PROTEINS: Structure, Function, and Bioinformatics 60:170–175 (2005)

The Targets of CAPRI Rounds 3–5

Joël Janin*
Laboratoire d'Enzymologie et de Biochimie Structurales, Gif-sur-Yvette, France

ABSTRACT **Ten protein–protein complexes have been offered by X-ray crystallographers as targets for structure prediction in Rounds 3–5 of the CAPRI experiment. They illustrate molecular recognition in several domains of biology: enzyme regulation, antigen–antibody recognition, signal transduction, and oligomer assembly. The targets presented various degrees of difficulty to the predictors, depending on their status (bound when components were taken from the complex, unbound when coming from independent structures of the free proteins), the amplitude of conformation changes, and the amount of biological information available. Predictors produced high-quality models of 6 of the targets, good models of 3 others, and failed only in 1 case, where the conformation change was particularly large. This result demonstrates significant progress relative to earlier rounds of CAPRI. Proteins 2005;60:170–175.** © 2005 Wiley-Liss, Inc.

Key words: protein–protein complexes; conformation changes; quaternary structure; epitope prediction; protein–protein docking

INTRODUCTION

The CAPRI community-wide experiment was designed in 2001 to assess blind predictions of the structure of target protein–protein complexes. In CAPRI, the 3-dimensional (3D) structure is predicted from that of the component proteins, and the models submitted by predictor groups are evaluated against experimental structures offered by X-ray crystallographers prior to publication. In 4 years, 5 rounds of CAPRI predictions have been performed on a total of 17 targets illustrating various aspects of protein–protein recognition in several domains of biology. Rounds 1 and 2 were completed in 2001–2002, and their results have already been described.[1–3] Rounds 3–5, completed in 2003–2004, included 12 targets. Two were cancelled; the other 10 are presented here. I discuss features that were relevant to CAPRI and could make a target difficult or easy to the predictors, who could submit up to 10 models of the target ranked 1 to 10. Other articles in this issue deal with the algorithms and methods used by the predictor groups that participated in Rounds 3–5, and the evaluation of the models by the CAPRI assessors.

Figure 1 illustrates the experimental structures of the targets, and Table I lists the relevant references, along with some of the target properties. The status of a target depends on which type of atomic coordinate sets was communicated to the predictor groups: "bound" means coordinates taken from the X-ray structure of the complex; "unbound," coordinates taken from an independent study of the free protein; "homology" is also unbound, but the free protein is homologous rather than identical, to a component of the complex that must be model-built. Due to conformation changes, the bound and unbound structures are generally different, and the RMSD (root-mean-square distance) between main-chain atom positions in the 2 sets gives an estimate of the amplitude of the change. The interface area is the sum of the solvent accessible surface areas of the 2 components of a complex less that of the complex, a convenient measure of the extent of the interaction. Other columns of Table I cite the number of predictor groups that submitted models of the target, and a summary of their collective achievement as evaluated by the CAPRI assessors.

TARGETS

T15–T18: Mitigated Success on Enzyme–Inhibitor Complexes

Enzyme–inhibitor interaction plays a part in many biological regulations. It is the earliest known mode of protein–protein recognition, and the first to have been analyzed by docking.[13] Targets T15–T18 of Round 5 were enzyme–inhibitor complexes. In T15, presented by M. Graille and H. van Tilbeurgh (LEBS-CNRS, Gif-sur-Yvette, France), the enzyme is the ribonuclease domain of colicin D, a toxin secreted by some strains of *Escherichia coli* to kill competing bacteria.[9] The inhibitor is the specific ImmD immunity protein produced by the same strains to protect themselves against the toxin. As both the colicin and the immunity protein were new structures, the target was presented to CAPRI participants as an exercise of bound docking with randomly oriented subunits and all surface side-chains truncated at Cβ. Thus, T15 was designed to test side-chain building methods as much as the docking algorithms. Unfortunately, the crystal structure of the complex became available on the Web a few days before the submission deadline of Round 5. The experiment was interrupted and models submitted after that date were not evaluated. Targets T16 and T17, two complexes that involved different xylanases bound to the same inhibitor, were cancelled for a similar reason, and

Grant sponsor: EIDIPP program of Action Concertée Incitative IMPbio.

*Correspondence to: Joël Janin, Laboratoire D'Enzymologie et Biochimie Structurales, UPR 9063, CNRS, 91198 Gif-sur-Yvette, France. E-mail: janin@lebs.cnrs-gif.fr

Received 7 January 2005; Accepted 11 January 2005

DOI: 10.1002/prot.20553

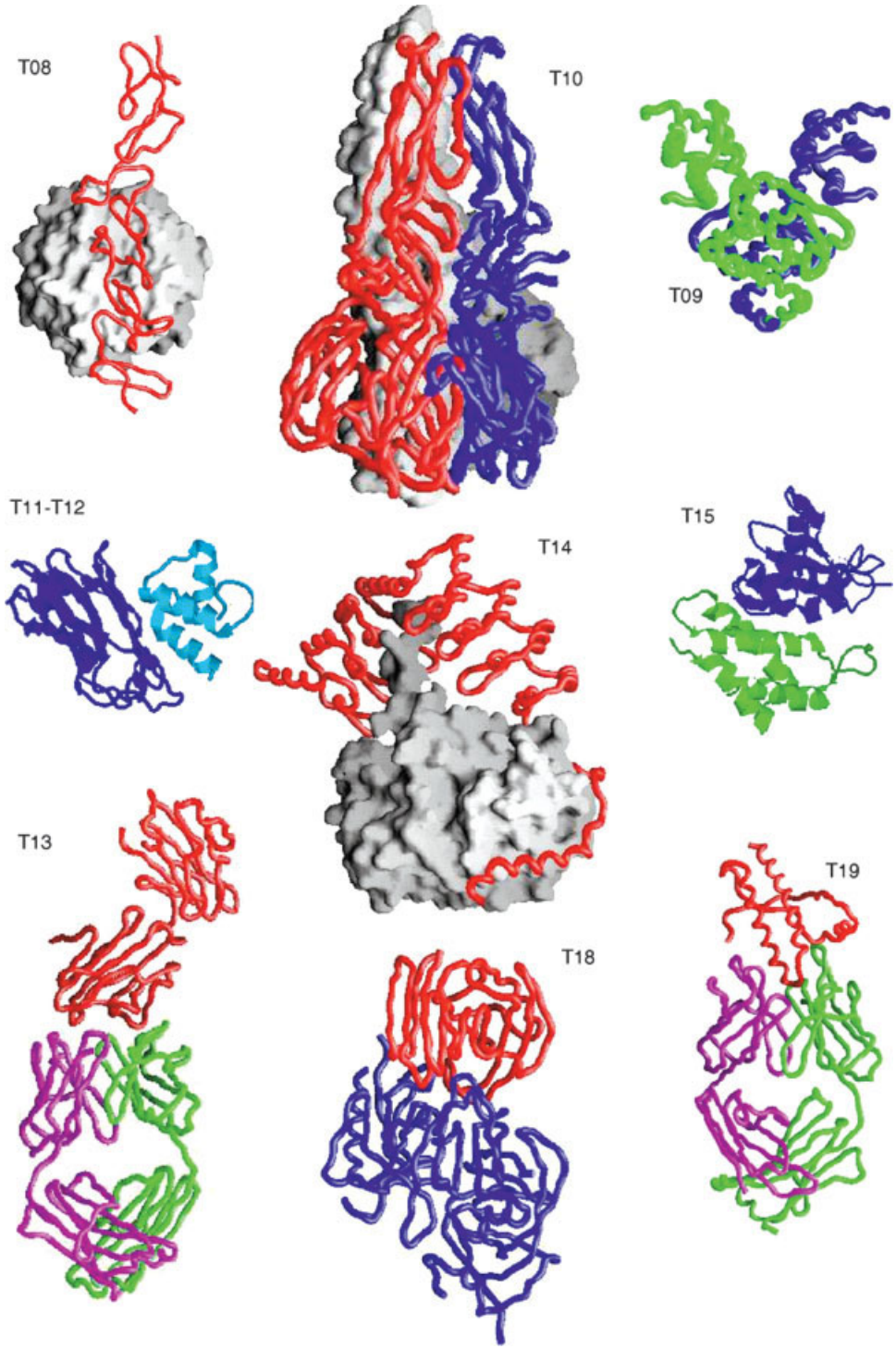

Fig. 1. The X-ray structures of the targets for Rounds 3–5. T08: Nidogen G3 (molecular surface)–laminin (red).[4] T10: Trimeric form of the tick-borne encephalitis virus envelope glycoprotein E.[5] T09: The wild-type LicT dimer.[6] T11–T12: Cohesin (blue)–dockerin (cyan).[7] T14: Protein phosphatase PP1δ (molecular surface)–MYPT1 (red).[8] T15: Colicin D (green)–Imm D (blue).[9] T13: SAG1 (red)–FAB (green/purple).[11] T18: Xylanase (red)–TAXI (blue).[12] T19: Ovine prion (red)–FAB (green/purple).[10]

only the predictions of T18, another xylanase–inhibitor complex, went to completion.

In all 4 complexes, the assumption that the inhibitor covers the active site of the enzyme was correct, but biochemical information was otherwise scarce. Bound docking in T15 gave a clear advantage to teams with experience in protein modeling. The groups of D. Baker (University of Washington, Seattle) and J. Gray (Johns

TABLE I. Characteristics of the Targets of Rounds 3–5

Target/ Reference	PDB entry	Name	Status	Target properties: RMSD[a] (Å[1])	Target properties: Interface area (Å[2])	Predictions: Submitting groups	Predictions: Model quality[b] High	Good	Acceptable
			Round 3 (Jan.–Feb. 2003)						
T08[4]	1npe	Nidogen G3–laminin	Bound–unbound	1.8	1710	16	2	9	16
T09[6]	1tlv	Wild-type LicT dimer	Oligomer prediction	12	3370/1690[c]	7	0	0	1
			Round 4 (Sept.–Oct. 2003)						
T10[5]	1urz	TBE virus protein E trimer	Oligomer prediction	11	4150/4100[d]	20	0	1	3
T11[7]	1ohz	Cohesin–dockerin	Unbound–homology	4.4	1500	19	0	11	31
T12[7]		Cohesin–dockerin	Unbound–bound	0.4	1500	22	21	0	14
T13[11]		SAG1–FAB	Unbound–bound	10	1740	21	6	6	7
			Round 5 (Jan.–March 2004)[e]						
T14[8]	1s70	Phosphatase PP1δ–MYPT1	Homology–bound	0.9	4950	25	16	20	32
T15[9]	1v74	Colicin D–ImmD	Bound[f]	(0)	1860	10	4	5	7
T18[12]	1t6g	Xylanase–TAXI	Unbound–bound	0.7	1920	26	0	6	4
T19[10]	1tpx	Ovine prion–FAB	Homology–bound	1.6	1900	24	1	10	9

[a]RMSD between the Cα atoms of the target and its components; the distance is zero for bound components.
[b]Each group could submit 10 models per target; quality criteria are defined in Méndez et al.[2]
[c]Interface area in the dimer of mutant/wild-type LicT.
[d]Interface area per subunit of the dimeric/trimeric protein E.
[e]T16 and T17 were cancelled, and only some T15 submissions were accepted due to early publication of the experimental structure.
[f]The subunit orientation was randomized and all surface side-chains truncated at Cβ.

Hopkins University, Baltimore), who are experts in that matter, obtained excellent results. They submitted a set of models that were within experimental error of the X-ray structure, accurately reproducing the contacts and side-chain conformations at the interface. The ClusPro Web server run by S. Comeau (Boston University) and the group of A. Bonvin (University of Utrecht) also submitted good models of the colicin D–ImmD complex. Other groups would presumably have done so if submissions had not been interrupted prematurely. This success on the only bound target so far in CAPRI confirms that docking procedures perform extremely well when the protein main-chain undergoes no change, but it also serves to remind us that bound docking is not a realistic prediction method.

All other CAPRI targets have had at least 1 unbound component, taken from a crystal or NMR structure of the free protein. In T18, the unbound component was the enzyme moiety of the complex, *Aspergillus niger* xylanase, whereas the *Triticum aestivum* xylanase inhibitor (TAXI) was a new structure given in its bound state for the prediction. S. Sansen and A. Rabijns (Leuwen University) who offered the target,[12] also determined an X-ray structure of the free inhibitor that shows very little change relative to the complex. TAXI is a catalytically inactive homologue of aspartate proteases. The homology might have been thought to relate to its inhibitory activity, but in practice, the face of the inhibitor that binds xylanase is opposite to the putative protease active site. This mode of binding was correctly predicted by the 6 groups that submitted good-quality models of T18 (R. Abagyan, Scripps Research Institute, La Jolla, CA; C. Camacho, Boston University; P. Fitzjohn/P. Bates, Cancer Research UK, London; I. Vakser, Stony Brook, NY; Z. Weng, Boston University; H. Wolfson, Tel Aviv University), and ranked first in at least 1 case. Thus, the prediction of T18 was rather successful, but it was not accurate. In the complex, the enzyme undergoes larger changes than the inhibitor, and loop movements affect its surface near the active site. As the models submitted by the predictors did not reproduce these changes correctly, they did not meet the stringent criteria set by the CAPRI assessors for high-quality predictions[2]: The model must contain at least 50% of the pairwise residue–residue contacts seen in the crystal structure (the native contacts), and place the regions of interaction (the epitopes) of both components within 1 Å of their native position. The best T18 submissions made the "good model" category (>30% of the native contacts, interface positioning error <2 Å), but those that had 50% of the native contacts also had many non-native contacts, and the geometry of the xylanase–TAXI interaction was mediocre.

T13–T19: Success With Antigen–Antibody Recognition

Like enzyme–inhibitor complexes, antigen–antibody complexes have long been favorite testing systems for docking procedures. Complexes of this type constituted most of the targets of Rounds 1 and 2 of CAPRI, but only 2 were offered in later rounds: T13 and T19. In all cases, the antigen moiety was unbound and the antibody moiety was the bound structure of a monoclonal antibody fragment. This implies that conformation changes in the antibody, especially movements of the complementarity determining region (CDR) loops, are ignored. As the CDR loops are known to constitute the antigen recognition site, targets of this type test the capacity of docking procedure to identify

the epitope region of the antigen surface and predict the geometry of the interaction.

In Rounds 1–2, predictor groups succeeded on 3 and failed on 2 of 5 antigen–antibody complexes.[2,3] In Rounds 3–5, they succeeded on both T13 and T19. T13, a target offered by M. Graille and F. Ducancel (CEA, Saclay, France), involves the major surface antigen 1 (SAG1) of the pathogen *Toxoplasma gondii*.[11] In T19, offered by F. Eghiaian and M. Knossow (LEBS-CNRS, Gif-sur-Yvette, France), the antigen is the globular domain of the ovine prion protein.[10] In both cases, the antigen undergoes significant changes relative to the free protein. The RMSD is large (10 Å) in SAG1, a protein with 2 flexibly connected domains, but it represents a rigid-body domain movement and does not affect the epitope, which is carried by only 1 domain of SAG1. High-quality models of this target were produced by 4 predictor groups (Bonvin, Camacho, Weng, and L. Ten Eyck, University of California, San Diego), and good models by 2 other groups. In T19, the ovine prion protein had to be model built from the NMR structure of the human homologue. The 1.6 Å RMSD cited in Table I between that structure and the bound prion, which reflects minor changes throughout the molecule, is the likely reason why only the Baker group produced a high-quality prediction of T19. Four groups submitted good models, and 5 others had models of the "acceptable" category (>10% of the native contacts, interface positioning error <4 Å). The ClusPro Web server of S. Comeau (Boston University), which cannot use biological information, made an acceptable prediction. In all these models, the epitope was correctly identified on the prion protein, but the mode of recognition by the antibody was imperfectly predicted. On the whole, these 2 targets confirm that existing procedures efficiently identify epitopes on an unbound antigen and can dock an antibody on it, but the effect of loop movements in the antibody remains to be tested.

T08–T14: Two Predictable Modes Of Association

Target T08 of Round 3 and T14 of Round 5 relate to cell biology and signal transduction. Targets of this type have been few in the past, yet docking procedures performed well on both. T08, offered by T. Springer (Harvard University, Boston), is a complex between domain G3 of nidogen and a fragment of laminin that comprises 3 EGF modules.[4] Nidogen and laminin are both large multidomain proteins, and their interaction is essential to the assembly of the basement membrane in animal tissues. The laminin fragment is a cylinder that fits in a groove on the surface of nidogen domain G3. Laminin undergoes only a local change upon binding, and domain G3 was a new structure given bound to the predictors. The complex has nanomolar affinity, the interface is comparable in size to antigen–antibody interfaces, some of the residues involved the interaction were known beforehand from biochemical data, and there is a high degree of evolutionary conservation at the interaction site. These features helped predictors to produce good models of the complex. The ClusPro Web server joined 8 other groups in achieving high- or good-quality predictions. Two groups (M. Eisenstein, Weizmann Institute, Rehovot; Gray/Baker, Seattle) submitted high-quality models as their first or second choice. In these models, the elongated laminin molecule was inclined relative to the orientation seen in crystal, but most of the pairwise contacts were correct and the epitopes were positioned to within 1 Å of their native position.

Target T14 of Round 5 also had features that helped its prediction by several groups, but compared to T08, it had the additional difficulty of requiring homology modeling. This target, offered by R. Dominguez (Boston Research Institute, Boston) is a complex between protein phosphatase-1δ (PP1δ) and the myosin phosphatase targeting subunit MYPT1, a regulatory protein that confers the enzyme specificity toward the myosin regulatory light chains during smooth muscle contraction.[8] Whereas MYPT1, a new structure, was given bound, PP1δ had to be model-built from the X-ray structure of the free PP1α isoform. PP1α and PP1δ have closely related sequences, but the free phosphatase has a disordered C-terminal segment. In the complex, this segment takes an extended conformation and makes major contacts with MYPT1. Few predictor groups attempted modeling the missing residues, and those who tried did poorly. On the other hand, residues 1–34 of MYPT1 were known from the literature to contribute to the recognition process, and the binding site of a conserved tetrapeptide had been identified. In the complex, these residues form a long arm that embraces the phosphatase. They would almost certainly be disordered in the free protein, and their presence in the bound structure certainly helped in making the docking prediction of T14 the most successful so far in CAPRI. The predictors also benefited from the very large interface formed by the 2 proteins and the fair amount of biochemical data that was available on their interaction. The best result was achieved by the Baker group, who submitted 10 essentially perfect models. Four other groups (Bonvin, Abagyan, Weng, and M. Zacharias, Bremen University) submitted at least 1 high-quality model, often ranked first, and 7 more groups made good-quality predictions.

T11–T12: Comparing Bound and Unbound Docking

Round 4 offered the opportunity to compare the performance of bound and unbound docking on the same system. The cohesin–dockerin complex of *Clostridium thermocellum*,[7] offered by A. Carvalho and M. Romao (University of Lisbon), was submitted twice for prediction as targets T11 and T12. Cohesin was an unbound X-ray structure, and in T11, dockerin had to be model-built from the NMR structure of a homologue. After T11 submissions were closed, the NMR structure was replaced by the dockerin moiety of the complex to form target T12. Dockerin and cohesin are 2 constituents of the cellulosome, a multiprotein assembly located on the bacterial cell surface and comprising a set of cellulose degradation enzymes held together by the CipA scaffold protein. Each enzyme component has a dockerin domain that interacts with one of the 9 cohesin repeats of CipA.

The interaction has been extensively studied by biochemists, but not all of that information proved helpful to the

predictors. Dockerin has approximate internal symmetry and contains 2 EF-hand Ca^{++} binding motifs, both of which have been implicated in cohesin recognition. However, only motif 2 interacts in the crystalline complex, so that the models that assumed a symmetrical mode of binding were incorrect. Still, the major problem here again was conformation changes. Only minor changes affect cohesin, but the NMR structure of dockerin used in T11 shows large helix displacements and a 4.4 Å RMSD relative to the bound molecule in T12. Although no high-quality model of T11 was produced, the prediction of the unbound target succeeded to a point, and it was a complete success on the bound target. Seven predictor groups (Abagyan, Baker, Bonvin, Gray, Fitzjohn/Bates, D. Ritchie (University of Aberdeen), and H. Umeyama (Kitasato University, Tokyo) submitted good models of T11, and 8 other groups had acceptable models. With T12, 11 groups, including the ClusPro Web server, produced high-quality predictions, and the best models (by the Baker group) were within experimental error of the X-ray structure.

T09–T10: Predicting Quaternary Structures

Targets T09 of Round 3 and T10 of Round 4 offered a completely new challenge to CAPRI participants: the prediction of subunit assembly in oligomeric proteins. The 2 targets were unbound and represented proteins that can take 2 different quaternary structures. T09, offered by M. Graille and H. van Tilbeurgh (LEBS-CNRS, Gif-sur-Yvette, France), was the wild-type Lic T antiterminator protein of *Bacillus subtilis*, the predictors being given the atomic coordinates of a mutant.[6] Lic T is a homodimer in both cases, but the subunits associate differently. T10, offered by S. Bressanelli and F. Rey (LVMS-CNRS, Gif-sur-Yvette, France), was the trimeric form of the envelope glycoprotein E of the tick-borne encephalitis virus. Protein E is a homodimer in the virus and also in crystals obtained at neutral pH, the form that was communicated to the predictors. At low pH, the dimers rearrange to form trimers that are able to interact with both the cellular and the viral lipid bilayers, promoting membrane fusion and cell infection.[5] The changes in quaternary structure are accompanied by large changes within subunits, which made T09 and T10 difficult targets for docking predictions.

The LicT protein of T09 has 3 domains, a N-terminal RNA-binding domain and 2 regulatory domains named PRD1 and PRD2. It interacts with a small antiterminator RNA to prevent the premature release of the RNA polymerase transcribing sugar utilization genes in *B. subtilis*. Lic T is inactive unless it is phosphorylated on 2 histidines of PRD2 by other components of the sugar transport system. The starting X-ray structure was a truncated form of LicT lacking the RNA-binding domain and bearing the H207D/H269D mutations that mimic in vivo the effect of histidine phosphorylation. The target was the same construct with the wild-type sequence. In the mutant, PRD1 and PRD2 of one subunit interact with their counterparts in the other subunit. In the wild-type, PRD2 undergoes a 180° rotation relative to PRD1, breaking the PRD2–PRD2′ contact and reducing the interface by half.[6] The biological effect of the point mutations, known at the time of Round 3, suggested a major change in structure, possibly a movement of the PRD2 domains bearing the mutations, but predicting its nature was a difficult task. Several groups attempted it, and the group of H. Wolfson (Tel Aviv University) did reproduce part of the movement by introducing a hinge at the right position between the domains, leading to an acceptable solution of T09, while all other predictors failed.

T10 looked even more intractable than T09, the viral protein E being much larger than LicT. Its subunits undergo large conformation changes and make completely different sets of contacts in the dimeric and the trimeric forms. The position of the C-terminal domain in the dimer interferes with the formation of the correct trimer. The predictors were informed that it should be deleted for docking, but many other changes take place in the subunit. Nevertheless, the Bonvin group found a good solution in which the geometry of the trimer was close to that of the X-ray structure, and 30% of the native residue-to-residue contacts were reproduced. Three other groups produced models with 10–20% of the native contacts, plus many non-native contacts. Albeit far from correct, such models would be useful starting points for site-directed mutagenesis, or for interpreting electron cryomicroscopy data. All correctly assumed that the trimer had 3-fold symmetry, which effectively reduces the rigid-body search from 6 to 4 degrees of freedom. Other factors that helped the predictors were the very large size of the interface and the fact that the conformation changes in the subunit were mostly domain movements.

CONCLUSION

In a couple of years, 10 protein–protein complexes have been offered to CAPRI participants for prediction by docking. They presented a variety of challenges that predictor groups have met by developing new methods, score functions, and docking procedures. No group succeeded on all targets, yet the experiment can be considered as a collective success. Moreover, the results indicate that significant progress has been made in a short time. In Rounds 1–2, prediction had failed on 3 out of 7 targets. In Rounds 3–5, high- or good-quality models were produced of all targets except T09. In most cases, these models were obtained by more than 1 method and more than 1 group. All the targets that did not show large conformation changes had high-quality predictions, and most of those with large changes had good predictions. Thus, improved docking methods can now handle such changes. The viral protein E trimer and the LicT dimer, which in previous years would have been beyond the reach of docking methods, had 1 good and several acceptable predictions. Much still needs to be done in that field, and we are still far from to fully sampling the diversity of molecular recognition by proteins in biology. This is one of several good reasons to go on with the CAPRI experiment, and to ask structural biologists to be as generous with targets in coming years as they have been in the recent past.

ACKNOWLEDGMENTS

We are grateful to all crystallographers who have provided targets for Rounds 3–5: Marcel Knossow, Frédéric Eghiaian, Herman van Tilbeurgh, Marc Graille (LEBS-CNRS, Gif-sur-Yvette, France), Felix Rey, Stéphane Bressanelli (LVMS-CNRS, Gif-sur-Yvette, France), Frédéric Ducancel (CEA-Saclay, France), Timothy Springer (Harvard Medical School, Boston, MA), Maria Romao, Anna Luisa Carvalho (University of Lisbon, Portugal), Roberto Dominguez (Boston Research Institute, Boston, MA), Stefaan Sansen, Anja Rabijns (Leuwen University, Belgium). Data in Table I are a courtesy of R. Méndez (Free University, Brussels, Belgium).

REFERENCES

1. Janin J, Henrick K, Moult J, Ten Eyck L, Sternberg M, Vajda S, Wakser I, Wodak SJ. CAPRI: A Critical Assessment of PRedicted Interactions. Proteins 2003;52:2–9.
2. Méndez R, Leplae R, De Maria L, Wodak SJ. Assessment of blind predictions of protein–protein interactions: current status of docking methods. Proteins 2003;52:51–67.
3. Wodak SJ, Mendez R. Predictions of protein–protein interactions: the CAPRI experiment, its evaluation and implications. Curr Opin Struct Biol 2004;14:242–249.
4. Takagi J, Yang Y, Liu JH, Wang JH, Springer TA. Complex between nidogen and laminin fragments reveals a paradigmatic beta-propeller interface. Nature 2003;424:969–974.
5. Bressanelli S, Stiasny K, Allison SL, Stura EA, Duquerroy S, Lescar J, Heinz FX, Rey FA. Structure of a flavivirus envelope glycoprotein in its low-pH induced membrane fusion conformation.. EMBO J 2004;23:728–738.
6. Graille M, Zhou CZ, Receveur-Bréchot V, Collinet B, Declerck N, van Tilbeurgh H. Activation of the LicT transcriptional antiterminator involves a domain swing/lock mechanism provoking massive structural changes. J Biol Chem 2005;280:14780–14789.
7. Carvalho AL, Dias FMV, Prates JAM, Nagy T, Gilbert HJ, Davies GJ, Ferreira LMA, Romao MJ, Fontes CMGA. Cellulosome assembly revealed by the crystal structure of the cohesin–dockerin complex. Proc Natl Acad Sci USA 2003;100:13809–13814.
8. Terrak M, Kerff F, Langsetmo K, Tao T, Dominguez R. Structural basis of protein phosphatase 1 regulation. Nature 2004;429:780–784.
9. Graille M, Mora L, Buckingham RH, van Tilbeurgh H, de Zamaroczy M. Structural inhibition of the colicin D tRNase by the tRNA-mimicking immunity protein. EMBO J 2004;23:1474–1482.
10. Eghiaian F, Grosclaude J, Lesceu S, Debey P, Doublet B, Tréguer E, Rezaei H, Knossow M. Insight into the PrPc→PrPsc conversion from the structures of antibody-bound ovine prion scrapie-susceptibility variants. Proc Natl Acad Sci USA 2004;100:10254–10259.
11. Graille M, Stura E, Bossus M, Muller BH, Letourneur O, Battail-Poirot N, Sibaï G, Gauthier M, Rolland D, Le Du MH, Ducancel F. Structure of the immunodominant epitope displayed by the surface antigen 1 (SAG1) of *Toxoplasma gondii* complexed to a monoclonal antibody. 2005. Submitted for publication.
12. Sansen S, De Ranter CJ, Gebruers K, Kristof Brijs K, Courtin CM, Jan A. Delcour, Rabijns A. Structural basis for inhibition of *Aspergillus niger* xylanase by *Triticum aestivum* xylanase inhibitor-I. J Biol Chem 2004;279:36022–36028.
13. Wodak S, Janin J. The structural basis of macromolecular recognition. Adv Protein Chem 2002;61:9–68.