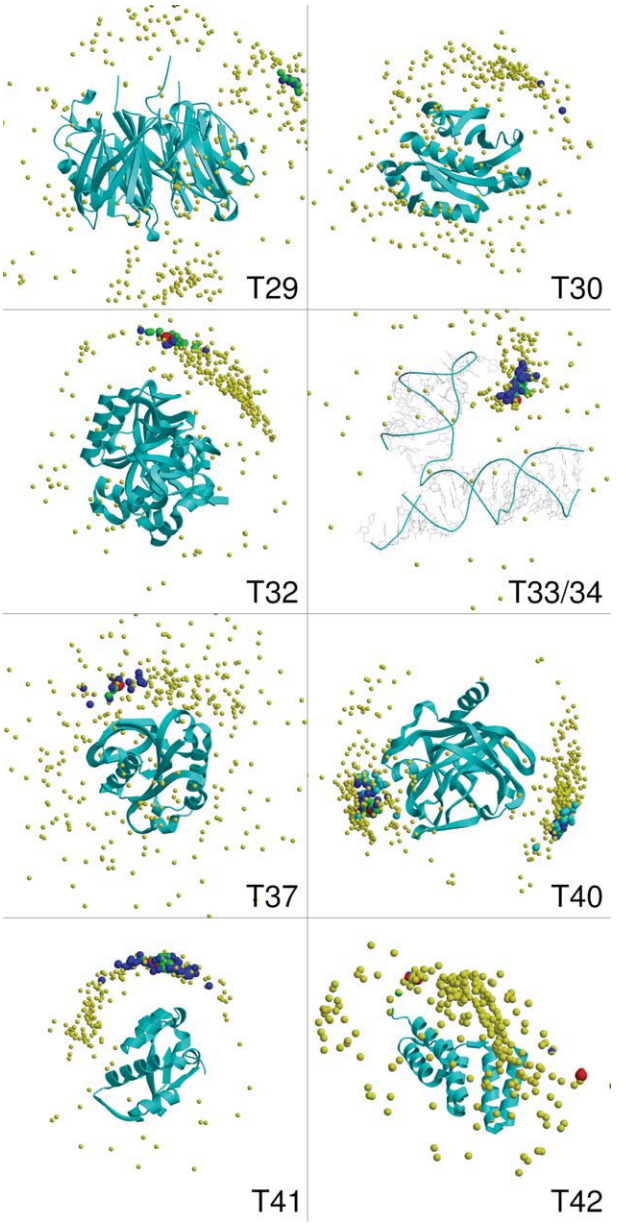

Figure 1
Pictorial views of CAPRI predictions for targets T29–T42. For each target complex (labelled by the target number), the geometric center of the ligand is represented by a red sphere, and the receptor (the largest of the two components) is shown using a ribbon representation. The shown targets are: 2VDU[19] (T29), 2REX[22] (T30), 3BX1[23] (T32), T33/34 (L. Renault *et al.*, unpublished), 2W83[26] (T37), 3E8L[29] (T40), 2WPT[31] (T41), and 2WQH[32] (T42). Targets with very few acceptable solutions are not displayed. For the ligand the positions of the geometric centers of the predictor-submitted models are plotted, relative to each receptor molecule. The definition of Receptor and Ligand for each target is given in Table I. Color-coding of the geometric centers is as follows: target and high-quality: red; medium-quality: green; acceptable: dark blue; incorrect: yellow.

model and six acceptable ones. The groups of Zhou and Fernandez-Recio submitted one medium accuracy model each, whereas Eisenstein, Smith, and Weng each submitted one acceptable model (see Supporting Information Table S1 for detailed summary).

Many more acceptable and higher accuracy models (two high accuracy; 78 and 87 medium and acceptable models, respectively) were among the set of 100 uploaded structures of as many as 11 groups, including six of the successful predictor groups. However, only a fraction of these correct uploaded models was reranked highly in the scoring experiment. Nevertheless, the scorers clearly outperformed the predictors for this target, with a total of one high quality model, by Fernandez-Recio, 13 medium quality models, identified by Bonvin, Weng, Bates, Camacho, Vajda, and Wolfson. The highest quality model identified by Fernandez-Recio (Table IV), approaches the quality of the models reconstructed from the unbound and bound monomers, as it identifies ~71% of the native contacts, and has L-rms and I-rms values of ~2.5 and 1 Å, respectively (see legend of Table II for definition).

Round 14: T30, T31

This Round comprised two targets, corresponding to complexes between two members of the Rho subfamily of G-proteins, Rnd1 and Rac1, respectively, and the Rho-binding domain (RBD) of plexin B1. The Rnd1/RBD and Rac1/RBD complexes were provided as targets for the same Round, but only the Rnd1/RBD complex (T30) was assessed, whereas T31 (Rac1/RBD) (Buck *et al.*, in press) is provisionally cancelled due to unforeseen difficulties in NMR structure determination of this complex, possibly owing to the fact that the RBD protein is a homodimer in solution[20] and in the crystal structure of the free form (PDB code: 2R2O).[21] In the crystal structure of the complex with Rnd1, the RBD proteins also form a dimer, but each monomer interacts with only one Rnd1 copy, forming identical interfaces of standard size (PDB code: 2REX).[22]

The poor results obtained for T30 (Tables III and IV and Supporting Information S2); a total of only two acceptable models predicted by the groups of Bates and Zacharias, are likely due to the difference in conformation between the unbound and bound structures (1.7 and 2.3 Å backbone rms for the Rnd1 and RBD components, respectively, with most of the conformational change occurring at the ligand side of the interface. Of the 1346 uploaded models for this target (Table I) only two were acceptable, with the best model (submitted by Zacharias), identifying only 43% of the native contacts, and displaying a significant displacement of the RBD components in the predicted versus target structures (L-rms of 7.8 Å) (Table IV).

Round 15: T32, T33/34, T35/36

The five targets in Round 15 were based on three complexes with different functions and properties: A complex between the subtilisin Savinase and the α-amylase subtilisin inhibitor BASI (T32),[23] the complex between an enzyme involved in RNA maturation and its RNA substrate (T33/34), which we cannot describe in further detail because the structure has not yet been published, and the association between the GH10 and CBM22 domains of the Xyn10B xylanase (T35/36).[18]

The prediction results for this Round very much depended on the target. For T32 as many as 15 high accuracy models, 4 medium accuracy, and three acceptable models were submitted by 11 predictors. Multiple high accuracy models were predicted by the groups of Gray (5) and Zacharias (4). Baker, Vajda, and Wolfson submitted two high accuracy models each, and three groups (Eisenstein, Camacho, Tovchigrechko) submitted medium accuracy models (see Supporting Information Table S3). The best model was submitted by Gray, with 75% of the native contacts correctly predicted, and only a small displacement relative to the bound BASI component (L-rms = 2.1 Å) (Table IV).

Only one of the successful predictor groups (Wolfson) was among the six groups uploading models. A very small fraction of these models (15 of ~600) were therefore acceptable or better, and it is hence not surprising that the 12 scorer groups had a hard time identifying good models (only two acceptable models, identified by Vakser).

Targets 33/34 derived from the protein–RNA complex posed a new challenge, that of predicting protein-RNA interactions. In T33, homology-built models had to be used in the docking calculations, and as the model for the RNA component in particular differed significantly from the bound structure, it is not unexpected that the submitted and uploaded models were all incorrect. The performance was dramatically improved for T34, when predictors were provided with the bound RNA structure. Dockers submitted 25 medium and 40 acceptable accuracy models. This time as well, several groups (eight in total) produced multiple medium accuracy models each (Supporting Information Table S4). The best model was submitted by Baker (Table IV), although others were of similar quality. Interestingly, two servers—HADDOCK[24] and CLUSPRO[25]—submitted respectively seven and one acceptable models. Only three groups uploaded structures and four groups participated in the scoring experiment (Table I). Of the 210 uploaded models for this target a majority (178) was of acceptable accuracy or better, but only 30 of these were reranked as correct by scorers. Interestingly, Bonvin who uploaded 13 medium accuracy models (the only ones in the entire set) identified only one of them as scorer (Supporting Information Table S4), highlighting yet again that scoring and ranking predicted complexes remains a challenge.

T35/36, the last pair of targets in Round 15, was the case were the information given to predictors about the association mode of the two xylanase domains was

potentially misleading. The collective "failure" to reproduce this association mode (only one acceptable model was amongst the ~300 submitted predictions), even when the bound form of the CBM22 module was provided (T36) (Supporting Information Table S5), reflects the problem and suggests that the covalently linked domains may not form a stable interface. We note however that the single acceptable model submitted for T36 by the Weng group had a similar fraction of correctly predicted native contacts (37%) as the reconstituted model using the unbound components but displayed a larger displacement of the domains relative to each other (Table IV). Anyhow, further experimental evidence is clearly needed to resolve this case.

Round 16: T37

T37 was an interesting target, involving the complex between the G-protein Arf6, and a 77-residue fragment of JIP4 (C-Jun-aminoterminal kinase-interacting protein 4).[26] The latter, denoted as JIP4-LZ2, forms a dimeric leucine zipper structure, analogous to that formed by GCN4, an unrelated transcriptional regulator in yeast, which the predictors were invited to use as a model for LZ2. In the crystal structure, two Arf6 molecules bind to both helices of the LZ2 dimer, forming interfaces of standard size, but do not interact with each other. It turned out that the unrelated GCN4 leucine zipper was quite an accurate model of the LZ2 dimer, and the Arf6 component changes little upon binding to LZ2. In addition, the geometry of JIP4-LZ2-Arf6 complex resembles that of other G-protein assemblies with helical effector domains. It was therefore quite satisfying to see the high success rate of the predictions for this target (Supporting Information Table S6). Five docking groups submitted good models: one high accuracy model by the Weng group and seven medium accuracy models by the groups of Elofsson, Bonvin, Zou, and Zacharias. Five additional groups and one server (SKE-DOCK)[27] submitted acceptable models. Interestingly, five groups uploaded one or two high accuracy models each, in addition to dozens of acceptable models. Several of these accurate models were highly ranked by scorers (Weng, Zou, and Bates). The best model for T37 (94% of native contacts, L-rms = 1.15 Å and I-rms = 0.77 Å) was produced by Zou in the scoring experiment (Table IV), with that of Weng being a close runner up.

Round 17: T38/39

The two targets of Round 17 are based on a complex involved in the same Arf6-dependent membrane trafficking pathway as in T37. Like the latter, it features a heterotetramer, which comprises a dimer of the FHA (forkhead-associated) domain of the KIF13B kinesin motor protein that interacts with two centaurin-α_1 proteins (PDB code 3FM8[28]). The three molecules form two distinct Centaurin/FHA interfaces, of which only one was large enough to be evaluated.[17]

In T38, predictors were required to model the complex on the basis of the unbound model of centaurin-α_1 (PDB code 3FEH) and models of FHA built by homology from other FHA templates. This turned out to be a very challenging docking problem, because the starting models for both components displayed significant conformational changes relative to the bound forms: 1.9 Å rms for the Centaurin moiety and 3.2 Å rms for the homology-built structure of the FHA domain. The latter featured large displacements of a loop contributing to the Centaurin/FHA interface. It is therefore not too surprising that no acceptable models were submitted for this target (Figure 1 and Tables III and IV).

Interestingly, information on the documented affinity of FHA domains for phosphothreonine peptide, which predictors might have used to guide their calculations, was not relevant to the binding mode observed in this complex. This fact also had an impact on the ability to predict this complex, but not for all groups. Indeed, when predictors were provided with the bound structure for FHA in T39, one high accuracy model (by Vajda) and two medium accuracy models by Wang and CLUSPRO were submitted (Tables III and IV and Supporting Information S7). However, of the 1400 uploaded models, all but four models were incorrect, and none of these four models were identified by the 14 scorer groups (Tables I and III).

Round 18: T40

The target in this Round is a ternary complex in which the double-headed arrowhead protease inhibitor API-A binds to two trypsin molecules, involving interactions of the active site of trypsin with two distinct binding sites (implicating respectively Leu 87 and Lys 145) on the API-A protein.[29] This was an easy target, because both API-A binding sites were known in advance. Furthermore, the two trypsin copies undergo negligible conformational change upon association, and the new API-A component was given in the bound form. Not unexpectedly therefore, submitted models for T40, which were evaluated against both binding modes, featured one of the highest success rates of all the Rounds evaluated here. In total 164 correct models (of which 79 of high and 54 of medium accuracy) were submitted by 27 predictor groups. Several of these groups submitted only (Bonvin) or mostly (Gray, Takeda-Shitaka) high accuracy models, and most of the groups were able to identify both interfaces (Tables III and IV and Supporting Information S8).

Many more good models were uploaded, and a sizable fraction of these were reranked by the scorers (39 high accuracy, 40 medium accuracy, and 13 acceptable models). In addition, six servers contributed high and me-

dium accuracy models for either interface, another confirmation that this target was straightforward.

The best models were submitted by the groups of Ten Eyck, the GRAMM-X server,[30] Nakamura and Eisenstein (Table IV).

Round 19: T41, T42

Round 19 had two targets: T41, a complex of the colicin E9 DNase domain with the Im2 immunity protein (PDB code 2WPT)[31] and T42, an oligomer comprising three copies of a designed idealized tetratricopeptide repeat, with tyrosine placed at three sites that should foster intermolecular contacts.[32]

In T41 both components were given in the unbound form, and the conformational changes relative to the bound components are not negligible (1.53 Å rms for E9 and 2.0 Å rms for Im2). This was nonetheless a relatively easy target as most predictors assumed that the complex involves the same interfaces as those conserved in complexes between related immunity proteins and other Colicin E DNase domains.[33] It was therefore comforting to see the high success rate of the submitted models for T41. As many as 22 predictor groups and four servers submitted a total of 149 models of acceptable quality of higher, with 24 high and 58 medium quality submissions (Tables III and Supporting Information S9). A total of 299 models of high, medium, and acceptable quality were uploaded (Table III), leading to a rather good performance by scorer groups (two high accuracy, and 16 medium accuracy models). As for T40, several groups (Takeda-Shitaka, Zou, Zacharias) submitted multiple high accuracy models.

A different prediction landscape emerged for T42. Although initially designed to form a trimer, the crystal structure contains two different dimers, each featuring a standard size interface. In one arrangement the monomers were related by two-fold symmetry and in the other by a screw transformation. In solution the protein forms large oligomers, possibly involving the association modes seen in the crystal. Predicted models were assessed against both modes. A total of nine groups submitted high and medium accuracy models for both interfaces, and seven groups submitted acceptable models. More high accuracy models were submitted for the two-fold dimer (7) than for the other version (2) (Supporting Information Table S10). Interestingly, none of the groups submitted good models for both association modes. Because of time constraints, it was unfortunately not possible to organize the scoring Round before publication of the target; it was therefore cancelled.

Prediction Results across Groups and Targets

Results of the docking predictions for all 13 assessed targets on CAPRI Rounds 13–19, obtained by all groups that submitted models for at least one target, are summarized in Supporting Information Table S11. As in previous summaries,[13,16] 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 one of the 13 targets assessed here. Of these groups, forty-three (including nine servers) submitted a model ranked acceptable or higher for at least one target, a significantly higher fraction of groups (~67%) than in Rounds 9–12 (49%).[13] It is noteworthy that nearly half of the successful groups/servers in Supporting Information Table S11 are different from those that submitted correct predictions in the 2007 assessment.[13] Most of these groups are new participants in the CAPRI experiment reflecting the steady growth of the docking field.

We find yet again that the success rate crucially depends on the target. High accuracy models were obtained by a number of groups for T32, T37, T39, T40, T41, and T42. In addition, medium accuracy predictions were made for T29 and T34. Four of these targets, the Trm8/Trm82 complex (T29), the protein/RNA complex (T34), Centraurin-α_1/FHA (T39), and Trypsin/API-A (T40) had a bound component. But the remaining four targets involved only unbound components that were either provided (T32, T41) or had to be modeled from a related structure (T37, T42). T29 involved both a homology-built and bound component. The significant success rate for these targets suggests that progress is being achieved in handling medium level conformational changes (1.5–2 Å rms) and in using modeled structures for docking.

On the other hand, it comes as no surprise that predictions failed altogether for the RNA/Protein complex of T33, where docking was performed using an RNA model that differed significantly from the bound structure, and for the Centaurin-α_1/FHA complex of T38, which involved docking an inaccurate model built by homology from a distantly related template. Of the remaining three targets with low prediction performance, only T30 (Rnd1/RBD complex) had disappointing results, likely attributable to the 2.3 Å conformational changes of the Rnd1 component upon binding, whereas the "failure" to predict the association of the CBM22 and GH10 domains of xylanase (T35/36) stems primarily from the ambiguity in interpreting the crystal structure and not from limitations of the docking methods.

Supporting Information Table S11 also enables to evaluate the success rate of individual groups. With 13 targets evaluated here, nearly twice as many as in previous

assessments, a clearer global picture emerges. But the same caveat of the limited sample size continues to apply to such evaluation. This notwithstanding, it is very encouraging to see that as many as eight groups can be ranked as top performers in the Rounds evaluated here. All eight groups submitted correct models for 6 of the 13 assessed targets. Among those, the groups of Vajda and Zacharias submitted high accuracy models for four targets, and medium accuracy models for two and one targets, respectively. The remaining top six predictor groups (Zou, Eisenstein Wolfson, Weng, Zhou, Bonvin) submitted between one (Bonvin) and three (Eisenstein) high accuracy models. The ninth top ranking performance is for the CLUSPRO server, with correct predictions for five targets, including three medium accuracy and one high accuracy models.

Another three groups, Fernandez-Recio, Gray, and Bates, but also the HADDOCK server, submitted correct models for four targets, and three more groups (Ritchie, Baker, and Nakamura) correctly predicted three targets. Nine additional groups, including two servers (GRAMM-X and SKE-DOCK), submitted correct predictions for two targets, whereas the remaining 16 groups made valid predictions for only one target (T40, T41 or T42). As in previous assessments, the lower prediction performance of the latter groups, as well as the failure to submit an acceptable model for even one target by the remaining 21 groups does not necessarily reflect shortcoming of the docking methods, as many of these groups submitted predictions for only a few targets.

Progress in Scoring Docked Models, but not in Ranking Them

Results of the scoring experiment are summarized in the Supporting Information Table S12. A total of 40 groups submitted 10 reranked models for at least one of the 12 assessed targets (no scoring experiment was held for T42). Nineteen of these groups submitted a correct model for at least one target.

Two groups, those of Bonvin and Bates, selected models of acceptable or higher accuracy, for 5 of the 12 assessed targets, comprising one high accuracy and several medium accuracy predictions. The groups of Zou, Weng, and Wang each submitted acceptable (or better) models for four targets. Many groups with lower prediction performance in Supporting Information Table S12 scored only a small subset of the targets, which were sometimes also the more difficult ones (T29, T30, or T38/39). Uploaded models for such targets were often of very poor quality, precluding the selection of good models by any scoring scheme.

On the other hand, scorers were usually quite successful when the uploaded set contained good models. This is illustrated in Figure 2, which contrasts the percentage

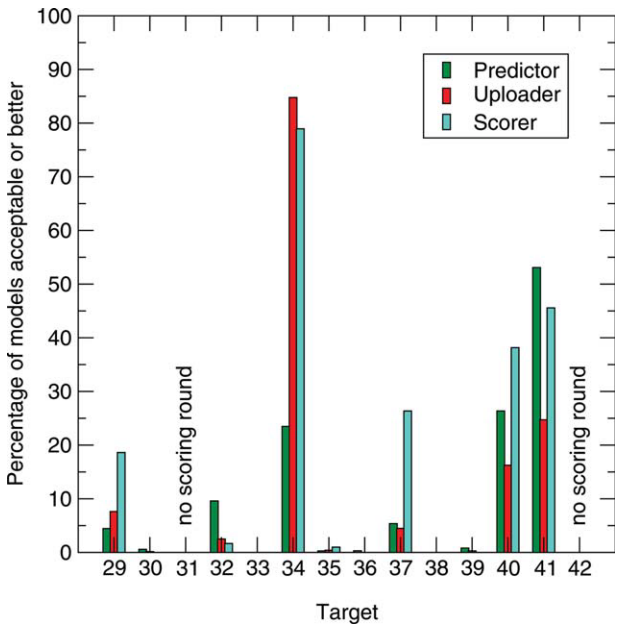

Figure 2

Relative success rates of CAPRI predictions for targets T29–T42. Fractions (%) of predicted models rated as acceptable accuracy or better, in CAPRI docking (predictor) and scoring experiments, for targets T29–T42. In the latter experiment the "Scorer" bars represent the results of models submitted by the scorer groups, whereas the "Uploader" bars represent the fractions of correct predictions (acceptable accuracy or better) in the sets of 100 structures uploaded in the scoring experiment.

of models of acceptable quality or better submitted in the docking and scoring experiments with that in the uploaded sets. We see that models submitted by scorers are usually significantly enriched in good models in comparison to those submitted by predictors. Exceptions to this trend occur for more difficult targets, such as T30, T32, or T39, once more because of the paucity of good models in the uploaded set.

For some targets correct models represent 20% or more of the docking submissions, and over 30% of the scoring submissions. In comparison, the fraction of correct models in the 10-fold larger uploaded sets (100 predicted models uploaded by each group, when compared with only 10 in the docking or scoring experiment) tends to be smaller and rarely exceeded ~15%.

Although the observed enrichment in correct models for the scoring experiment does suggest progress in scoring methods, it is important to remember that the uploaded set from which scorers select their best models represents a limited ensemble of putative complexes (totaling a couple of thousands at best (Table I). In comparison, the docking experiment may generate millions of solutions from which promising candidates need to be singled out, hence significantly decreasing the odds that such candidates can be identified by chance alone, compared with the scoring experiment.

Table V
Prediction Performance of Web-Servers

Target	29	30	32	33	34	35	36	37	38	39	40	41	42
ClusPro	0	0	0	0	1*	0	0	0	0	1**	2/1**	1**	1***
FiberDock												10/1***	0
FireDock			0	0	0	0	0	0	0	0	2/1***		
GRAMM-X	0	0	0	0	0	0	0	0	0	0	2***	1***	0
HADDOCK			0	0	7*	0	0	0	0	0	1***	4/1**	1*
SKE-DOCK	0	0	0	0	0	0	0	2*	0	0	2/1***	0	0
Top down								0	0		2/1**	0	0

Performance is indicated by the number of submitted models of acceptable accuracy or better, specifying after the slash the number of models with the indicated accuracy ("**" for medium accuracy, and "***" for high accuracy). A zero entry indicates that no acceptable model was submitted, whereas an empty entry indicates no participation for that target.

As previously observed,[13] scorer groups often tend to select models uploaded by other groups, rather than their own, even when their own models are more accurate. This remains puzzling, and likely indicates that singling out the better models from sets containing a larger proportion of good models requires finer discrimination criteria than those used in docking. We also observe that scorers rarely improve upon the quality of the uploaded model in the scoring process, for example as a result of further refinement.

The same limitations also adversely affect the ability to correctly rank submitted models in both the scoring and docking experiments. This ranking is often based on the scoring scheme that predictors use. But some predictors trust less their scoring function, and use ad hoc criteria instead. In either case, the position of a model in the ranked list of 10 predictions submitted for each target reflects the degree of confidence in that model, with high confidence models appearing at top of the list. In line with previous findings,[13] we see no obvious correlation between the ranks of models and their accuracy as determined here for the 13 evaluated targets.

Performance of Docking Servers

The steady increase in the number of participating docking servers is a very welcome development in CAPRI, as it heralds wider access to docking procedures by nonexperts. Servers are operated completely automatically and their allowed turn around time in CAPRI is 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. Also, because servers need to behave robustly, they often implement less recent but more extensively tested versions of the docking and scoring methods.

It is therefore not unexpected that the performance of servers is not on par with that of human dockers or scorers. Table V summarizes the prediction results of servers that submitted models of acceptable quality or higher for at least 1 of the 13 evaluated targets in the docking experiment. Seven of the 12 servers have entries in this Table. Two of these, ClusPro, and HADDOCK, currently outperform their counterparts. ClusPro submitted correct models for 5 of the 13 targets, and HADDOCK for four. The remaining five servers submitted correct models for at most two targets, although not always the easier ones. FiberDock and Top Down submitted predictions for only two and five targets, respectively.

DOCKING METHODS: WHAT IS NEW?

CAPRI continues to fulfill one of its major goals in remaining a fertile testing ground for new docking methods, which predictors develop and apply to meet new challenges posed by the increasingly diverse and realistic targets offered to them.

Rigid-body search algorithms remain the well-established core component of most docking procedures.[34,35] Several docking packages implement efficient versions of these algorithms, (HEX, ZDock, HADDOCK, ClusPro) and have become increasingly popular outside the docking community. New CAPRI participants also tend to build on these packages, enabling them to concentrate on other aspects that are crucial to successful docking predictions.

The docking community currently recognizes three main aspects that need addressing: (1) improving the criteria for singling out promising solutions, (2) modeling conformational changes, and (3) incorporating restraints on the basis of data from different sources.

Criteria for Singling Out Promising Docking Poses

Recent approaches for selecting promising docking poses focus on improving the selection efficiency during both the coarse-grained rigid body search and in subsequent refinement steps.

An increasing number of docking procedures consider ensembles of docking poses. They cluster solutions obtained in the rigid-body search, and process further models corresponding to the most densely populated clusters. Several procedures also compile statistics on how frequently individual residues appear in the interfaces of computed docking solutions, and then select candidate docking poses that involve residues occurring frequently in these interfaces (Zhou).

Groups continue to test and apply new scoring functions specifically designed to discriminate between correct and incorrect binding modes. These functions tend to include residue or atom pair potentials, as well as a variety of other structural or physicochemical features derived from known protein interfaces. Various flavors of these functions are used in both the rigid-body search and refinement steps.

In the refinement step, the sampling of internal degrees of freedom, using various techniques (see below), is closely

intertwined with the application of more sophisticated scoring functions. It is therefore very difficult to comment on the effectiveness of these functions beyond what we did here, namely showing that the ability of predictor groups to correctly rank or rerank models in both the docking and scoring calculations, remains limited. A much more quantitative assessment is reported in a recent study,[36] which evaluated the performance of 10 established scoring functions for protein–protein interactions, by comparing computed scores with experimentally measured binding affinities for 81 protein complexes from the Protein–protein docking Benchmark 3.0.[37]

Lastly, a simulation framework very different from the typical docking protocols, and much more computationally demanding, offers completely new ways of singling out promising docking solutions. This framework is geared at searching for encounter complexes,[38] and simulates interactions events between multiple copies of the component proteins in a crowded environment. The frequency of encounter events and dynamic parameters, such as the time interval that any two proteins remain in contact during the simulations (retention time), seem to be promising discriminators between specific and nonspecific complexes. This is an important new research area that will clearly be very relevant to CAPRI in the future.

Modeling Conformational Changes

Modeling conformational changes involving backbone and loop rearrangements remains a major stumbling block that methods-developers are striving to overcome.[39] The changes that need to be modeled may be those undergone by the protein components upon association, or result from using inaccurate models built by homology in the docking calculations, as was the case for several of the targets evaluated here (T29, T35, and T38). To address this issue, an increasingly wider pallet of strategies is being devised. Docking ensembles of conformers generated for the individual components is quite popular. Such ensembles may be derived by molecular dynamics or related methods (Bonvin, Bates, Zacharias), by Monte Carlo Sampling (Gray), or by simulating motion along normal modes (Sternberg, Zacharias). An interesting alternative proposed by Mitchell uses structural homologs of the components in the PDB to derive large sets of conformational ensembles that are then docked to one another.[40]

In most cases, these various approaches are time consuming and therefore used mainly in the refinement step. Recently however, several groups directly incorporate them into the rigid-body sampling step (Zacharias, Gray).

Incorporating Restraints on the Basis of Data from Different Sources

Predictors usually exploit as much as possible biochemical information on the residues involved in the interaction, or data on residue conservation across related proteins, to bias docking calculations or filter solutions. Structural data on related or similar complexes, when available, can be particularly helpful, as was the case for the JIP4-LZ2/Arf6 complex in T37 and that of colicinE9/Im2 in T41.

Methods that more fully exploit such data are also gaining ground. These methods use templates from the compendium of known protein complexes to derive the association modes of two protein components. This can be done by employing homology modeling[41] or threading techniques[42] or by using structural alignments, preferably focused on interface residues.[43] Currently, however, this type of approaches remains limited by the relative paucity of known complexes and by the low accuracy of the predicted models.

Difficult targets often turn out to be those for which such Supporting Information cannot be obtained. To improve docking results in such cases, predictors recently proposed to use computational methods to infer candidate binding surfaces from structural properties of the individual proteins, and then use those to bias the docking calculations.[44,45] It is unclear however, how useful such methods really are in light of their generally poor performance record.[46]

On the other hand, particularly noteworthy is the progress achieved by docking methods that integrate restraints based on data derived from multiple sources. These so-called data-driven docking methods are exemplified by the HADDOCK package[47] and by Attract,[44] which can integrate a wide range of data types, and by software such as MultiFit, which uses low resolution cryo-EM maps as restraints in docking atomic resolution models.[48] These methods are powerful tools when the experimental information they rely on is correct. But they are also particularly vulnerable to incorrect or ambiguous data on the interfaces involved, on the stoichiometry, or the symmetry of the complex. The lower performance of these and other docking methods for T35/36 may be an example of such instances.

CONCLUDING REMARKS

In this fourth edition of the CAPRI assessment, we evaluated over 18,000 predicted models for 13 targets, corresponding to the three-dimensional structures of 11 distinct complexes. Most targets represented challenging aspects for the docking and scoring experiments, which were evaluated in parallel. For the first time in CAPRI, predictors had to model protein–RNA interactions. They also had to consider docking solutions that involve multiple interfaces between a protein pair and about half the targets involved the use of modeled three-dimensional structures.

It was therefore most encouraging to see that high and medium accuracy models were submitted for as many as

eight targets, with a significant number of groups submitting multiple models of acceptable quality or better for several of these. It is noteworthy that many of these groups are newcomers to CAPRI. Several automatic servers, veteran as well as new ones, were also quite successful. With some well-established docking programs and servers available for use by anyone, new groups can readily join the CAPRI community in addressing various aspects of the docking problem that need improvement. The frequent requirement to dock homology-built models is not a deterrent for groups with little experience in building such models, as they can count on experienced members of the CAPRI community to build these models for them.

The thrust of recent methods developments is on modeling protein plasticity, improving the criteria for singling out specific association modes from nonspecific ones, and on integration of data from different sources to restrain the search of docking solutions. Although the challenges remain significant, many creative approaches are being tested and clear progress is being achieved.

In closing, we reiterate our call upon X-ray crystallographers and NMR experts to trust us with their structures. The submission of a target to CAPRI does not jeopardize the confidentiality of the work, since submitted atomic coordinates remain confidential until released by the author or by the PDB. On the other hand, it enhances work done by the structural biologists and in some instances can contribute to the interpretation of the structural data.[13]

ACKNOWLEDGMENTS

The authors express the gratitude of all CAPRI groups to the structural biologists who provided the targets of rounds 13–19: N. Leulliot (Université Paris-Sud, Orsay, France), Y. Tong, W. Tempel, C. Arrowsmith, and H. Park (Structural Genomics Consortium, University of Toronto, Canada), M. Buck (Case Western Reserve University, Cleveland, USA), L. de Maria (Novozymes, Bagsvaerd, Denmark), L. Renault (LEBS-CNRS, Gif-sur-Yvette, France), S. Najmudin, and C. Fontes (University of Lisbon, Portugal), J. Ménétrey (Institut Curie, Paris, France), R. Bao, and Y. Chen (University of Science and Technology of China, Hefei, China), C. Kleanthous and N. Meenan (University of York, United Kingdom). Finally our thanks also go to the CAPRI management team and all predictor groups for stimulating discussion, valuable input, and cooperation.

REFERENCES

1. 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.
2. Krogan NJ, Cagney G, Yu HY, Zhong GQ, Guo XH, Ignatchenko A, Li J, Pu SY, Datta N, Tikuisis AP, Punna T, Peregrin-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B, Richards DP, Canadien V, Lalev A, Mena F, Wong P, Starostine A, Canete MM, Vlasblom J, Wu S, Orsi C, Collins SR, Chandran S, Haw R, Rilstone JJ, Gandi K, Thompson NJ, Musso G, St Onge P, Ghanny S, Lam MHY, Butland G, Altaf-Ui AM, Kanaya S, Shilatifard A, O'shea E, Weissman JS, Ingles CJ, Hughes TR, Parkinson J, Gerstein M, Wodak SJ, Emili A, Greenblatt JF. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature 2006;440:637–643.
3. Kuhner S, van Noort V, Betts MJ, Leo-Macias A, Batisse C, Rode M, Yamada T, Maier T, Bader S, Beltran-Alvarez P, Castano-Diez D, Chen WH, Devos D, Guell M, Norambuena T, Racke I, Rybin V, Schmidt A, Yus E, Aebersold R, Herrmann R, Bottcher B, Frangakis AS, Russell RB, Serrano L, Bork P, Gavin AC. Proteome Organization in a Genome-Reduced Bacterium. Science 2009;326:1235–1240.
4. Yu HY, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, Hao T, Rual JF, Dricot A, Vazquez A, Murray RR, Simon C, Tardivo L, Tam S, Svrzikapa N, Fan CY, de Smet AS, Motyl A, Hudson ME, Park J, Xin XF, Cusick ME, Moore T, Boone C, Snyder M, Roth FP, Barabasi AL, Tavernier J, Hill DE, Vidal M. High-quality binary protein interaction map of the yeast interactome network. Science 2008;322:104–110.
5. Breitkreutz BJ, Stark C, Reguly T, Boucher L, Breitkreutz A, Livstone M, Oughtred R, Lackner DH, Bahler J, Wood V, Dolinski K, Tyers M. The BioGRID interaction database: 2008 update. Nucleic Acids Res 2008;36:D637–D640.
6. Kim PM, Lu LJ, Xia Y, Gerstein MB. Relating three-dimensional structures to protein networks provides evolutionary insights. Science 2006;314:1938–1941.
7. Kundrotas PJ, Zhu ZW, Vakser IA. GWIDD: genome-wide protein docking database. Nucleic Acids Res 2010;38:D513–D517.
8. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The protein data bank. Acta Crystallogr D Biol Crystallogr 2002;58(Pt 6 No 1):899–907.
9. Dutta S, Berman HM. Large macromolecular complexes in the Protein Data Bank: a status report. Structure 2005;13:381–388.
10. Pieper U, Eswar N, Davis FP, Braberg H, Madhusudhan MS, Rossi A, Marti-Renom M, Karchin R, Webb BM, Eramian D, Shen MY, Kelly L, Melo F, Sali A. MODBASE: a database of annotated comparative protein structure models and associated resources. Nucleic Acids Res 2006;34:D291–D295.
11. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. Annu Rev Bioph Biom 2000;29:291–325.
12. Yamaguchi A, Iwadate M, Suzuki E, Yura K, Kawakita S, Umeyama H, Go M. Enlarged FAMSBASE: protein 3D structure models of genome sequences for 41 species. Nucleic Acids Res 2003;31:463–468.
13. Lensink MF, Mendez R, Wodak SJ. Docking and scoring protein complexes: CAPRI 3rd edition. Proteins 2007;69:704–718.
14. Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJ, Vajda S, Vakser I, Wodak SJ. CAPRI: a critical assessment of predicted interactions. Proteins 2003;52:2–9.
15. 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.
16. 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.
17. Janin J. The targets of CAPRI rounds 13–19. Proteins 78;15:000–000.
18. Najmudin S, Pinheiro BA, Prates JAM, Romao MJ, Fontes CMGA. Putting an N-Terminal End to the Clostridium thermocellum Xyla-

nase Xyn10B Story: Crystallographic Structure of the CBM22–1-GH10 Modules Complexed with Xylohexaose. J Struct Biol; doi: 10.1016/j.jsb.2010.07.009.
19. 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.
20. Tong YF, Hughes D, Placanica L, Buck M. When monomers are preferred: a strategy for the identification and disruption of weakly oligomerized proteins. Structure 2005;13:7–15.
21. 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.
22. Tong Y, Tempel W, Shen L, Arrowsmith CH, Edwards AM, Sundstrom M, Weigelt J, Bochkarev A, Park H. Crystal structure of the effector domain of PLXNB1 bound with Rnd1 GTPase, in press.
23. Micheelsen PO, Vevodova J, De Maria L, Ostergaard PR, Friis EP, Wilson K, Skjot M. Structural and mutational analyses of the interaction between the barley alpha-amylase/subtilisin inhibitor and the subtilisin savinase reveal a novel mode of inhibition. J Mol Biol 2008;380:681–690.
24. de Vries SJ, van Dijk AD, Bonvin AM. The HADDOCK web server for data-driven biomolecular docking. Nature Protoc 2010;5:883–897.
25. Comeau SR, Kozakov D, Brenke R, Shen Y, Beglov D, Vajda S. ClusPro: performance in CAPRI rounds 6–11 and the new server. Proteins 2007;69:781–785.
26. 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.
27. Terashi G, Takeda-Shitaka M, Kanou K, Iwadate M, Takaya D, Umeyama H. The SKE-DOCK server and human teams based on a combined method of shape complementarity and free energy estimation. Proteins 2007;69:866–872.
28. Shen L, Tong Y, Tempel W, MacKenzie F, Arrowsmith CH, Edwards AM, Bountra C, Weigelt J, Bochkarev A, Park H. Crystal structure of full length centaurin alpha-1 bound with the FHA domain of KIF13B, in press.
29. 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.
30. Tovchigrechko A, Vakser IA. GRAMM-X public web server for protein-protein docking. Nucleic Acids Res 2006;34 (Web Server issue):W310–W314.
31. 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.
32. Krachler AM, Sharma A, Kleanthous C. Self-association of TPR domains: lessons learned from a designed, consensus-based TPR oligomer. Proteins 2010;78:2131–2143.
33. 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.
34. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 2002;47:409–443.
35. Smith GR, Sternberg MJ. Prediction of protein-protein interactions by docking methods. Curr Opin Struct Biol 2002;12:28–35.
36. Kastritis PL, Bonvin AM. Are scoring functions in protein-protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. J Proteome Res, in press.
37. Hwang H, Pierce B, Mintseris J, Janin J, Weng Z. Protein-protein docking benchmark version 3.0. Proteins 2008;73:705–709.
38. Li X, Moal I, Bates PA. Detection and refinement of encounter complexes for protein protein docking: taking account of molecular crowding. Proteins 78;15:000–000.
39. Lensink MF, Mendez R. Recognition-induced conformational changes in protein-protein docking. Curr Pharm Biotechnol 2008;9: 77–86.
40. Demerdash O, Buyan A, Mitchell J. ReplicOpter: a replicate optimizer for flexible docking. Proteins 78;15:000–000.
41. Aloy P, Russell RB. Interrogating protein interaction networks through structural biology. Proc Natl Acad Sci USA 2002;99:5896–5901.
42. Lu L, Lu H, Skolnick J. MULTIPROSPECTOR: an algorithm for the prediction of protein-protein interactions by multimeric threading. Proteins 2002;49:350–364.
43. Sinha R, Kundrotas P, Vakser I. Docking by strutural similarity at protein-protein interfaces. Proteins 78;15:000–000.
44. Fiorucci S, Zacharias M. Binding site prediction and improved scoring during flexible protein protein docking with attract. Proteins 78;15:000–000.
45. Gong J, Wei T, Stark RW, Jamitzky F, Heckl WM, Anders HJ, Lech M, Rossle SC. Inhibition of Toll-like receptors TLR4 and 7 signaling pathways by SIGIRR: a computational approach. J Struct Biol 2010;169:323–330.
46. Ezkurdia I, Bartoli L, Fariselli P, Casadio R, Valencia A, Tress ML. Progress and challenges in predicting protein-protein interaction sites. Brief Bioinform 2009;10:233–246.
47. Karaca E, Melquiond AS, de Vries SJ, Kastritis PL, Bonvin AM. Building macromolecular assemblies by information-driven docking: introducing the HADDOCK multi-body docking server. Mol Cell Proteomics, in press.
48. Lasker K, Topf M, Sali A, Wolfson HJ. Inferential optimization for simultaneous fitting of multiple components into a CryoEM map of their assembly. J Mol Biol 2009;388:180–194.
49. MacLachlan AD. Rapid comparison of protein structures. Acta Crystallogr A 1982;38:871–873.