Assessment of Blind Predictions of Protein–Protein Interactions: Current Status of Docking Methods

Raúl Méndez,[1,2] Raphaël Leplae,[1] Leonardo De Maria,[1,3] and Shoshana J. Wodak[1]*

[1]*Service de Conformation de Macromolecules Biologiques, et Bioinformatique, Centre de Biologie Structurale et Bioinformatique, CP 263, BC6, Université Libre de Bruxelles, Bruxelles, Belgium*

[2]*Grup de Modelització Estructural i Funcional de Sistemes Biològics, Institut de Neurociències, Unitat de Bioestadística, Facultat de Medicina, Universitat Autonòma de Barcelona Bellaterra, Spain*

[3]*Centro Internacional de Fisica, A.A. 4948, Bogota DC, Colombia*

ABSTRACT **The current status of docking procedures for predicting protein–protein interactions starting from their three-dimensional structure is assessed from a first major evaluation of blind predictions. This evaluation was performed as part of a communitywide experiment on Critical Assessment of PRedicted Interactions (CAPRI). Seven newly determined structures of protein–protein complexes were available as targets for this experiment. These were the complexes between a kinase and its protein substrate, between a T-cell receptor β-chain and a superantigen, and five antigen-antibody complexes. For each target, the predictors were given the experimental structures of the free components, or of one free and one bound component in a random orientation. The structure of the complex was revealed only at the time of the evaluation. A total of 465 predictions submitted by 19 groups were evaluated. These groups used a wide range of algorithms and scoring functions, some of which were completely novel. The quality of the predicted interactions was evaluated by comparing residue–residue contacts and interface residues to those in the X-ray structures and by analyzing the fit of the ligand molecules (the smaller of the two proteins in the complex) or of interface residues only, in the predicted versus target complexes. A total of 14 groups produced predictions, ranking from acceptable to highly accurate for five of the seven targets. The use of available biochemical and biological information, and in one instance structural information, played a key role in achieving this result. It was essential for identifying the native binding modes for the five correctly predicted targets, including the kinase-substrate complex where the enzyme changes conformation on association. But it was also the cause for missing the correct solution for the two remaining unpredicted targets, which involve unexpected antigen-antibody binding modes. Overall, this analysis reveals genuine progress in docking procedures but also illustrates the remaining serious limitations and points out the need for better scoring functions and more effective ways for handling conformational flexibility. Proteins 2003;52:51–67.**

Key words: structure predictions; residue-residue contacts; interface residues

INTRODUCTION

Intermolecular interactions, particularly those between proteins, have become a central focus in postgenomic biology.[1–8] Tens of thousands of gene products are known or suspected to interact with many others, from genetic, biochemical, or bioinformatics studies forming millions of putative complexes.[9–12] However, a very small fraction of these complexes will in the near future be available for analysis, let alone for crystallization. Therefore, predicted modes of association between macromolecules should be a useful guide for genetic and biochemical experiments, but the prediction methods, commonly called docking algorithms, must first be extensively tested and their validity assessed.

Docking algorithms, first suggested in 1978,[13,14] operate on the atomic coordinates of two individual proteins usually considered as rigid bodies and generate a large number of candidate association modes between them. These candidates are then ranked by using various criteria, used independently or in combination. The criteria generally include geometric and chemical complementarity measures, electrostatics, H-bonding interactions and solvation, and use empirical potentials or database-derived functions. A number of algorithms and many different scoring functions have been developed in the last 10 years, as recently reviewed by several authors,[2,15] and the field has become extremely active.

Authors developing docking procedures usually test them on crystal structures of complexes deposited in the Protein Data Bank, which so far have been protease-inhibitor and antigen-antibody complexes. In the most favorable case, the best prediction is reasonably close to

Grant sponsor: Marie Curie Host Training; Grant number: QLK3-1999-51297.

*Correspondence to: Shoshana J. Wodak. Service de Conformation de Macromolecules Biologiques, et Bioinformatique, Centre de Biologie Structurale et Bioinformatique, CP 263, BC6, Université Libre de Bruxelles, Blvd. du Triomphe, 1050 Bruxelles, Belgium. E-mail: shosh@scmbb.ulb.ac.be

Received 20 December 2002; Accepted 30 December 2002

the crystal structure,[16–27] but none of the docking procedures achieves this on all test complexes. Furthermore, the procedures perform poorly when the docking calculations use the unbound molecules, which often display some differences in conformation relative to the bound ones, either in their side-chains, or main-chain, or both.

Previous attempts critically assessed docking procedures by comparing their performance in blind trials, in which the X-ray structures of the components are known and that of the complex were made available only at the time of evaluation.[28–30] But these attempts were limited in scope, with only two available targets and less than a handful of participating groups.

Here we present the results from an independent evaluation of blind predictions of protein–protein interactions. These predictions were performed as part of the Critical Assessment of PRedicted Interactions (CAPRI), a communitywide experiment analogous to CASP,[31] but aimed at assessing the performance of procedures for predicting the mode of association of two proteins based on their three-dimensional (3D) structure.

Seven protein–protein complexes were available as targets for the first two rounds of CAPRI evaluated here. These comprised a complex between a kinase and its protein substrate, that of a T-cell receptor β-chain and a superantigen, and five antigen-antibody complexes. In several of the complexes, one of the partners was a large multimeric protein.

For each target, the experimental structures of the free components, or of one free and one bound component in a random orientation, were made available to the predictors, but the structure of the complex was revealed only at the time of the evaluation.

Nineteen groups submitted a total of 465 predicted complexes, with not all the predictor groups submitting predictions for every target. No limits were set on the source or type of additional information (homologue proteins, biochemical data on interacting residues) that the predictors could use to guide their docking calculations.

The presented evaluation was performed with no knowledge of the identity of the predictors. It is based on a number of simple criteria, agreed on by the CAPRI management team and approved by the participating groups during the first CAPRI evaluation meeting held in France on September 19–21, 2002 (see http://capri.ebi.ac.uk). The results are presented in ways that should enable easy perusal and comparison of the predictions (a set of simple numbers and pictures), with the following main goals: assessing the performance of available docking procedures, identifying useful applications of available methods in structural genomics and proteomics efforts, and evaluating the prospects of making progress in the field in the near future.

THE TARGET COMPLEXES AND THEIR COMPONENTS

The seven targets of this first CAPRI experiment are denoted T01–T07 in Table I of Janin et al. (this issue). T01 was the complex between the protein kinase from *Lactobacillus casei* HprK, and its protein substrate Hpr (from *B. subtilis*).[32] Targets T02–T06 were the following five different antigen-antibody complexes. The complex between the bovine rotavirus VP6 protein with Fab fragments from a monoclonal antibody (Rey et al., unpublished), the complex between the influenza hemagglutinin and an Fab fragment that prevents the hemagglutinin low pH transition,[33] and the complexes between porcine pancreatic α-amylase and three different variants of the single VHH domain of camelid antibodies.[34] Target 7 (T07) was the complex between the mouse T-cell receptor β-chain and the superantigen Streptoccoccal pyrogenic exotoxin A.[35]

The atomic coordinates of the targets were provided to the assessors before publication to conduct the independent evaluation reported here.

To perform the prediction for each target, predictors were given the atomic coordinates of the two components to be used in their docking calculations. The 3D structures of these components, with the corresponding PDB codes when applicable, are listed in Table I of Janin et al. (this issue). For the five antigen-antibody complexes, no 3D structures were available for the unbound Fab fragments. Therefore, the predictors were given the atomic coordinates of the bound Fab molecules, but after their overall orientation was randomized. This was not considered a serious bias given that antibodies generally undergo limited conformational changes on antigen binding.

The seven targets cover a range of biological systems and represent different degrees of difficulty. As seen from the Table I of Janin et al. (this issue), the interfaces in these complexes bury surface areas in the range of 1200–2300 Å^2, similar to those encountered in many protein–protein complexes.[36] These interfaces involve between 17 and 34 residues on each protein, and those are engaged in 37–65 residue–residue contacts. The types of residues involved in these interfaces (charged, polar, and hydrophobic) are also typical.

The HprK-Hpr complex (T01) was clearly one of the most difficult problems. The free HprK protein is a hexamer and although the complex is known to involve one Hpr molecule per kinase subunit, the predictors had to make a choice between using the entire hexamer, a trimer only, or just the monomer, with the latter being a perilous choice because it excluded the possibility for the Hpr molecule to simultaneously interact with two kinase subunits, as it is in fact the case.

Because most docking programs are still limited in their capacity in dealing with conformational adjustments, a further difficulty arose from the fact that kinase conformations in the free and bound forms differed. The main-chain of individual bound and unbound monomers superimpose with a root-mean-square deviation (RMSD) of 1.96 Å, with a significant movement of the C-terminal helix (from residue 287 on), which makes contact with the Hpr molecule in the complex. The monomers move moreover relative to one another.

In addition, the structures of the unbound HprK molecules available in the PDB (see Table I in Janin et al.) had missing coordinates for the loop comprising residues 241–

252. Therefore, predictors needed to model the structure for this loop, to avoid the risk of missing solutions that involve interactions with it.

However, there were some helpful clues as well. It was known that the phosphorylation site of the Hpr molecules was Ser-46, and not His-15, and hence that the kinase P-loop needed to be close to this residue in the cognate complex. Ample information on amino acid sequences of related kinases and Hpr molecules was also available.

The problem of conformational changes was negligible for the remaining targets. The largest structural changes between the unbound and bound molecules are for target T07 where the backbones of the variable domains (bound/unbound) display an RMSD of 1.08 Å (Table I Janin et al.). However, the large size of the virus antigen moieties in targets T02 and T03 was a drawback for most docking programs. To deal with such large molecules, predictors often use common knowledge or biochemical information to limit the search space around regions of the proteins that are most likely to interact. For these targets such regions were the complementarity determining regions (CDRs) of the antibody and the portions of the antigen that would be accessible to an antibody on the virus envelope. But restricting the search space bears the danger of missing the correct solution, as reported by some predictors (Eisenstein). As seen below, the use of prior assumptions was a serious problem for the two simpler complexes of the porcine α-amylase with the camelid VHH domains (T04 and T05), where the antibodies bind in a rather unexpected mode involving a large number of framework residues.

In terms of potentially helpful information, a low-resolution structural description[37] was available for T02, and a crystal structure of a related complex (1jck) with the same mode of interaction was available for the TCRβ-SpeA complex.

THE DOCKING PROCEDURES

Table I gives a short summary of the docking procedures and the assumptions and constraints used by the different groups in predicting the CAPRI target complexes.

A sizeable fraction of the procedures uses a cubic grid representation of the protein surface and fast Fourier transform search algorithms, following earlier work. Another involves alternative representations of the protein models and different algorithms for sampling interaction modes, which include geometric hashing, Monte Carlo, genetic algorithm and Molecular, or Brownian-type, mechanics procedures.

The criteria used to single out the correct interaction from the large number of incorrect solutions are also rather diverse. Some still rely essentially on purely geometric criteria, but the more recent methods make increasing use of composite scoring functions, which involve contributions from H-bonds, electrostatics, and solvation. However, usually these contributions are severely approximated and often used only to rescore a subset of solutions selected on the basis of geometric complementarity. But in some more recent algorithms, they are directly incorporated in the sampling step.

The ability of most procedures to handle conformational changes that may occur when the two proteins interact is still insufficient. Several allow for limited adjustments of the side-chains and backbone conformations by relaxation via molecular mechanics during a refinement step or account for flexibility by a built-in "softness" of the potential. However, some completely repack side-chains at the interface, whereas others dock sets of conformations from molecular dynamics simulations, rather than a single structure, to mimic conformational flexibility.

It is noteworthy that at least six procedures used in this first CAPRI challenge are novel or have novel aspects, and many of these are still in the development stage. Hence, their present performance is in no way indicative of their intrinsic value. It should also be emphasized that the docking procedures differ significantly in calculation speed. Some take minutes, whereas others take hours or days for the same problem. However, comparing the speeds of different procedures was not possible at this stage because of the large variety of hardware and scoring functions used.

THE EVALUATION PROTOCOL

To assess the quality of the predicted complexes, their atomic coordinates were first processed to correctly align the various chain segments to their counterparts in the targets. Procedures implemented in the BRUGEL package[39] were then used to compute a set of parameters aimed at evaluating different aspects of the predicted complexes, which might be useful for different applications. For example, accurate modeling of the predicted interface for use in designing inhibitors requires that residue–residue contacts and, if possible, atomic contacts be correctly identified, whereas information of the interface residues might be sufficient for mutagenesis experiments looking for mutants that inhibit the interaction. On the other hand, fitting an atomic model into a low-resolution electron density map derived from electron microscopy[40] is not sensitive to the exact nature of the interacting residues but rather to the overall geometry of the complex (relative orientation and position of the receptor and ligand molecules). In the following paragraphs, the parameters used in the CAPRI evaluation are described in detail.

Residue–Residue Contacts and Interface Residues

Two quantities pertaining to residue–residue contacts between the docked proteins were computed for each predicted complex. A pair of residues on different sides of the interface was considered to be in contact if any of their atoms were within 5 Å.

One quantity is the fraction of native contacts f_{nat} defined as the number of native (correct) residue–residue contacts in the predicted complex divided by the number of contacts in the target complex. The other is the fraction of non-native contacts $f_{non\text{-}nat}$, defined as the number of non-native (incorrect) residues–residue contacts in the predicted complex divided by the total number of contacts

TABLE I. Procedures for Protein–Protein Docking Used in CAPRI

Predictor	Program	Algorithm	CAPRI specifics
Scripps, US (*Abagyan*)	ICM-DISCO	Rigid body pseudo-Brownian MC with grid-based energy function; refinement: grid-based Biased Probability MC Minimization using ICM (internal coordinate mechanics).	T01: filtered for H15 or D46 in contact with phosphate and ranked according to score. T02–T06: filtered, for CDRs in contact; T07: filtered for VHH CDRs in contact, and no clashes with TCR-α.
Boston U. US (*Camacho*)	SmoothDock	Clustering, refinement and discrimination protocol: constrained minimization using CHARMM; Coulomb and solvation.	T01: restricted search to presumed phosphate binding region. T04: restricted to solutions involving CDRs. T07: ranked first solution as in homologue.
Weizmann Inst. IL (*Eisenstein*)	MolFit	Weighted geometric docking with FFT followed by clustering and filtering. Scoring by weighted geometric complementarity.	T01: up-weight conserved residues in HPrK and HPr; filtering for solutions with D46 directed to P-loop. T02–T06: up-weight interactions with CDRs; T04–T06 filter against solutions with non-CDR contacts. T07: homologue used, down-weight residues in TCRα/β contacts.
Cancer Research UK LRI (*Fitzjohn / Bates*)	Guided Docking	Rigid body molecular mechanics using CHARMM22 force field. Flexible refinement step in some cases. Solvation used in final energy ranking.	Only T04–T07 predicted. Partial manual ranking, T07: homologue (1JCK)used
Sheffield U. UK (*Gardiner*)	GAPDOCK	GA for sampling different relative rotations and clustering. Scoring based on shape correlation and buried area, as per CCP4 package; some clash checks.	T01, T02 only. Search restricted to expected contact regions. Final selection, manual. T01: selected solutions with Ser 46 contacting ligands on kinase. T02: limited search space, and selected solution containing CDRs.
U. Washington US (*Gray / Baker*)		Monte Carlo and clustering, side-chain repacking, rigid body minimization. Fitted multiterm scoring function, mainly vdW orientation-dependent hydrogen bonding; implicit solvation.	T01–T03: no minimization; T01: search restricted to HPr. Ser46-HPrK. Asp157 in contact, kinase conserved residues. T02–T06: favored contacts with CDRs; T02: only outer cap.; T07: used homologue, manual selection except for model1
Aberdeen U. UK (*Mustard*)	CONCOORD[43]+*Hex*	Dock 500 essential dynamics structures from CONCOORD using *Hex*.	T07: 500 SpeA structures docked onto rigid TCR. Search constrained to TCR hypervariable loops.
Columbia U. US (*Norel*)	PPD	Geometric hashing; multiple rescoring.	
Scripps US (*Olson / Norledge*)	Surfdock[46]	Fourier correlation of spherical harmonics. Scoring by electrostatic, hydrophobic, van der Waals (buried surface area) interactions and shape complementarity. Clashes determined by overlap of the docked surfaces.	Only T01–T03 predicted. Manual inspection of top scoring solutions. T01: Hpr S46 constrained near kinase active site + manual check of kinase binding region. T02–T03: Ag epitopes from other structures. Penalization of Ag regions inaccessible in virion. Manual filtering of symmetric solutions.
Universidade Nova De Lisboa (*Palma / Krippahl*)	BiGGER (Chemera)	Binary grids; scoring with geometric, contacts counts, electrostatics and solvation, no clash checks.	Some manual ranking. T07: TCR. truncated at D118
Aberdeen U. UK (*Ritchie*)	*Hex*	Spherical polar Fourier correlation of shape and electrostatics, followed by soft rigid body OPLS minimization.	T01: search restricted to HPrK P-loop. T02–T07: search restricted to antibody and TCR hypervariable loops. Visual inspection and selection of orientations for T01–T03.
North Western U. US (Shoichet)	Northwestern DOCK	Search for hot-spot correspondences between receptor and ligand to calculate orientations, precalculation of ensemble of ligand conformations, receptor held rigid, energy evaluation using electrostatics and van der Waals terms.	Only T01 predicted
Imperial Coll. UK (*Sternberg)*	3D-DOCK MULTIDOCK	FFT; rescore: residue potentials; flexible refinement: mean-field side-chain multicopy, solution clustering	Search restricted to expected interacting regions in all cases (CDRs for targets T02–T06); manual selection from highest ranking solutions.
UCSD US (*Ten Eyck*)	DOT	FFT for shape complementarity and Poission–Boltzmann electrostatics.	Ranking by shape complementarity (FADE). T02–T07 screened for CDRs. Manual ranking for T01 and T02. Cluster analysis for T04–T07.
Kitasato U. JP (*Umeyama / Komatsu*)	TSCF	Solvent cluster fitting; refinement by molecular mechanics and molecular dynamics using AMBER force field.	No biochemical data, or bioinformatics used. Some manual ranking.
SUNY/StonyBrook, US (*Tovchigrechko, Vakser*)	GRAMM	FFT for shape complementarity; softer potential for conformational changes; no clash check; symmetry of multimeric receptors used to enhance sampling; docking of ligand to overlapping fragments of receptor for speed.	Only T01–T03 predicted. T01: filtered for HPr H15 or D46 close to kinase active site. T02[1]: filtered for for CDRs[44] in contact + constraints on epitope residues. T03: Filtered for CDRs[45] in contact + constraints on epitope
U. Autonoma Madrid, SP (*Valencia*)		Neural network-based predictions of interactions sites, using information on related sequences.	Only T01 predicted
Boston U. US (*Weng*)	ZDOCK	FFT with scoring by pairwise shape complementarity, solvation and electrostatics; clustering.	Blocked non-CDR residues for T02–T06. T01: Ser 46 of Hpr within 7 Å from P-loop. T02: distal half of VP6. T07: locked residues according to homology; manual selection from best solutions
Tel Aviv U., IL NIH, US (*Wofson / Nussinov*)	PatchDock (T04–07) BUDDA (T01–07) PPD (T01)	Geometric docking: matching of local shape features and geometric hashing, fast geometric scoring and search, avoids exhaustive orientation search.	T01: HPr Ser46 close to P on kinase. T02–T06: CDRs only. T04–T06: preference for high-sequence variability regions in mammalian amylases. T07: used interface from homologue.

[1]Figure 6 of this reference.[44]

The left-most column lists the affiliation of the predictor group and the group principal investigator(s). The second column gives the acronym of the docking software. Column 3 lists key features of the docking algorithm and the right-most column lists details on the various assumptions, constraints, and additional information used in applying the different procedures to the CAPRI targets. The format of this table was inspired from (http://www.bmm.icnet.uk/~smithgr/soft.html) with permission from G. Smith.

in that complex. This latter quantity is informative because a predicted complex in which a small number of native contacts is embedded in a large number of nonnative ones is much less useful than a complex in which the number of false-positive contacts is small.

When several quasi-symmetrical docking solutions were submitted, as was the case for targets T01–T03 in which one of the components is a multisubunit protein, f_{nat} was evaluated for all the submitted solutions and the one with the largest f_{nat} value was used for further analysis.

To evaluate the extent to which a prediction identifies the native interaction surfaces or "epitopes" on either or both proteins, independently of its ability to produce the native contacts between them, we also computed the fraction of native interface residues f_{IR}. This fraction, computed for each of the interacting proteins, is defined as the number of native residues in the predicted interface divided by the total number of native residues in the corresponding target interface. Here, the interface residues were defined as those that lose accessible surface area when the two proteins associate.

Backbone RMSDs and Rigid Body Transformations

To evaluate the overall geometric fit between the 3D structures of the predicted and observed complexes, several parameters were computed. One is the RMSD of the ligand (the smaller of the two proteins) in the predicted versus target complexes after the receptor (the larger of the two proteins) was superimposed.[41] This ligand RMSD, denoted L_RMS, and the superpositions were both computed on backbone atoms (N,Cα,C,O). To further characterize the global fit between the predicted and observed ligand positions, we also computed the residual rigid body rotation angle θ_L and the residual translation vector of the geometric centre d_L required to superimpose the ligand molecules once the receptors have been superimposed.

Because some predictors chose to perform their docking calculations by using only part of the component molecules (e.g., the hypervariable domains of the Fab fragments), the L_RMS values were computed for the molecular portions that were common to all the predictors. In target T07, where the two domains of the TCR receptor move relative to one another on complex formation, the L_RMS was computed on the backbone of the variable domain only.

When several quasi-symmetrical docking solutions were possible or actually submitted, as was the case for targets T01–T03, the receptor subunit involved in the maximum number of residue–residue contacts with the ligand was successively superimposed onto the different receptor subunits of the target. The superposition yielding the best fit of the ligands was then used to computed the L_RMS and other related parameters.

The L_RMS is a global measure that depends on the size of the ligand. Thus, it may not always provide a good picture of the fit at the interface, especially when the ligand is large and oriented slightly differently in the predicted model than in the target. Therefore, it was agreed to evaluate also the backbone RMSD of interface residues only (I_RMS) and use it as a measure of how the model fits the target structure in the interface region, realizing well that such measure includes contributions from rigid body motion as well as local conformational changes. To that end, the interface residues in the target were redefined as those having at least one atom within 10 Å of an atom on the other molecule. This is twice the threshold used in defining the residue–residue contacts described above. The backbone of these residues were then superimposed on their equivalents in the predicted complex to compute the I_RMS.

TABLE II. Criteria for Ranking the CAPRI Predictions

Rank	f_{nat}	L_rms	or I_rms
High	≥0.5	≤ 1.0	or ≤ 1.0
Medium	≥0.3	1.0 < x ≤ 5.0	or 1.0 < x ≤ 2.0
Acceptable	≥0.1	5.0 < x ≤ 10.0	or 2.0 < x ≤4.0
Incorrect	<0.1		

Column 1 lists the rank of the predictions, and columns 2 and 3 list the parameter ranges defining the ranks. These parameters are the fraction of native contacts f_{nat} defined as the number of native residue–residue contacts in the predicted complex divided by the number of native contacts in the target (see text). L-rms is the backbone root-mean-square displacement of the ligands in the predicted versus the target structures computed after the receptors of these structures have been superimposed, and the corresponding coordinate transformation has been applied to superimpose the two complexes. The I-rms, is the root-mean-square displacement of the backbone of the interface residues only, in the predicted versus the target complexes (see text).

Atomic Clashes

Early during the analysis, it became obvious that several prediction procedures produced complexes that featured a very large number of atomic clashes due to appreciable interpenetration of the two molecules and that such interpenetration often artificially increased the fraction of native contacts f_{nat}. Therefore, it was agreed to impose a threshold on the allowed number of clashes for the predictions submitted for a given target. To that end, we computed the average and standard deviation of the number of such clashes in all the predictions for each target and rejected from any further assessment predictions where the number of clashes was 2 SD away from the average. Clashes were defined as contacts between nonhydrogen atoms separated by <3.0 Å.

Ranking the Predictions

As initially suggested by members of the CAPRI management committee, a major criterion for ranking the predicted complexes was the fraction of native residue–residue contacts (f_{nat}). But because this quantity is sensitive to the number of atomic clashes and depends on the chosen distance cutoff parameter, a detailed evaluation was performed on all the predictions with $f_{nat} \geq 0.1$, whereas those with lower f_{nat} values were not evaluated further because they were considered to represent random solutions.

The evaluated predictions were in turn grouped into four categories on the basis of the three parameters f_{nat}, L_RMS, and I_RMS. These categories are highly accurate, medium accuracy, acceptable, and incorrect, with predic-

TABLE III. CAPRI Prediction Results for Individual Targets

A. Target T01
Lactobacillus HPrK - *B. subtilis* HPr complex

Predictor groups	16
Evaluated predictions	69
High accuracy(***)	0
Medium accuracy(**)	0
Acceptable (*)	8
Incorrect	55
Predictions with clashes	6
Average no. of clashes (SD)	80.4 (84.3)

Model no. (category)	Predictors	f_{nat}	$f_{non\text{-}nat}$	f_{IR_L}	f_{IR_R}	N_{clash}	L_rms (Å)	I_rms (Å)	θ_L (°)	d_L (Å)
5 (*)	U Sheffield UK (*Gardiner*)	0.3	0.9	0.5	0.6	215	12.0	3.1	49.4	9.5
1 (*)	Weizmann Inst. IL (*Eisenstein*)	0.2	0.8	0.6	0.4	27	7.5	2.5	34.1	4.8
1 (*)	U. Autonoma Madrid SP (*Valencia*)	0.2	0.9	0.7	0.8	140	8.5	3.9	47.9	3.7
1 (*)	Boston U. US (*Camacho*).	0.2	0.9	0.7	0.6	103	9.4	4.0	56.6	2.4
3 (*)	Tau IL/NIH US (*Wolfson*)	0.2	0.8	0.7	0.8	45	10.2	3.4	43.1	7.6
1 (*)	Scripps US (*Olson*)	0.1	0.9	0.6	0.6	76	10.0	3.3	44.9	7.6
3 (*)	UCSD US (*Ten Eyck*)	0.1	0.9	0.5	0.3	46	15.9	3.9	64.1	12.2

B. Target T02
Bovine Rotavirus VP6-FAB complex

Predictor groups	15
Evaluated predictions	70
High accuracy (***)	0
Medium accuracy(**)	1
Acceptable (*)	6
Incorrect	60
Predictions with clashes	3
Average no. of clashes (SD)	59 (46.6)

Model no. (category)	Predictors	f_{nat}	$f_{non\text{-}nat}$	f_{IR_L}	f_{IR_R}	N_{clash}	L_rms (Å)	I_rms (Å)	θ_L (°)	d_L (Å)
3 (**)	Boston U. US (*Weng*)	1.0	0.3	1	1	86	2.3	1.2	4.8	2.1
3 (*)	U. Sheffield UK (*Gardiner*)	0.5	0.6	0.8	0.8	84	9.2	2.4	29.9	6.1
3 (*)	SUNY/MUSC US (*Vakser*)	0.4	0.6	0.8	0.8	12	12.9	3.9	38.8	8.1
1 (*)	Weizmann IL (*Eisenstein*).	0.4	0.7	0.8	0.9	33	8.5	3.4	15.5	7.5
1 (*)	Imperial Coll. UK (*Sternberg*)	0.4	0.4	0.6	0.8	8	11.8	3.7	36.7	7.0
4 (*)	UCSD US (*Ten Eyck*)	0.3	0.8	0.9	0.9	99	12.9	4.0	31.6	10.2

TABLE III. (Continued)

C. Target T03
Influenza hemagglutinin-FAB complex

Predictor groups	13
Evaluated predictions	62
High accuracy(***)	0
Medium accuracy(**)	2
Acceptable (*)	0
Incorrect	58
Predictions with clashes	2
Average no. of clashes (SD)	138.3 (124.4)

Model no. (category)	Predictors	f_{nat}	$f_{non\text{-}nat}$	f_{IR_L}	f_{IR_R}	N_{clash}	L_rms (Å)	I_rms (Å)	θ_L (°)	d_L (Å)
2 (**)	Scripps US (*Abagyan*)	0.7	0.3	0.5	0.7	22	4.6	1.9	11.7	3.4
4 (**)	Aberdeen U. UK. (*Ritchie*)	0.7	0.5	0.5	0.8	100	7.4	1.7	20.4	6.1

D. Targets T04,T05
Porcine α-amylase-CAbAM-D10 complex
Porcine α-amylase-CAbAM-07 complex

	T04	T05
Predictor groups	13	13
Evaluated predictions	65	64
High accuracy (***)	0	0
Medium accuracy (**)	0	0
Acceptable (*)	0	0
Incorrect	60	60
Predictions with clashes	5	4
Average no. of clashes (SD)	34.2 (37.4)	37.9 (37.4)

E. Target T06
Porcine α-amylase-CAbAM-D9 complex

General summary

Predictor groups	13
Evaluated predictions	65
High accuracy (***)	4
Medium accuracy (**)	4
Acceptable (*)	0
Incorrect	54
Predictions with clashes	3
Average no. of clashes (SD)	34.5.7 (32.2)

Model no. (category)	Predictors	f_{nat}	$f_{non\text{-}nat}$	f_{IR_L}	f_{IR_R}	N_{clash}	L_rms (Å)	I_rms (Å)	θ_L (°)	d_L (Å)
2 (***)	Imperial Coll. UK (*Sternberg*)	0.9	0.2	0.9	1.0	24	0.8	0.6	2.7	0.6
1 (***)	Boston U. US (*Camacho*)	0.8	0.2	0.9	0.9	10	2.4	0.8	6.3	2.1
5 (***)	Aberdeen U. UK (*Ritchie*)	0.8	0.1	0.9	0.9	20	2.2	0.8	7.8	1.5
5 (**)	UCSD US (*Ten Eyck*)	0.8	0.3	0.9	1.0	59	2.0	1.1	8.9	1.1
1 (**)	U. Washington US (*Gray/Baker*)	0.7	0.3	0.9	0.9	28	1.5	1.1	3.0	1.3
1 (***)	Scripps US (*Abagyan*)	0.7	0.4	0.9	0.9	4	1.3	0.9	5.6	0.7
5 (**)	U. Lisbon PT (*Palma/Kripphal*)	0.7	0.6	0.8	0.8	62	6.2	1.6	19.1	5.0

TABLE III. (Continued)

F. Target T07 TCRβ-SpeA complex	
Predictor groups	14
Evaluated predictions	70
High accuracy (***)	5
Medium accuracy (**)	7
Acceptable (*)	8
Incorrect	47
Predictions with clashes	3
Average no. of clashes (SD)	33.8 (44.9)

Model no. (category)	Predictors	f_{nat}	$f_{non\text{-}nat}$	f_{IR_L}	f_{IR_R}	N_{clash}	L_rms (Å)	I_rms (Å)	θ_L (°)	d_L (Å)
1 (**)	Boston U. US (*Weng*)	0.8	0.4	1.0	0.9	42	3.9	1.2	12.0	3.3
1 (***)	Weizmann IL (*Eisenstein*)	0.8	0.2	1.0	0.8	12	2.1	0.9	7.2	1.3
1 (***)	Cancer Research UK (*Fitzjohn/Bates*)	0.8	0.2	1.0	0.9	3	2.6	1.0	10.8	1.7
1 (***)	Boston U. US (*Camacho*)	0.8	0.1	1.0	0.9	2	1.5	0.9	7.0	0.4
1 (***)	Tau IL/NIH US (*Wolfson*)	0.7	0	1.0	0.9	5	1.3	0.6	3.4	0.6
1 (***)	U. Washington US (*Gray/Baker*)	0.6	0.2	0.9	0.9	5	2.2	0.7	7.1	1.5
5 (**)	Scripps US (*Abagyan*)	0.6	0.5	1.0	0.9	3	6.0	1.9	22.6	3.9
1 (*)	U. Lisbon PT (*Palma/Krippahl*)	0.5	0.4	0.9	0.8	13	7.5	2.3	31.6	4.8
3 (*)	Imperial Coll. UK (*Sternberg*)	0.3	0.8	0.6	0.8	15	8.5	3.3	38.1	4.5

Sections (a)–(f) of this table are devoted to the results for individual targets T01–T07. Each section is divided into two main parts. The top part gives a general summary of the predictions, and the bottom part lists the key parameters of the best predictions ranked as acceptable or higher submitted by each group.

The submitted predictions were divided into four categories as detailed in Table II. Predictions with a number of clashes exceeding a defined threshold were not evaluated. Clashes were defined as those between two non-hydrogen atoms on each side of the interface whose distance was $<$ 3 Å. The threshold was taken as 2 SD, plus the average of the number of clashes in all the predictions submitted for a given target.

Detailed results for the best predictions for each participant, which were of acceptable quality or better (bottom), ranked as indicated in Table II. Column 1 lists the model number (1–5) and the rank of the prediction (high accuracy (***), medium accuracy (**), and acceptable (*). Column 2 lists the participant groups in order of decreasing native contact fraction f_{nat} (column 3). Column 3 lists the fraction of non-native contacts $f_{non\text{-}nat}$, defined as the number of non-native contacts over the total number of contacts in the predicted complex. $f_{IR\_L/R}$ is defined as the number of native residues in the predicted interface over the total number of interface residues in the target, computed for both the R (receptor) or L (ligand) molecules. Column 7 (N_{clash}) lists the number of atomic clashes in the predicted complex. Columns 8 and 9 list the RMSD values. L_RMS is the backbone RMSD (Å) of the ligand molecules in the predicted versus the target complexes after the receptor moieties have been superimposed. The I_RMS is the interface RMSD (Å) computed by superimposing only the backbone of the interface residues from the target complex onto their counterparts in the predicted complex. The last two columns list the residual rigid body rotation (θ_L) and translation (d_L) of the ligand in the predicted versus the target complexes after the corresponding receptor molecules have been superimposed. For further details on how the various parameters were computed, see the text.

tions of the latter category being considered as equivalent to random solutions. The parameter ranges used to define these categories are summarized in Table II. As expected, we could observe a correlation between the different parameters, particularly between the native contact fraction and the I_RMS, although mainly for predictions reasonably close to the correct solution. But this correlation tended to break down for poorer predictions.

RESULTS AND DISCUSSION

In this section, the CAPRI evaluation results are presented as follows. First, we describe the results obtained for individual targets. Second, an overview is presented of the results across predictors and targets, respectively, and third, an attempt is made to relate the results of the different groups with the methodology used.

A detailed description of all the parameters and all the results obtained for the seven targets can be found on the CAPRI web site (http://capri.ebi.ac.uk).

Predictions for Individual Targets

Target T01: HPr-K/P-Hpr complex

Table III(a) summarizes the prediction data for Target 01, the Hpr-K/P-Hpr complex. Sixteen groups submitted a total of 69 nonredundant predictions for this target (predictions not related by quasi-symmetry). Of these, six had too many atomic clashes (more than the threshold of 228 atom–atom contacts within 3 Å) and were not analyzed further. Of the remaining 63 predictions, only 8 were of acceptable quality (with $f_{nat} \geq 0.1$, L_RMS $\leq$ 10 Å and I_RMS $\leq$ 4 Å), and none were of medium or high accuracy.

The best of the acceptable predictions obtained for this target by each group and their quality measures are listed in Table III(a) (bottom). We see that the prediction with the highest native contact ratio (f_{nat} = 0.3), by U. Sheffield (*Gardiner*), has an appreciable number of clashes (215), and a large fraction of non-native contacts ($f_{non\text{-}nat}$ = 0.9). It also features a quite large L_RMS (12 Å) of the Hpr molecule whose overall rotation and translation are shifted by ~49° and 9.5 Å, respectively [last two column of Table III(a)]. Despite this, the I_RMS value is relatively low (~3.1 Å), reflecting the fact that it measures the fit of the predicted versus target backbones for interface residues only and is not affected by parts of the molecule far way from the interface.

One of the best predictions listed in Table III(a) is that by the Weizmann (*Eisenstein*) group. However, this prediction has a somewhat lower native contact fraction (f_{nat} = 0.2) but also a slightly lower, although still too high, fraction of non-native ($f_{non\text{-}nat}$ = 0.8). It has furthermore significantly lower RMSDs (L_RMS = 7.5 Å, I_RMS = 2.45 Å), the deviation of the rigid body rotation and translation parameters of the Hpr molecule are smaller (34° and 4.8 Å), and the number of clashes is low (27). Except for the poorer prediction, at the bottom of the list, the remaining ones listed in Table III(a) are of roughly comparable quality, although several have a notable number of clashes.

A pictorial overview of all the predictions provided for this target is presented in Figure 1. This figure shows the positions of the ligand HPr molecules and their geometric centers relative to the receptor surface in the target (Fig. 1), and the same centers in all the predicted complexes after the receptor backbones in these complexes have been superimposed onto their counterpart in the target, as described above [Fig. 1(b)]. For the acceptable predictions, the geometric centers of the ligand cluster more clearly around the observed positions, whereas in incorrect predictions ($f_{nat} \leq 0.1$) they sample a much wider range of positions, corresponding to random solutions. Some of the depicted solutions lie near the three-fold axis of the kinase trimer or at the interface between the top and bottom trimer layers.

Thus, we see that even the best predictions for this target are of relatively poor quality, as judged by most of the parameters listed in Table III(a). The fact that f_{nat} is between ~0.1 and 0.3 in these predictions and that the false-positive contacts represent as much as 80% of the predicted ones in the best prediction, suggests that the models would not be too useful for any detailed analysis of the interactions. However, considering the large size of the receptor and the appreciable difference between the bound and unbound kinase conformations, the fact that some solutions approach, albeit imperfectly, the correct mode of association should be viewed as very encouraging. On the other hand, it is important to mention that this encouraging, although limited, success is primarily the fruit of using biochemical data (on the residues expected to be in contact) or clues from patches of conserved residues in multiple alignments of related kinases and Hpr proteins to guide or filter the docking predictions (Table I and other predictor reports in the present issue). An extreme case of this approach is the prediction by the Madrid (*Valencia*) group, which involved no docking calculations at all but was entirely based on their neural network analysis of the residue conservation patterns, and used standard modelling to generate the structure of the complex.

Target T02: Rotavirus VP6-Fab Complex

Table III(b) summarizes the prediction data for this target. Fifteen groups submitted a total of 70 predictions. Of those, six were of acceptable quality, one was of medium accuracy, and none of high accuracy. Three predictions were rejected from the evaluation because of too many clashes.

The best prediction for this target, by the Boston U. (*Weng*) group [bottom part of Table III(b)], is of medium accuracy. It retrieves all the native contacts (f_{nat} = 1), together with about 30% of non-native contacts. The L_RMS and I_RMS for this predictions are 2.33 Å and 1.16 Å, respectively, indicating a rather good fit of the predicted and target ligand molecules, as well as a close to excellent fit of the backbone of the interface residues. The rigid body rotation and translation of the ligand are also correspondingly small (4.8° and 2.05 Å). Only the number of clashes is a little high (87).

All the other predictions listed in Table III(b) are of acceptable quality only. They have f_{nat} values between 0.3 and 05, I_RMS values between 2.4 and 4 Å and L_RMS values in the range of 8.5–12.9 Å.

In all the predictions for this target, the requirement to involve the CDRs of the Fab molecule in the interaction was used to limit the search space or to filter solutions (Table I and predictor reports in this issue). Several of the predictors also restricted the search to regions of the receptor that were more likely to be accessible to antibodies on the virus envelope. However, these restrictions were obviously not sufficient to direct most of the prediction procedures to correct solution. Therefore, the single medium accuracy prediction for this complex should be viewed as quite an achievement.

Figure 2 presents a pictorial overview of all the predictions for this target, which illustrates well the differences in the positions of the geometric centres of the correct and incorrect solutions, and particularly the close fit of the geometric centres of the single medium accuracy prediction for this target.

Target T03: Influenza Hemagglutinin-Fab Complex

Table III(c) summarizes the prediction data for this target. Thirteen predictor groups submitted a total of 62 nonredundant predictions. Of those, two were eliminated because of clashes, two were ranked medium accuracy, and the rest as incorrect. Key parameters of the two correct predictions are listed in bottom part of Table III(c).

The best prediction, by the Scripps (*Abagyan*) group, identifies 70% of the native contacts, and the fraction of non-native contacts in the predicted structure is rather low (30%). The RMSD values (L_RMS = 4.6 Å, I_RMS = 1.9 Å) and the rigid body shifts (11.7°, 3.4 Å) are

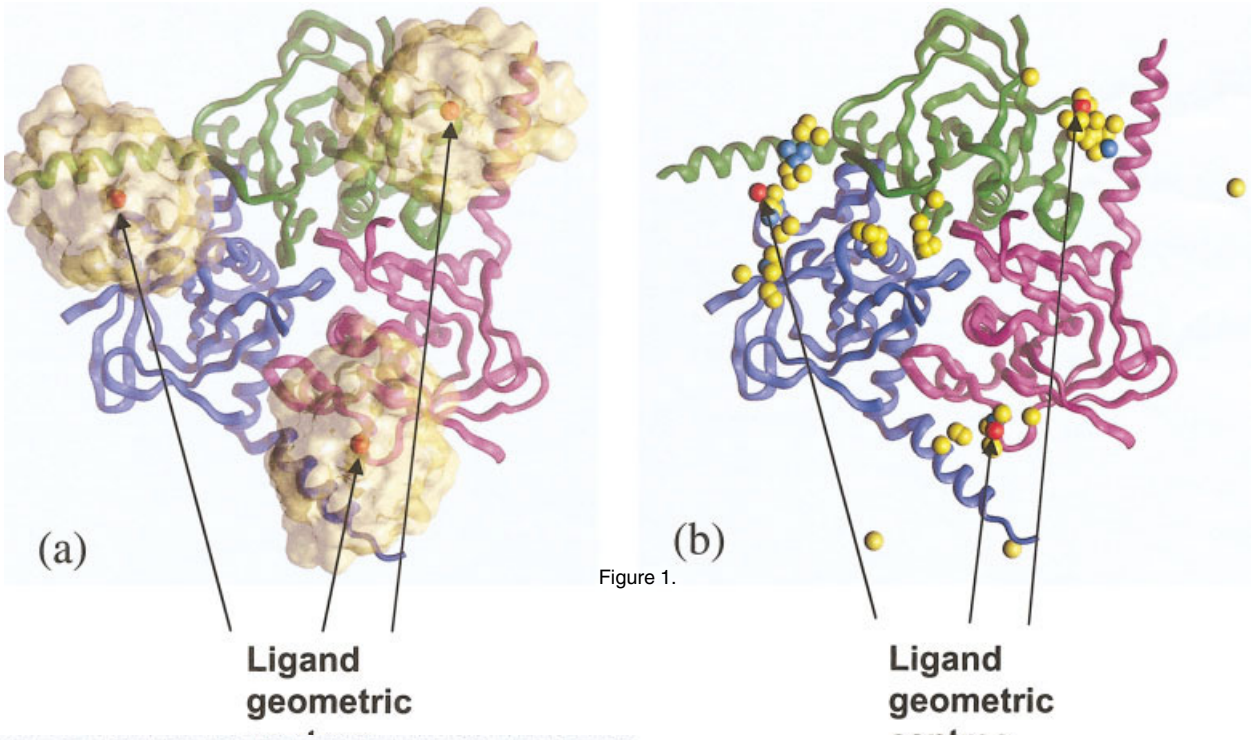

Figure 1.

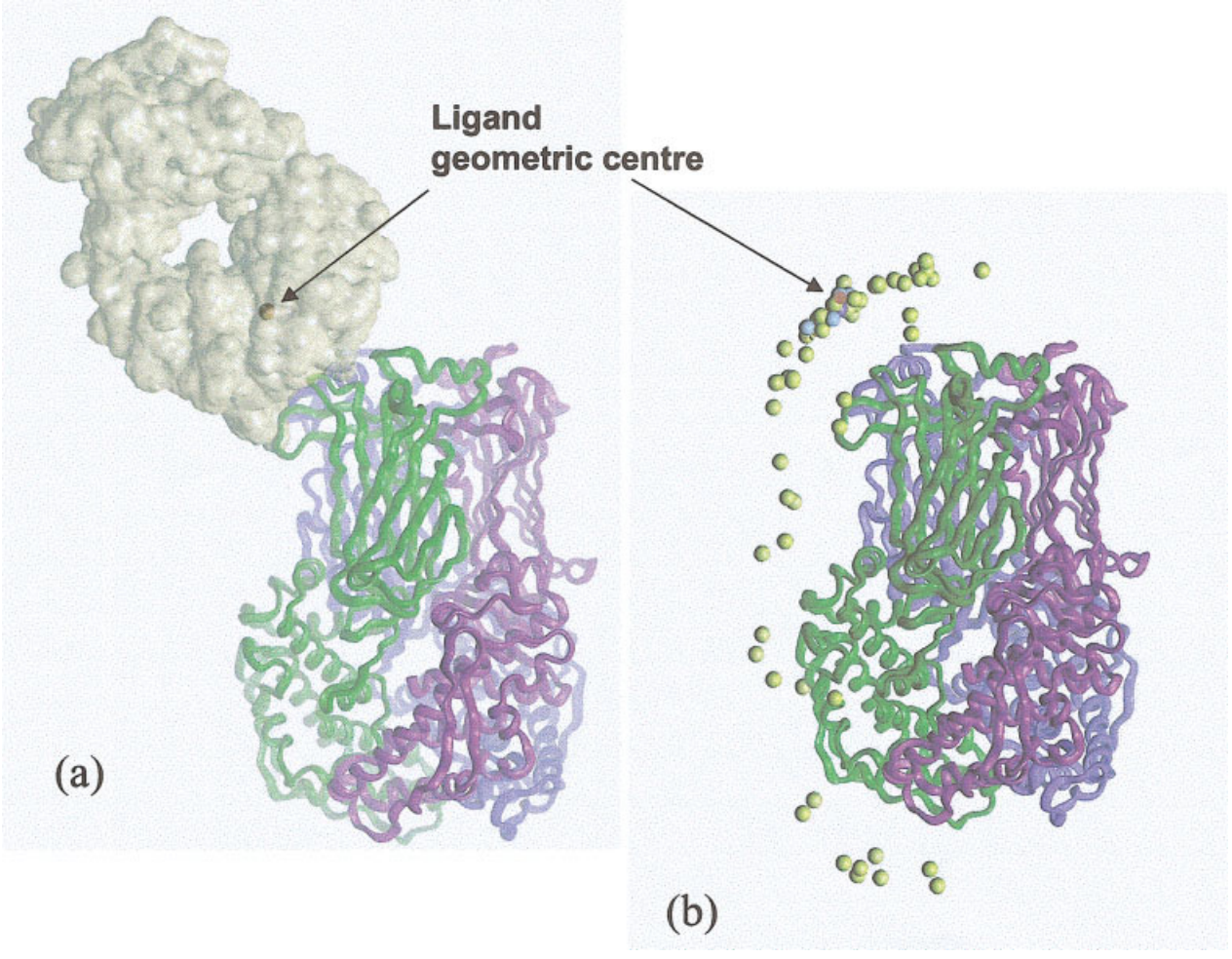

Figure 2.

relatively small too, with only 22 clashes. The second acceptable prediction has the same fraction of native contacts, but the fraction non-native contacts is higher (50%), and so are the RMSD values and the number of clashes. It is rather reassuring to see that both predictions are of higher quality than the blind predictions submitted for a similar target in a previous contest 6 years earlier.[30]

Two other groups submitted one prediction each that qualified as acceptable, but these predictions were not assessed here because they were not among the top five solutions submitted by these groups.

Here too, most predictors limited the search regions to the loops of the hypervariable regions of the antibody and to the accessible portions of the antigen. But as for target 2, the correct solutions certainly cannot be attributed solely to these restrictions, especially because virtually all the predictors used them.

A pictorial presentation of the target and a summary of the predictions for this target is presented in Figure 3. It shows that the scatter of the ligand geometric centers is similar to that in target T02.

Target T04-T06: Complexes Between Pancreatic α-Amylase and Camelid Antibodies

Respectively, 65 and 64 predictions were submitted for targets T04 and T05. All of them were qualified as incorrect, with only three predictions of target T05 having $0 < f_{nat} \leq 0.1$, and the remaining ones all had $f_{nat} = 0$ [Table III(d)].

This poor result is primarily due to the fact that the intermolecular contacts in the crystal structure involve the framework regions of the camelid VHH domain and not mainly the CDR region, as expected. Indeed, most predictors either restricted the search space to solutions involving only the CDR regions or filtered out solutions that did not involve them in a postprocessing step, thereby effectively eliminating any complexes with alternative interaction modes. It is of interest that a handful of submitted solutions did involve the CDR regions of the VHH domain and the amylase active site and thus provided acceptable predictions for target T06, which does in fact involve such interaction (see below). But these solutions were not assessed here. In contrast to the previous targets, these negative results illustrate how restrictions and filters applied on the basis of prior knowledge can just as effectively confine the search to incorrect solutions.

A much better performance was observed for Target T06, which features the expected mode of interactions with the VHH domain. Thirteen groups provided a total of 65 predictions for this target [Table III(e)] and three were rejected as having too many clashes. As many as four predictions were ranked as highly accurate, four as medium accuracy models, and the remainder as incorrect solutions.

The best prediction of acceptable quality or better provided for this target by each group, together with their quality measures, are listed in Table III(e). The most accurate prediction of the lot, and for that matter of the entire challenge, was that by the Imperial Coll. (*Sternberg*) group. This highly accurate prediction retrieves 90% of the native contacts, with a mere 20% of false positives, has RMSD values below 1 Å (L_RMS = 0.8 Å, and I_RMS = 0.6 Å) and only 24 clashes. Three other predictions in Table III(e), by Boston U. (*Camacho*), Aberdeen U. (*Richie*), and Scripps (*Abagyan*), are also of high accuracy. They identify between 70 and 80% of the native contacts and have only 10–40% false-positive contacts. Although their L_RMS's range between 1.3 and 2.4 Å, their I_RMSs are all <1 Å. Therefore, all are rather accurate predictions by our standards.

The remaining three predictions listed in Table III(e) are of medium accuracy and hence also rather good. Thus, looks like submitted solutions for this target were either correct or random solutions, with few solutions in between. A possible reason for this two-tier performance might be that in this case the receptor is a single protein whose surface is less complex and features fewer major grooves than the surfaces of the multiprotein receptors of targets T01–T03. In addition, some predictors may have introduced additional constraints on the basis of evidence from biochemical studies[34] that, unlike the two monoclonal camelid antibodies of targets T04 and T05, this particular one inhibits enzymatic activity and is hence likely to bind to the enzyme active site.

Figure 4 highlights the ligand geometric centre positions in target T06 [Fig. 4(a)] and provides a pictorial summary of the predictions [Fig. 4(b)], which illustrate rather well the good fit of the geometric centres.

Target T07: TCRβ-SpeA Complex

Fourteen groups submitted 70 predictions for this target. Five were of high accuracy, seven of medium accuracy, eight were qualified as acceptable, and three were dis-

Fig. 1. Pictorial overview of the prediction results for target T01, the Hpr-K/P-Hpr complex. **a:** A top view of the target complex showing the Hpr-K/P trimer (receptor) in ribbon representation with each subunit represented in a different colour. The three bound Hpr ligand molecules are represented by their molecular surface envelope, and the red spheres on each molecule depict the geometric centre of the ligands. **b:** Positions of the geometric center of the ligands relative to the kinase trimer in submitted predictions. To produce this picture, the receptor trimers in each predicted complex were superimposed onto their counterpart in the target and the corresponding rigid body transformation was applied to the entire complex (see text). The turquoise spheres correspond to acceptable solutions, and the gold spheres to incorrect solutions, following the ranking criteria in Table II. The red spheres represent the relative positions of the geometric centres of the ligands in the target structure. This figure and Figures 2–5 were generated by using the Insight software.[47]

Fig. 2. Pictorial overview of the prediction results for target T02, the rotavirus VP6-Fab complex. **a:** The rotavirus receptor structure is in ribbon representation with each subunit in the trimer represented in a different color (bottom), and one bound Fab molecule represented by its molecular surface envelope (top). The red sphere indicates the geometric center of the variable domain of the antibody used in analyzing the correspondence between the predicted and observed ligand positions. **b:** Positions of the geometric centers of the ligands relative to the VP6 trimer in predicted complexes ranked as being of medium accuracy (blue spheres), acceptable (turquoise spheres) and incorrect (gold spheres). The relative position of the geometric center of the ligand in the target is shown as a red sphere. This picture was produced as described in the legend of Figure 1(b).

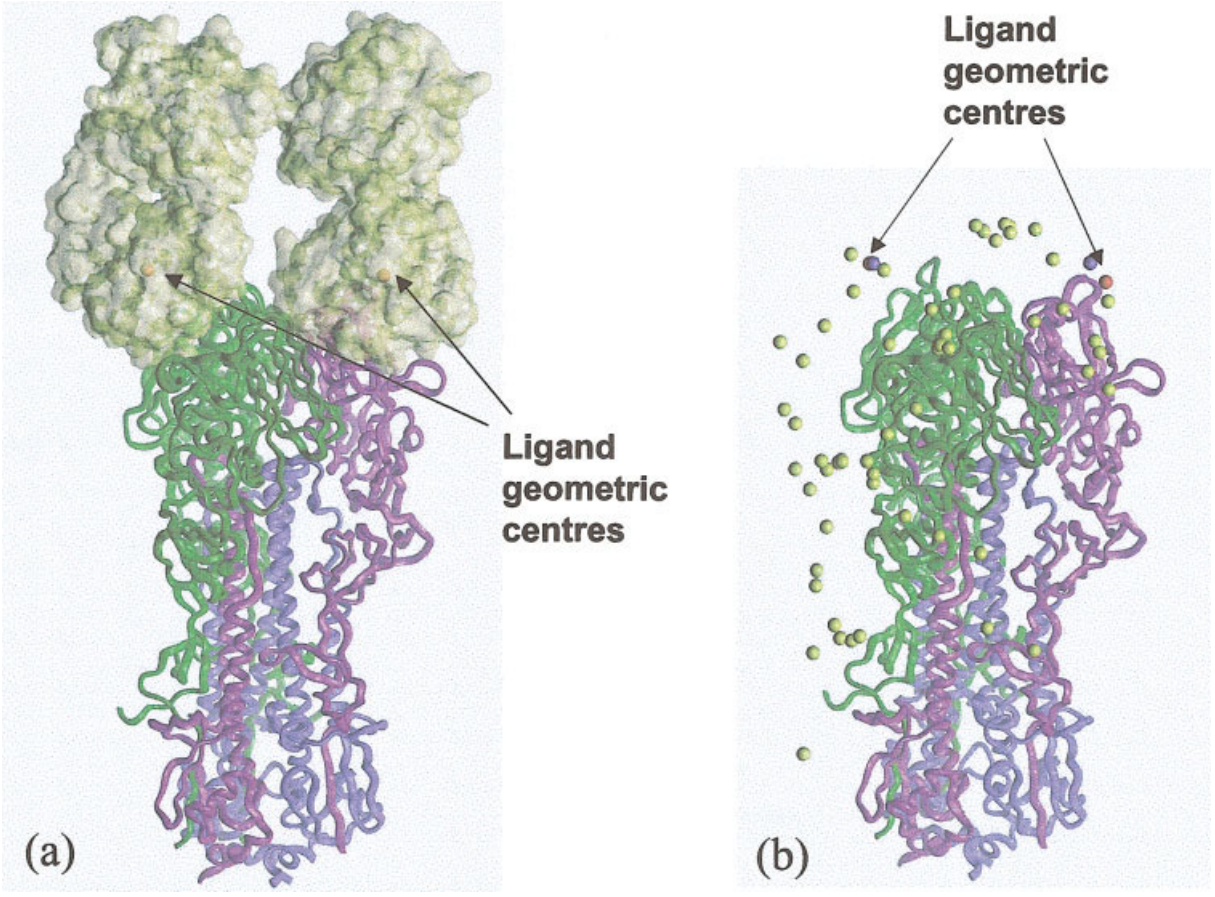

Figure 3.

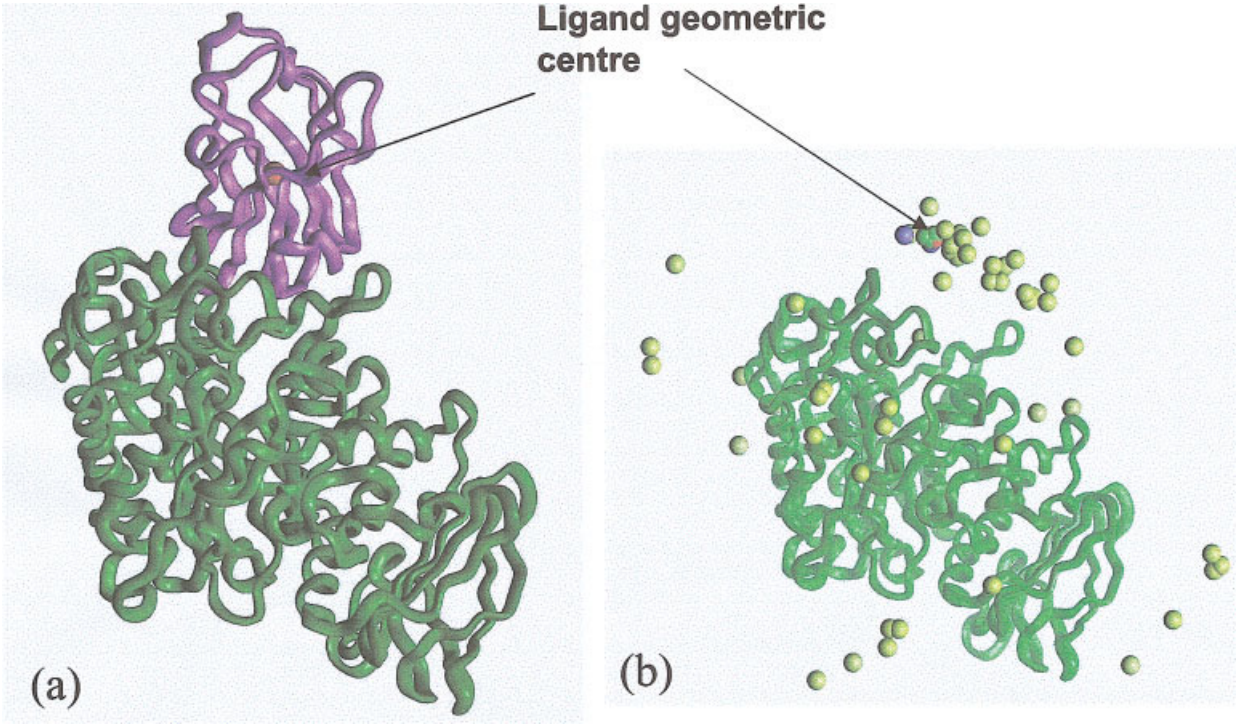

Figure 4.

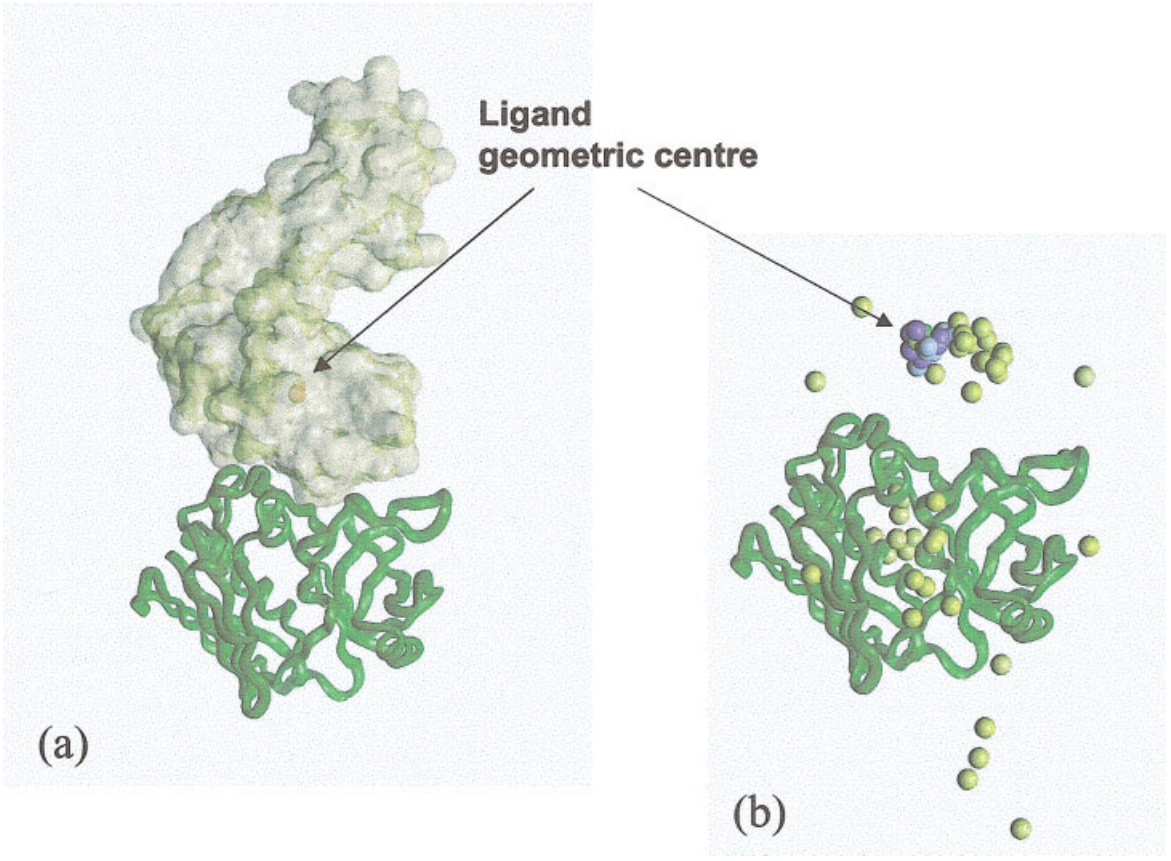

Fig. 5. Pictorial overview of the prediction results for target T07, the TCRβ-SpeA complex. **a:** The target complex, depicting ribbon drawings of the receptor (SpaA molecules) in green, and the TCR β chain ligand in purple. The geometric center of the ligand portion used for the analysis (TCRβ variable domain) is depicted as a red sphere. **b:** Positions of the geometric center of the ligands relative to the superantigen in predictions ranked as highly accurate (green spheres), medium accuracy (blue spheres), acceptable (turquoise spheres), and incorrect (gold spheres). For all other details, see legend of Figure 1(b).

carded for having too many clashes [Table III(f)]. It is of interest for this target more than any other, several of the groups submitted more than one correct prediction, and for two groups Tau (*Wolfson*) and Boston U. (*Weng*), all the five submitted predictions were of acceptable quality or better.

Fig. 3. Pictorial overview of the prediction results for target T03, the influenza Hemagglutinin-Fab complex. **a:** The influenza hemagglutinin structure is in ribbon representation, with each subunit in the trimer represented in a different color (bottom) and the bound Fab molecules represented by their molecular surface envelopes (top). The red spheres indicate the geometric center of the variable domains of the antibody used in analyzing the correspondence between the predicted and observed ligand positions. **b:** Positions of the geometric center of the ligands relative to the hemagglutinin trimer in predictions ranked as being of medium accuracy (blue spheres) and incorrect (gold spheres). For other details see legend of Figure 1(b).

Fig. 4. Pictorial overview of the prediction results for target T06, the porcine α-amylase-CAbAM D9 complex. **a:** The target complex, depicted as ribbon drawings of the porcine α-amylase receptor in green, and the CabAM-D9 VHH ligand in purple. The geometric center of the ligand is depicted as a red sphere. **b:** Positions of the geometric center of the ligands relative to the enzyme in predictions ranked as being highly accurate (green spheres), medium accuracy (blue spheres), acceptable (turquoise spheres), and incorrect (gold spheres). For all other details see legend of Figure 1(b).

Table III(f) (bottom) lists the details for the best prediction ranked acceptable or higher from each group. The five highly accurate predictions identify between 60 and 80% of the native contacts and between 0 and 20% of false-positive contacts. Their L_RMSs are between 1.3 and 2.6 Å and I_RMSs are all 1 Å or less. The two medium accuracy predictions by the Boton U. (*Weng*) and Scripps (*Abagyan*) groups have native contacts fraction of 80% and 60%, respectively, but a larger proportion of false-positive contacts (40–50%) and higher L_RMS values (3.9–6.0 Å). The Scripps prediction is a borderline case, which is probably better ranked as acceptable, because it resembles more the last two acceptable predictions listed in Table III(f) than the better models.

The prediction performance for this target (20 models of acceptable or higher quality out of a total of 70 models) is thus relatively high and clearly higher than for targets T01–T06. Inspection of the prediction methods summary in Table I reveals that all the groups that produced at least one high or medium accuracy prediction for target T07 used the information on the known 3D structure of the homologue complex to guide, or filter, their solutions. The groups that did not use this information produced lower

TABLE IV. Summary of Docking Predictions

Predictor group	T1	T02	T03	T04	T05	T06	T07	Predictor summary
Scripps US (*Abagyan*)	0	0	**	0	0	***	**	3/2**/1***
Boston U. US (*Camacho*)	*	0	0	0	0	***	***	3/2***
Weizmann Inst. IL (*Eisenstein*)	*	*	0	0	0	0	***	3/1***
Imperial Coll. UK (*Sternberg*)	0	*	0	0	0	***	*	3/1***
UCSD, US (*Ten-Eyck*)	*	*	0	0	0	**	0	3/1**
Sheffield U. UK (*Gardiner*)	*	*	—	—	—	—	—	2
U. Washington US (*Gray/Baker*)	0	0	0	0	0	**	***	2/1**/1***
U. Lisbon,Pt (*Palma/Krippahl*)	—	0	—	0	0	**	*	2/1**
Aberdeen U. UK. UK (*Ritchie*)	0	0	**	0	0	***	0	2/1**/1***
Boston U. US (*Weng*)	0	**	0	0	0	0	**	2/2**
TAU IL/NIH US (*Wolfson/Nussinov*)	*	0	0	0	0	0	***	2/1***
Cancer Research UK (*Fitzjohn/Bates*)	—	—	—	0	0	0	***	1/1***
Scripps US (*Olson*)	*	0	0	—	—	—	—	1
U. Autonoma Madrid SP (*Valencia*)	*	—	—	—	—	—	—	1
SUNY/MUSC US (*Vakser*)	0	*	0	—	—	—	—	1
Target summary	7	6/1**	2/2**	0	0	7/3**/4***	9/2**/5***	

This table summarizes the results obtained by all the groups that submitted one or more predictions of acceptable quality or better for at least one target. Column 1 lists the group's affiliation and the last name of the principle investigator. The next seven columns list the results obtained for each of the seven targets. The right-most column summarizes the results per predictor group, and the bottom row summarizes the results per target. 0, none of the submitted predictions was of acceptable quality;—,no predictions were submitted; *, at least one of the submitted predictions was in the acceptable range; **, at least one of the submitted predictions was of medium accuracy; and ***, at least one prediction was of high accuracy. See Table II for the definition of the parameters range used to rank the predictions. The summary entries list the total number of acceptable predictions, followed by the number of predictions of medium and high accuracy denoted by a ** and ***, respectively.

quality or incorrect predictions. The pictorial summary of the predictions for this target is given in Figure 5.

Overview of the Results Across Predictor Groups and Targets

Table IV summarizes the results obtained for all seven targets by all the groups that submitted at least one prediction ranking acceptable or better. We see that 14 of the 19 groups taking part in this first CAPRI experiment have an entry in this table. Inspection of Table IV reveals moreover that five of the seven targets were correctly predicted, albeit, with varying accuracy, by at least one group.

The best any group has done is to provide at least one acceptable prediction for three of the seven targets available for this experiment, although obviously several of the groups listed in Table IV did not submit predictions for every target. The first five groups listed in Table IV achieve this performance. Three of these provide at least one acceptable, or better, prediction for targets T06 and T07. One provides a medium accuracy prediction for target T03, whereas four have one acceptable prediction for either target T01 or T02, or both.

A further six groups submitted acceptable predictions for only two targets, and for half of these, the best prediction is of medium or high accuracy. The Sheffield (*Gardiner*) group submitted predictions for only two targets (T01 and T02) and those contained acceptable solutions for both.

The last four entries in Table IV are for groups that submitted predictions for only a small subset of the targets. Thus, it is not too surprising that they produced only one acceptable or better prediction. The Madrid (*Valencia*) group predicted only target T01 because it seemed to be the only case amenable to their neural network approach. The Scripps (*Olson*) and SUNY (*Vakser*) groups did not submit predictions to the second round of CAPRI (targets T04–T07), whereas the Cancer Research (*Bates*) group joined only the second CAPRI round

and hence provided no predictions of the first three targets.

It can also be seen that no highly accurate predictions were obtained for the first three targets, which represent the more difficult prediction problems either due to the large size of the molecules, or to conformational differences between the bound and unbound components, or both. On the other hand, highly accurate predictions were achieved by several groups for targets T06 and T07.

The excellent results obtained for Target T07 were clearly the consequence of using structural information on a related complex. But those for target T06 made use only of biochemical information on the regions likely to interact (VHH CDRs with the enzyme active site) and are therefore more directly attributable to the docking procedures themselves. The use of this type of information, as well as data on sequence conservation, were most probably key to having as many as seven groups producing acceptable predictions for Target T01 despite the important conformational differences between the bound and unbound kinase molecules.

Finally, a comment is warranted on the two groups Columbia (*Norel*) and Kitasato U. (*Umeyama*) who submitted predictions for all seven targets but produced acceptable predictions for none. The main feature common to their approaches seemed to be that they made little or no use of biochemical information. Otherwise, their docking procedures are very different, with the Columbia team using a more classical FFT procedure, whereas the Kitasato team used a new hitherto untested algorithm (Table I). Nothing can be said about the two remaining groups with no acceptable predictions, Aberdeen (*Mustard*), and Northwestern (*Shoichet*), because both submitted predictions for a single target, T01 and T07, respectively.

Current Status of Docking Procedure and Their Potential Applications

This first CAPRI experiment with 465 predictions for seven targets, obtained by using a wide variety of methods, provides a much better overview of the performance of current docking procedures than hitherto available. The good quality predictions obtained here for a number of difficult targets clearly shows that progress has been achieved since docking procedures have last been assessed about 6 years ago.[28–30] But the performance of current procedures taken globally is by no means reliable enough to allow their use as a routine tool. For many real-world problems, as in the case of targets T01–T06, only around 10% of the submitted predictions are of acceptable quality or better. Only when detailed structural information is available on a related complex in which the interaction mode is conserved, as for Target 07, do we see a higher success rate of 30% correct predictions.

Prior knowledge and information on the regions likely to interact have been used whenever possible on all targets by most predictors (Table I). The results were positive in most cases, and predictors who did not consider them clearly lost out. A further indication that using such information was important can be obtained by comparing the descriptions of the docking procedures in Table I with the results summary in Table IV. We see indeed that the most successful procedures, those that produced at least three predictions of acceptable or higher quality, are based on very different search algorithms (FFT, rigid body mechanics molecular mechanics) and scoring functions (geometric only, different empirical, and database-derived force fields). Because several other procedures, which produced less successful results, used similar algorithms or scoring function in combination with biochemical knowledge, it is tempting to conclude that a successful achievement depends on the manner in which this knowledge is integrated into the calculation procedure. Thus, more work in this area is clearly a direction for future research.

A legitimate question to address at this stage is under which circumstances and for what purpose can current docking procedures be useful?

Having found that accurate predictions can be obtained for only a subset of problems (molecules are small, limited conformational changes, biochemical information available, etc.), we can conclude that, at least for such problems, current docking procedures are capable of producing models accurate enough for guiding drug design or for rational mutagenesis studies. Given that usually these models accurately predict the relative ligand-receptor position and orientation as well, they should also be useful in structural studies for deriving phase information by molecular replacement[42] or for the interpretation of low-resolution electron density maps obtained by electron microscopy.[40]

Medium-accuracy predictions seem to be reachable for a wider range of problems, including those with quite large receptor molecules. But here too, biochemical information and prior knowledge are needed to guide the calculations. However, the resulting models often contain between 30 and 60% of non-native contacts, and although the interface RMSDs are within 2 Å, detailed side-chain modeling might not be straightforward. Therefore, such models are probably not accurate enough for guiding rational drug design or structural studies but should be useful for exploratory mutagenesis and functional analyses.

To enable such applications, however, the correct model should be among the top ranking solutions produced by the docking calculations, a condition largely fulfilled by the acceptable solutions analyzed here, because those were among the five top predictions of each group examined in this evaluation. Moreover, of the 24 high- and medium-accuracy predictions produced for the seven targets, a sizeable fraction (60%) corresponds to the top ranking solutions (models 1 or 2). However, one should recall that many predictors reranked their predictions manually (Table I).

Finally, many of the lower accuracy predictions, somewhat generously qualified as acceptable in this evaluation, might also be very helpful. We find that many such predictions, as well as some of the incorrect predictions, often identify at least 50% of the native residues that participate in the interaction on the receptor, the ligand or both [f_{IR_L} and f_{IR_R} in Table III(a–f)]. Hence, many of the

predictions may be helpful in delineating regions on the protein surface that are likely to interact. On the other hand, it is not clear from the present analysis to what extent current docking procedures can be used for predicting if two proteins that are not known to interact beforehand are likely to do so or not. This requires a systematic comparison of scores of good solutions with those of random ones, something that was not performed here, and undertaken by a few groups only (Eisenstein et al. this issue).

Prospects for Improving Docking Methods

The very wide panorama of methods used in this first CAPRI experiment offers unprecedented opportunity for some useful insights into the directions for future progress. Not too surprising is the recognition that the ability of docking methods to single out the correct modes of protein–protein interactions from a large number of incorrect ones is still limited. However, appreciable efforts in this direction are being made already, as judged by the increased sophistication of the scoring functions used in several of the more recent procedures described in Table I. Further progress should be expected as our capacity to judiciously approximate the physical contributions to the free energy of these interactions improves.

Another key component of docking procedures is the algorithms for sampling rigid body and conformational degrees of freedom. Some of these, in particular the spherical polar Fourier correlation or geometric hashing algorithms, are quite efficient. But often the procedures that implement them are not the same as those with the most advanced scoring functions. Therefore, notable progress could result in the near future from combining the best of both worlds.

Handling conformational flexibility remains nonetheless a major challenge. However, several completely novel procedures evaluated in this first CAPRI experiment were capable of handling some conformational flexibility and simulating more physical docking trajectories. Some of these seemed quite promising because they performed no worse than the more classical approaches despite being still in an early stage of development.

Thus, there appears to be not only a lot of room for improvements in docking algorithms but also concrete perspectives in achieving them. Large-scale blind predictions and their objective assessment, as done here, are clearly key in monitoring progress in this field. Therefore it is of paramount importance that other CAPRI experiments should follow this one to create a similar momentum as that produced by CASP for protein structure prediction methods.

ACKNOWLEDGMENTS

We acknowledge the Marie-Curie Host Training grant (contract num, QLK3-1999-51297), to Raúl Méndez for support. Thanks go to Jean Richelle for help with computer systems and to the rest of the SCMBB (Service de Conformation de Macromolecules Biologiques, et Bioinformatique) group members for their kind support and friendship. We also thank all the participants and the management team of this first CAPRI experiment for helpful discussions and valuable input and cooperation. Last but not least, we express gratitude to all the crystallographers that provided their X-ray structures before publication.

REFERENCES

1. Kleanthous C. Protein-protein recognition, Oxford University Press; 2000.
2. Janin J, Wodak SJ. The structural basis of macromolecular recognition. Adv Protein Chem 2002;61:9–73.
3. Jones S, Thornton JM. Principles of protein-protein interactions. Proc Natl Acad Sci USA 1996;93:13–20.
4. Marcotte EM, Pellegrini M, Thompson MJ, Yeates TO, Eisenberg D. A combined algorithm for genome-wide prediction of protein function. Nature 1999;402:83–86.
5. Xenarios I, Rice DW, Salwinski L, Baron MK, Marcotte EM, Eisenberg D. DIP: the database of interacting proteins. Nucleic Acids Res 2000;28:289–291.
6. Enright AJ, Iliopoulos I, Kyrpides NC, Ouzounis CA. Protein interaction maps for complete genomes based on gene fusion events. Nature 1999;402:86–90.
7. Bader GD, Donaldson I, Wolting C, Ouellette BF, Pawson, T, Hogue CW. BIND—the biomolecular interaction network database. Nucleic Acids Res 2001;29:242–245.
8. Park J, Lappe M, Teichmann SA. Mapping protein family interactions: intramolecular and intermolecular protein family interaction repertoires in the PDB and yeast. J Mol Biol 2001;307:929–938.
9. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 2000;403:623–627.
10. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci USA 2001;98:4569–4574.
11. Hazbun TR, Fields S. Networking proteins in yeast. Proc Natl Acad Sci USA 2001;98:4277–4278.
12. Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, Lenzen G, Petel F, Wojcik J, Schachter V. The protein-protein interaction map of Helicobacter pylori. Nature 2001;409:211–215.
13. Wodak SJ, Janin J. Computer analysis of protein-protein interaction. J Mol Biol 1978;124:323–342.
14. Janin J, Wodak SJ. Reaction pathway for the quaternary structure change in hemoglobin. Biopolymers 1985;24:509–526.
15. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 2002;47:409–43.
16. Jiang F, Kim SH. "Soft docking": matching of molecular surface cubes. J Mol Biol 1991;219:79–102.
17. Cherfils J, Duquerroy S, Janin J. Protein-protein recognition analyzed by docking simulation. Proteins 1991;11:271–280.
18. Shoichet BK, Kuntz ID. Protein docking and complementarity. J Mol Biol 1991;221:327–346.
19. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. Proc Natl Acad Sci USA 1992;89:2195–2199.
20. Norel R, Fischer D, Wolfson HJ, Nussinov R. Molecular surface recognition by a computer vision-based technique. Protein Eng 1994;7:39–46.
21. Totrov M, Abagyan R. Detailed ab initio prediction of lysozyme-antibody complex with 1.6 A accuracy. Nat Struct Biol 1994;1:259–263.
22. Vakser IA. Protein docking for low-resolution structures. Protein Eng 1995;8:371–377.
23. Meyer M, Wilson P, Schomburg D. Hydrogen bonding and molecular surface shape complementarity as a basis for protein docking. J Mol Biol 1996;264:199–210.
24. Gabb HA, Jackson RM, Stemberg MJ. Modelling protein docking using shape complementarity, electrostatics and biochemical information. J Mol Biol 1997;272:106–120.
25. Ackermann F, Hermann G, Posch S, Sagerer G. Estimation and filtering of potential protein-protein docking positions. Bioinformatics 1998;14:196–205.

26. Camacho CJ, Weng Z, Vajda S, DeLisi C. Free energy landscapes of encounter complexes in protein-protein association. Biophys J 1999;76:1166–1178.
27. Ritchie DW, Kemp GJ. Protein docking using spherical polar Fourier correlations. Proteins 2000;39:178–194.
28. Strynadka NC, Eisenstein M, Katchalski-Kazir E, Shoichet BK, Kuntz ID, Abagyan R, Totrov M, Janin J. Zimmerman F. Molecular docking programs successfully predict the binding of a beta-lactamase inhibitory protein to TEM-1 beta-lactamase. Nat Struct Biol 1996;3:233–239.
29. Dixon JS. Evaluation of the CASP2 docking session. Proteins 1997;Suppl 1:198–204.
30. Vakser IA. Evaluation of GRAMM low ressolution docking methodology on the hemaglutinin-antibody complex. Proteins 1997; Suppl 1:226–230.
31. Moult J. Predicting protein three-dimensional structure. Curr Opin Biotechnol 1999;10:583–588.
32. Fieulaine S, Morera S, Poncet S, Mijakovic I, Galinier A, Janin J, Deutscher J, Nessler S. X-ray structure of a bifunctional protein kinase in complex with its protein substrate HPr. Proc Natl Acad Sci USA 2002;99:13437–13441.
33. Barbey-Martin C, Gigant B, Bizebard T, Calder LJ, Wharton SA, Skehel JJ, Knossow M. An antibody that prevents the hemagglutinin low pH fusogenic transition. Virology 2002 Mar 1;294:70–74.
34. Desmyter A, Spinelli S, Payan F, Lauwereys M, Wyns L, Muyldermans S, Cambillau C. Three camelid VHH domains in complex with porcine pancreatic alpha-amylase. Inhibition and versatility of binding topology. J Biol Chem 2002;277:23645–23650.
35. Sundberg EJ, Li H, Llera AS, McCormick JK, Tormo J, Schlievert PM, Karjalainen K, Mariuzza RA. Structures of two streptococcal superantigens bound to TCR beta chains reveal/diversity in the architecture of T cell signaling. Structure (Camb) 2002;10:687–699.
36. Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. J Mol Biol 1999;285:2177–2198.
37. Thouvenin E, Schoehn G, Rey FA, Petitpas I, Mathieu M, Vaney MC, Cohen J, Kohli E, Pothier P, Hewat E. Antibody Inhibition of the transcriptase activity of the rotavirus DLP: a structural view. J Mol Biol 2001;307:161–172.
38. Wang H. Grid-search molecular accessible surface algorithm for solving the protein docking problem. J Comp Chem 1991;12:746–750.
39. Delhaise P, Bardiaux M, De Maeyer M, Prevost M, Van Belle D, Donneux J, Lasters I, Van Cutsem E, Alard P, Wodak SJ. The BRUGEL package: toward computer-aided design of macromolecules. J Mol Graphics 1988;6:219.
40. Wriggers W, Milligan RA, McGammon JA. Situs: a package for docking crystal structures into low-resolution maps from electron microscopy. J Struct Biol 1999;125:185–195.
41. McLachlan AD. Gene duplications in the sructural evolution of chymotrypsin. J Mol Biol 1979;128:49–79.
42. Navaza J. Implantation of molecular replacement in AMoRe. Acta Crystallogr D. Biol Crystallogr 2001;57:1367–1372.
43. de Groot BL, van Aalten DMF, Scheek RM, Amadei A, Vriend G, Berendsem HJC. Prediction of protein conformational freedom from distance constraints. Proteins 1997;29:241–251.
44. Thouvenin E, Schoehn G, Rey F, Petitpas I, Mathieu M, Vaney MC, Cohen J, Kohli E, Pothier P, Hewat E. Antibody inhibition of the transcriptase activity of the rotavirus DLP: a structural view. J Mol Biol 2001;307:161–172.
45. Gigant B, Barbey-Martin C, Bizebard T, Fleury D, Daniels R, Skehel JJ, Knossow. A neutralizing antibody Fab-influenza haemagglutinin complex with an unprecedented 2:1 stoichiometry: characterization and crystallization. Acta Crystallogr D 2000;56: 1067–1069.
46. Duncan B, Olson A. Applications of evolutionary programming for the prediction of protein-protein interactions. Evolutionary programming V. Proceedings of the 5th annual conference on evolutionary programming. Boston, MA: MIT Press; 1996.
47. Dayringer, H.E., Tramontano A., Sprang, S.R., Fletterick, R.J., Interactive program for visualization and modeling of proteins, nucleic-acids and small molecules. J Mol Graph 1986;4:82.