PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Strengths and weaknesses of data-driven docking in critical assessment of prediction of interactions

Sjoerd J. de Vries,[1] Adrien S. J. Melquiond,[1] Panagiotis L. Kastritis,[1] Ezgi Karaca,[1] Annalisa Bordogna,[2] Marc van Dijk,[1] João P. G. L. M. Rodrigues,[1] and Alexandre M. J. J. Bonvin[1]*

[1] NMR Research Group, Bijvoet Center for Biomolecular Research, Utrecht University, 3584 CH Utrecht, The Netherlands
[2] Dipartimento di Scienze dell'Ambiente e del Territorio, Universita degli Studi di Milano-Bicocca, Piazza della Scienza 1, 20126 Milano, Italy

ABSTRACT

The recent CAPRI rounds have introduced new docking challenges in the form of protein-RNA complexes, multiple alternative interfaces, and an unprecedented number of targets for which homology modeling was required. We present here the performance of HADDOCK and its web server in the CAPRI experiment and discuss the strengths and weaknesses of data-driven docking. HADDOCK was successful for 6 out of 9 complexes (6 out of 11 targets) and accurately predicted the individual interfaces for two more complexes. The HADDOCK server, which is the first allowing the simultaneous docking of generic multi-body complexes, was successful in 4 out of 7 complexes for which it participated. In the scoring experiment, we predicted the highest number of targets of any group. The main weakness of data-driven docking revealed from these last CAPRI results is its vulnerability for incorrect experimental data related to the interface or the stoichiometry of the complex. At the same time, the use of experimental and/or predicted information is also the strength of our approach as evidenced for those targets for which accurate experimental information was available (e.g., the 10 three-stars predictions for T40!). Even when the models show a wrong orientation, the individual interfaces are generally well predicted with an average coverage of 60% ± 26% over all targets. This makes data-driven docking particularly valuable in a biological context to guide experimental studies like, for example, targeted mutagenesis.

Proteins 2010; 78:3242–3249.

Key words: Haddock; multi-body docking; web server; scoring; experimental restraints; interface prediction.

INTRODUCTION

Over the last years, molecular docking methods have gained in importance to provide atomic insight into the structure of biomolecular complexes, especially when experimental structural information is supporting the modeling. They can basically be divided into three different classes[1,2]: global methods based on geometric hashing or fast Fourier transforms (FFTs),[1] Monte Carlo techniques,[3] and data-driven docking such as our High Ambiguity Driven biomolecular DOCKing approach, HADDOCK.[4,5]

The performance of data-driven docking relies most on the amount and quality of the available experimental information. HADDOCK can deal with a versatile range of data ranging from molecular (stoichiometry, symmetry, orientation), through residue-level (identification of residues at the interface) to atomic resolution information (binding mechanism, specific contacts, distance restraints from NMR).[6,7] The use of restraints provides a valuable benefit for all docking cases but above all for the difficult ones, such as docking performed from homology models and/or structures involving (large) conformational changes. By using a variety of restraints, HADDOCK decreases the complexity of the interaction space search, limiting it ideally to the interfacial regions, which allows a thorough refinement of the models. On the other hand, misleading information can introduce a strong bias that HADDOCK might not be able to overcome. Although nearly all docking methods can optionally use information to restrict the docking solutions, restraints play a central role in data-driven docking. Therefore, using

Additional Supporting Information may be found in the online version of this article.
Abbreviations: AIRs, ambiguous interaction restraints; CAPRI, critical assessment of prediction of interactions; HADDOCK, high ambiguity driven biomolecular DOCKing; PDB, protein data bank; RMSD, root mean square deviation.

The authors state no conflict of interest.
Grant sponsor: Netherlands Organization for Scientific Research; Grant number: VICI 700.56.442
*Correspondence to: Alexandre M. J. J. Bonvin, NMR Research Group, Bijvoet, Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.
E-mail: a.m.j.j.bonvin@uu.nl
Received 1 April 2010; Revised 23 June 2010; Accepted 26 June 2010
Published online 6 July 2010 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.22814

reliable information is of the highest importance in our approach since high quality information can lead to high quality models, whereas wrong information can prevent the generation of acceptable solutions.

The last CAPRI rounds have introduced new challenges with a special category for servers, the inclusion of protein-RNA complexes, symmetrical targets, multi-component targets, and finally the discrimination of a designed interface from decoys (not included in this evaluation). Over the last years, HADDOCK has shown a rather strong and constant performance in CAPRI, belonging to the most performing methods, with a noteworthy success rate when restricted to difficult cases.[5,8] We present here our results in CAPRI rounds 13–19, both from manual submissions and from the HADDOCK server.[9] We discuss in particular the strengths and weaknesses of data-driven docking, highlighting the uniqueness of our approach and providing a guideline for structural biologists to avoid possible pitfalls in molecular docking.

MATERIALS AND METHODS

Docking approach

All docking runs were performed with a local version of HADDOCK2.1[5] or its web server implementation (http://haddock.chem.uu.nl/services/HADDOCK/haddock.php).[9] HADDOCK supports the docking of different types of molecules: proteins, peptides, nucleic acids, and even small ligands. The multi-body docking interface of HADDOCK[10] allows the user to supply up to six molecules. Several types of restraints can be used within HADDOCK, the most commonly used ones being Ambiguous Interaction Restraints (AIRs). Via AIRs, the interface information can be incorporated into the docking procedure. AIRs are composed of active and passive residues; active residues are defined as the interface residues that should be at the interface and passive residues correspond to their solvent accessible neighbours. Typically, 50% of the restraints are discarded at random for each docking trial. The standard docking procedure of HADDOCK is composed of three consecutive steps: rigid body energy minimization (it0), semi-flexible refinement in torsion angle space (it1), and a final explicit solvent refinement (water). After each step, the structures are ranked according to the HADDOCK score.[5] Residues identified as being part of the interface within a given model are automatically treated as flexible during the it1 and water steps. The final solutions are clustered using a 7.5Å cutoff based on their pairwise interface ligand-RMSD values (see de Vries *et al.*[5]; the interface ligand-RMSD is similar to the ligand-RMSD, but only those backbone atoms of the receptor and ligand are taken into account that are in the interface of at least one docking model) and ranked based on the average score of the four best-score cluster members. Compared to default settings, the number of models generated was usually increased to 10,000 (rigid-body) of which the top 400 were subjected to the semi-flexible refinement.

For the server submission, we selected the four models of the top-ranking cluster as given by the server and then distributed, according to cluster ranking, the remaining six models over all other clusters.

Homology modeling

For targets 33/34, 37, and 42, we used Modeller 9v3.[11] We built typically 50 standard models with the automodel script, followed by loop refinement using the loopmodel class.[12] The resulting models (typically 500) were ranked by their objective function and the 10 lowest energy models were used for docking predictions. For target 37, we took advantage of the multiple templates functionality of MODELLER to generate an ensemble of 16 structures (see text for details). Finally, we used the I-TASSER server[13] for target 38.

Information used to drive the docking

Detailed information about the restraints used for the various CAPRI targets is provided in the Supporting Information.

Scoring experiment

The scoring of the models uploaded by the CAPRI participants was performed using HADDOCK scripts. Missing atoms were added automatically and the side-chains were subjected to 50 steps of energy minimization. The resulting models were clustered and scored using the standard final HADDOCK score, excluding the AIR energy term:

$$\text{HADDOCK} - \text{score} = 1.0^{*}E_{\text{vdW}} + 0.2^{*}E_{\text{elec}} + 1.0^{*}E_{\text{desolv}}$$

where E_{elec} and E_{vdW} correspond to the Coulomb and Lennard-Jones energy terms calculated with a 7.5Å cutoff using the OPLS non-bonded parameters[14] and E_{desolv} is an empirical desolvation energy term.[15] Our submission consisted typically of several models of the top ranking cluster and additional models from the remaining clusters, filtered based on the available experimental or predicted information.

RESULTS

The results of our manual submission and of the HADDOCK server are summarized in Table I. Our manual submissions from HADDOCK were successful for 6 out of 9 complexes (T29, T34, T37, T40, T41, and T42), as T33 and T34 (respectively T38 and T39) correspond to the same complex. The HADDOCK server, which participated for 7 complexes, was successful for 4 of them

Table I
Performance of Our Group in the Recent CAPRI Rounds

	Target class	Exp. data[a]	Prediction performance	Prediction experiment	Uploaded structures	Scoring	HADDOCK SERVER
T29	ub/b	No		6/1**	89/46**, 2***	9/5**	Did not participate
T30	ub/ub	No		0	0	0	Did not participate
T32	ub/ub	No		0	0	0	0
T33	h/h	Yes		0	0	0	0
T34	h/b			6/3**	92/13**	7/1**	7/7*
T37	ub/h	No		2/2**	12/5**, 2***	2/1**	0
T38	ub/h	Yes[a]		0	0	0	0
T39	ub/b			0	0	0	0
T40	ub/b	Yes		5/5***	50/10**, 39***	6/4**	1/1***
				5/5***	40/40***	4/2**, 2***	0
T41	ub/ub	Yes		10/2**	17/2**	10/10*	4/1**
T42	h/h	Yes[a]		1/1*	Not assessed		1/1*[b]

Dark, medium, and light gray boxes correspond to high, medium, and acceptable models, respectively, and the white boxes indicate incorrect models, according to the CAPRI criteria. The target class describe the origin of the starting structures used for the docking prediction (ub: unbound; b: bound; h: homology model). Models are assessed with respect to the CAPRI star definition (*: acceptable, **: medium, ***: high quality prediction). Columns 4 and 5 correspond to the manual submission performance whereas columns 6–8 correspond to correct uploaded structures for the scoring experiment, performance in the scoring experiment, and performance of our server, respectively.

[a]The biochemical information for T38/T39 turned out to be incorrect for the crystal structure. The information for T42 was only partially correct.

[b]The HADDOCK multi-body server was able to predict one acceptable model for this target. Interestingly, this complex spans the second interface in the crystal structure whereas the acceptable model from the manual submission predicted the first interface.

(T34, T40, T41, and T42). In the scoring experiment, we successfully identified correct predictions for 6 targets (T29, T34, T37, T40, T41, and T42), which corresponds to the largest number of successfully scored targets of all scorer groups. Next to measuring success in term of acceptable, medium, and high quality predictions, we also analyzed how well our models predicted the individual interfaces (Table II). In the following, we discuss briefly the various complexes.

Target 29: (unbound–bound) was a yeast Trm8/Trm82 tRNA guanine-*N*(7)-methyltransferase complex[16] with Trm82 given in the bound form. Both Trm82 (bound) and Trm8 (free) structures were supplied by the organizers. As there was no other information available, we defined our active residues based on WHISCY interface predictions[17] (http://nmr.chem.uu.nl/Software/whiscy/index.html). For this target, we submitted six models from the highest-ranking cluster, a two-star and five one-star models. However, among the 100 structures we uploaded for the scoring experiment, two were three-star models, the only three-stars submitted by any group. These originated from a separate docking run, driven by contact restraints based on the highest-ranking cluster. For this run, 98.5% of the models were of two-star quality or better, 28.5% were of three-star quality, and the best-scoring structure was also a three-star. However, the structures from this run had a less favorable HADDOCK score and were not selected for the final submission. In the scoring experiment, we were again unable to select them, despite the fact that they belong to the top ranking cluster. One of them was however selected by the groups of Fernandez-Recio and Yaoqi Zou.

Target 30: (unbound–unbound) was given as "Rnd1-GTP bound to the RBD dimmer."[18] Because of this, we assumed that both RBD monomers were involved in the interaction. Both structures were supplied in their free forms by the organizers. As there was no other information available, we defined our active residues based on WHISCY interface predictions.[17] The Rnd1-GTP restraints were defined ambiguously to both monomers, driving both monomers to make contact. As the RBD dimer interface is very small, we anticipated conformational changes upon complex formation, and therefore performed a three-body docking (Rnd1 and both RBD subunits separately), defining additional restraints between the dimer interface residues of the two RBD subunits.

Table II
Average Fraction of Correct Interface Coverage f(IR) and Over-Predicted Interface f(OP) for the Submitted HADDOCK Solutions[a]

	p(IR)		p(OP)	
Target	Ligand	Receptor	Ligand	Receptor
T29	0.77 ± 0.11	0.82 ± 0.12	0.34 ± 0.05	0.38 ± 0.08
T30	0.73 ± 0.15	0.62 ± 0.14	0.52 ± 0.07	0.57 ± 0.10
T32	0.78 ± 0.05	0.62 ± 0.09	0.42 ± 0.06	0.49 ± 0.05
T33	0.42 ± 0.02	0.36 ± 0.02	0.36 ± 0.03	0.49 ± 0.03
T34	0.47 ± 0.13	0.75 ± 0.15	0.26 ± 0.20	0.09 ± 0.06
T37	0.26 ± 0.38	0.75 ± 0.09	0.79 ± 0.31	0.42 ± 0.07
T38	0.58 ± 0.11	0.00 ± 0.00	0.66 ± 0.05	1.00 ± 0.00
T39	0.47 ± 0.25	0.00 ± 0.00	0.71 ± 0.16	1.00 ± 0.00
T40	0.89 ± 0.07	0.90 ± 0.04	0.12 ± 0.08	0.07 ± 0.03
T41	0.82 ± 0.12	0.81 ± 0.10	0.11 ± 0.06	0.15 ± 0.08
T42	0.68 ± 0.11	0.68 ± 0.11	0.28 ± 0.05	0.28 ± 0.05

[a]The correct interface coverage f(IR) is calculated as the number of residue correctly in the interface in the model divided by the total number of interface residues in the reference complex. The over-predicted interface f(OP) is defined as the number of incorrectly predicted interface residues divided by the total number of predicted interface residues. Interface residues are defined based on loss of solvent accessibility. The values were taken from the CAPRI evaluation reports of the various models.

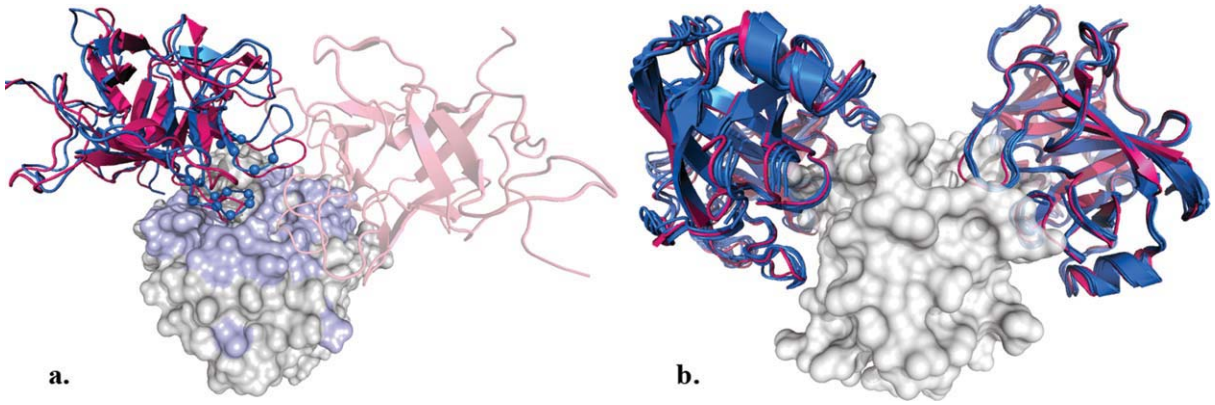

Figure 1

Overlay of our predictions (blue) onto the reference complex (pink) for two selected targets. In both cases, the receptor is shown in surface representation and the ligand in cartoon representation. (**a**) Best CAPRI redocking solution for target 32 (three-body docking with a second inhibitor molecule defined as fully passive). The area defined by the active residues is shown in light blue on the receptor and the active residues indicated by spheres on the Ca atoms for the ligand. The second inhibitor in the crystal structure is shown with a transparency effect. (**b**) Our 10 three-star CAPRI submissions for target 40.This figure was generated with PyMol.[51]

Unfortunately, all of our models showed the wrong orientation with respect to the crystal structure. This target was very difficult, which is most likely caused by a loop rearrangement (1809–1814) in RBD, and the lack of biochemical data. Note that only two groups predicted a single one-star model for this target. Still, HADDOCK was successful in obtaining very good estimates of the interface. The interface predictions that drove the docking covered 76% of the interface for both proteins, but the predicted surface area was more than twice as large as the true interface area. In contrast, the resulting HADDOCK models have interfaces similar in size to the crystal structure that cover 73–84% of the true interface (Table II).

Also, for the complex between savinase and the BASI inhibitor[19] in target 32 (unbound–unbound), our docking solutions correctly identified the interface: 87% for the inhibitor and 75% for the enzyme (Table II). This was due to the fact that the interface predictions for this complex were very good. For the first time, we used our recently developed consensus interface prediction algorithm, CPORT[20] that combines the results of six interface predictors (WHISCY,[17] PIER,[21] PROMATE,[22] cons-PPISP,[23] SPPIDER,[24] PINUP[25]). Predictions were made for savinase, all of which were centered on the active site. The same was done for the inhibitor, but we manually restricted the predictions to the inhibitory loop (residues 85–91). The complex was an unbound–unbound structure, with the structure of BASI coming from a complex with a different enzyme, Barley alpha-amylase[26] (PDB entry 1AVA_C).

However, the crystal structure reveals two inhibitor molecules in contact with the enzyme, each occupying a part of the active site. This crystal structure is in full agreement with our interface predictions, but our lack of knowledge on the correct stoichiometry prevented us from predicting the correct model since the entire interface was included in the AIRs definition. Therefore, our submitted docking models were unable to improve upon the interface predictions. In contrast, after the crystal structure was released, we performed a three-body docking, with exactly the same docking parameters, but with a second inhibitor molecule defined as fully passive. This resulted in a two-star docking solution [Fig. 1(a)], stressing the importance of proper information for data-driven docking.

Targets 33 and 34: (homology–homology/homology–bound) were a complex between a RNA methyl transferase and a 74 nucleotides RNA transcript containing three ribosomal RNA hairpins. Currently, this complex has not yet been published, and therefore, by necessity, our description of the docking procedure below lacks certain details and references to the literature.

The methyl transferase was built by homology from using MODELLER,[11] keeping the co-substrate and the zinc ions. For target 33, the RNA transcript was built manually by mutating the bases, using as a template a ribosomal *Escherichia coli* segment (71% sequence identity). In addition, one of the hairpins was modeled based on an NMR structure (100% sequence identity). The resulting models were subjected to a refinement in water using HADDOCK. The bound form of the rRNA was given for target 34. For both targets, the docking was driven by center of mass restraints, an unambiguous restraint between the co-substrate involved in the methyl group transfer and the atom N1 of a specific hairpin guanine of the RNA (found in the literature), ambiguous restraints derived from published NMR titration data for the RNA, and from a consensus interface prediction using RISP,[27] RNA-BindR,[28] and Pprint[29] for the protein. In addition, Watson-Crick and backbone dihedral restraints were introduced to maintain the RNA structure.

Unfortunately, while the biochemical data from the literature was accurate, inaccuracies in the homology model of the RNA prevented us from obtaining correct models for target 33. In contrast, several two-star solutions were generated for target 34 for which the bound form of the RNA was provided. The HADDOCK server, which participated for the first time to CAPRI in this round, generated seven one-star models out of 10 submitted models.

Targets 35 and 36: (homology–homology/homology–bound) consisted of two domains of xylanase Xyn10B from C. thermocellum, the polysaccharide binding module CBM22 and the catalytic module GH10, both of which had to be generated by homology modeling for T35. For T36, the bound form of CBM22 was provided. These domains are adjacent and covalently linked by an 8-residue linker (information given by the organizers). However, analysis of the crystal structure [PDB entry 2W5F (Najmudin *et al.*, unpublished data)] reveals a distance of more than 30 Å between the C-terminus of the first domain and the N-terminus of the other, which is simply incompatible with a covalent linkage (which we wrongly imposed as restraint in our docking!) Considering that the provided information was misleading, we will exclude this target from our analysis and overall performance assessment.

Target 37: (unbound–homology) was one of the most challenging targets. It was a complex between Arf6 and a 71-residue leucine zipper from JIP4.[30] For Arf6, the unbound form of the protein was known. There was no further experimental information on the interface.

We modeled the leucine zipper using MODELLER based on two templates: 2ZTA[31] (31 residues, 19% sequence identity) and 1GK6[32] (59 residues length, 20% sequence identity with the target), generating 16 models that were subsequently used as starting ensemble for docking. The periodicity of the leucine residues in the leucine zipper sequences (every $n + 7$ steps) allowed us to produce a multiple alignment by using multiple copies of the template sequence shifted by a running window of seven residues. Using this approach, we could model the 71 residues coiled coil homodimer from shorter, overlapping templates (eight times 2ZTA and four times 1GK6).

A particular challenge for this target was that the leucine zipper has helical repeats and we did not know which stretch would bind to the protein. We set up therefore a series of docking runs, with, for Arf6, active residues defined from CPORT predictions and, for the zipper, passive residues spanning various shifted 15-residue segments. We also found a homolog of Arf6 bound to a leucine zipper (PDB entry 2D7C[33]), and defined another series of docking runs based on the contacts observed in this complex, shifting again the alignment of the zippers for each docking run. All resulting runs were pooled and clustered and the top structures of the most favored clusters were submitted.

Strikingly, despite the challenging nature of this target, two of our submitted structures were of two-star quality. In fact, among the 100 structures we uploaded for the scoring experiment, two were of three-star quality! These were successfully picked up by the groups of Zhiping Weng and Yaoqi Zou. Our success in this target was largely due to the very accurate homology models that we were able to generate: the best model was only 0.95Å backbone RMSD from the crystal structure, and 10 out of 16 models were at less than 2Å from the crystal structure. We shared our models with the community and other groups could generate good quality predictions using them.

Some of our submitted solutions were derived from the docking based on the 2D7C structure, but it turned out that its binding mode was perpendicular to the Target 37 crystal structure. For the HADDOCK server, the time frame and server rules prevented the extensive docking approach described above, making this target beyond the abilities of our server.

For target 38/39 (unbound-homology/unbound-bound), we were misled by biochemical information[34] that Tyr232 was involved in the interaction that turned out to be inconsistent with the reference crystal structure, so that none of our submissions were correct. These targets were the unbound–homology and unbound–bound docking, respectively, of the same complex between centaurin and an FHA domain. For target 38, the FHA domain had to be modeled, we used I-TASSER[13] in this case, while for target 39 it was given in the bound form. The docking was performed using interface prediction from CPORT,[20] center of mass restraints, and forcing a contact between Tyr232 of the centaurin and the His45 of FHA, which turned out to be incorrect.

We note that also the CPORT interface predictions were unable to identify the correct interface for one of the proteins. This clearly shows the limitation of data-driven docking in the absence of reliable data.

For target 40 (unbound-bound), the trypsin-inhibitor complex, we correctly assumed from the literature that each inhibitor molecule would bind two trypsin molecules in a canonical binding mode.[35,36] This target was an enzyme-inhibitor complex between the unbound structure of the trypsin (PDB entry 1BTY[37]) and the bound form of API-A inhibitor.[38] For this target, five of our submissions were generated without any active or passive residues definition. Instead we imposed ambiguous distances between the OG and N of Ser195, the O of Ser214 and the N and O of Gly216 for the serine protease and the inhibitory loops 83–88 and 143–148 on the API-A side.[35,36] The five other submissions were driven using ambiguous interaction restraints for the trypsin (derived from the interfaces of known serine protease-inhibitor complexes) and the inhibitory loop 143–148.

Using this information in combination with CPORT interface predictions, we obtained highly specific results and submitted, for each binding mode, five models from the top ranking cluster, thus a unique solution for each binding mode. All 10 submitted models were of three-star quality, a first in CAPRI history [Fig. 1(b)]! The server

prediction, with restraints only from CPORT, also contained a three-star model for one of the two interfaces.

Target 41 (unbound–unbound) was a non-cognate complex between Colicin E9 DNase and immunity protein Im2 (PDB entry 2WPT).[39] We used the free form of colicin E9 and a NMR ensemble of Im2 [PDB entry 2NO8 (Boetzel *et al.*, unpublished data)]. The list of active and passive residues was derived after superimposition of Im2 and E9 onto the known colicin-Im protein complexes E9-Im9 (PDB entry 2K5X[40]) and E7-Im7 (PDB entry 7CEI[41]). This information was complemented by the definition of the binding pocket of Im2 derived from alanine scanning mutagenesis.[42] In addition, we imposed the stacking of the rings of Phe86 of E9 and Tyr54 of Im2 as described in the literature.[43]

The cognate binding mode for colicin-immunity protein complexes is well known, and we correctly assumed that it would be the same for T41. For both the manual prediction and the server, we submitted two-star and one-star models with near-complete interface coverage (Table II). However, many other groups were able to generate three-star models by direct homology modeling from the cognate E9-Im9 complex. Clearly, for cases like target 41, homology modeling is superior to unbound docking in obtaining high-accuracy predictions.

Target 42 (homology–homology) was an oligomeric assembly of a designed tetratricopeptide protein (PDB entry 2WQH),[44] originally specified by the organizers as a trimer. We developed especially for this target a multi-body version of the HADDOCK server, constructed within the 24 h time frame given by the organizers. However, the target was later re-specified as a dimer, with two possible interfaces. The model was generated with MODELLER using an idealized TPR motif (PDB entry 1NA0[45]) that differs only by one residue per TPR repeat (D39Y, D73Y, D107Y) from the target sequence. Nine of our manual submissions were obtained using non-crystallographic symmetry and C2 symmetry[10] with only three residues defined as active residues (Tyr39, Tyr73, Tyr107). The last submission was coming from a two-body docking of a super-helical trimer; this structure was generated with the multi-body server using the information available on TPR super helices.[45,46]

From literature information, we assumed that the target would be a superhelix, which turned out to be false. Nevertheless, our interface definitions were good and we were able to generate one-star predictions for both the manual and the server submission, with highly accurate interface coverage (Table II).

DISCUSSION

Our results in CAPRI rounds 13–19 show that for data-driven docking with HADDOCK, the use of experimental information is both a strength and a weakness. Indeed, T40, and to a lesser extent T34 and T41, illustrate the strength of HADDOCK when accurate experimental restraints are available. Likewise, T32 and T38/T39 indicate that this can also be a weakness in the case of misleading information, for example, about the stoichiometry (T32) or the interface (T38/T39).

In the absence of restraints, we have relied on bioinformatic interface predictions. These were an essential factor to our success for targets 29 and 37: HADDOCK proved remarkably successful in dealing with such fuzzy data and selecting the correct residues out of a large over-predicted interface. Strikingly, this is also the case for two of our "failed" cases (T30 and T32), where HADDOCK models, although in the wrong orientation, showed an accurate characterization of the individual interfaces. In particular, for the case of T30, the interface coverage by HADDOCK reveals a considerable improvement over the underlying interface predictions that drove the docking. This means that, even if the HADDOCK models do not have the correct orientation, they show a high fraction of correct interface residues and can thus still be used as a reliable guide for mutagenesis experiments. It is worth pointing out that the interface coverage of the HADDOCK models has a great degree of consistency: indeed, similar interface coverage is shown by all of our 10 submitted models, regardless of their quality (stars) (except, obviously, for the case of target 37, with its many possible zipper segments). In contrast, solutions obtained by *ab initio* methods often show much more spread over the surface and, as such, are less informative in the context of predicting the interface and directing mutagenesis experiments.

When it comes to scoring, HADDOCK is well capable of discriminating correct from incorrect models, as shown by our performance in both the prediction experiment (we belong to the group with the largest number of correctly predicted targets) as in the scoring experiment (we achieved the highest number of correctly predicted complexes). It is worth noting that, in our high resolution scoring, models from *ab initio* FFT-based methods often have very poor scores due to intermolecular clashes that a simple energy minimization cannot alleviate. Such models might cluster with correct solutions, but are never selected. When it comes to finding highly accurate (three-star) complexes, there is still room for improvement since we were unable to pick up in some instances our own three-star predictions. Here, others have performed better than we have. This is apparent in both the prediction experiment (the Vajda and Zacharias groups) and the scoring experiment (the Yaoqi Zou group), and particularly for targets 29 and 37. Part of our problem can be due to the usually large number of acceptable solutions obtained and the flexible refinement stage leading to an optimization of any interface (even wrong ones), challenging our current scoring scheme.

CONCLUSION

In conclusion, HADDOCK is a highly successful docking method in cases where experimental information is available or bioinformatics predictions can be obtained. Even for "failed" cases showing a wrong orientation of the components, HADDOCK models are usually highly reliable in identifying interface residues, which is a valuable property to guide experimental work. The main weakness of our approach, as illustrated here, is its vulnerability for wrong experimental data related to the interface or the stoichiometry of the complex. Further, compared to other docking methods, we still have difficulty in discriminating highly accurate (three-star) from accurate (two-star) models. Finally, as all docking methods currently in use, complexes undergoing large conformational changes remain a challenge. Our ability to perform, differently than other existing approaches,[47–50] simultaneous multi-body docking that allows the user to model arbitrary large macromolecular assemblies,[10] however, provides a possible solution to dealing with domain motions by cutting a protein into interconnected sub-domains (unpublished results). For such systems, however, the quality of the available information will be even more crucial in order to guide the conformational changes towards the bound structure.

REFERENCES

1. Ritchie DW. Recent progress and future directions in protein-protein docking. Curr Protein Pept Sci 2008;9:1–15.
2. Vajda S, Kozakov D. Convergence and combination of methods in protein-protein docking. Curr Opin Struct Biol 2009;19:164–170.
3. Hart TN, Read RJ. A multiple-start Monte Carlo docking method. Proteins 1992;13:206–222.
4. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc 2003;125:1731–1737.
5. de Vries SJ, van Dijk AD, Krzeminski M, van Dijk M, Thureau A, Hsu V, Wassenaar T, Bonvin AMJJ. HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. Proteins 2007;69:726–733.
6. van Dijk AD, Boelens R, Bonvin AM. Data-driven docking for the study of biomolecular complexes. FEBS J 2005;272:293–312.
7. Melquiond ASJ, Bonvin AMJJ. Data-driven docking: using external information to spark the biomolecular rendez-vous. In: Zacharias M, editor. Protein-protein complexes: analysis, modeling and drug design. London: Imperial College Press; 2010. pp 183–209.
8. van Dijk AD, de Vries SJ, Dominguez C, Chen H, Zhou HX, Bonvin AM. Data-driven docking: HADDOCK's adventures in CAPRI. Proteins 2005;60:232–238.
9. de Vries SJ, Van Dijk M, Bonvin AMJJ. The HADDOCK web server for data-driven biomolecular docking. Nat Protocols 2010;5:883–897.
10. Karaca E, Melquiond ASJ, de Vries SJ, Kastritis PL, Bonvin AMJJ. Building macromolecular assemblies by information-driven docking: introducing the HADDOCK multi-body docking server. Mol Cell Proteomics 2010.
11. Sali A, Blundell TL. Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
12. Fiser A, Do RK, Sali A. Modeling of loops in protein structures. Prot Sci 2000;9:1753–1773.
13. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 2008;9.
14. Jorgensen WL, Tirado-rives J. The OPLS Potential functions for proteins. Energy minimizations for crystals of cyclin peptides and crambin. J Am Chem Soc 1988;110:1657–1666.
15. Fernandez-Recio J, Totrov M, Abagyan R. Identification of protein-protein interaction sites from docking energy landscapes. J Mol Biol 2004;335:843–865.
16. Leulliot N, Chaillet M, Durand D, Ulryck N, Blondeau K, van Tilbeurgh H. Structure of the yeast tRNA m7G methylation complex. Structure 2008;16:52–61.
17. de Vries SJ, van Dijk AD, Bonvin AM. WHISCY: what information does surface conservation yield? Application to data-driven docking. Proteins 2006;63:479–489.
18. Tong Y, Chugha P, Hota PK, Alviani RS, Li M, Tempel W, Shen L, Park HW, Buck M. Binding of Rac1, Rnd1, and RhoD to a novel Rho GTPase interaction motif destabilizes dimerization of the plexin-B1 effector domain. J Biol Chem 2007;282:37215–37224.
19. Micheelsen PO, Vevodova J, De Maria L, Ostergaard PR, Friis EP, Wilson K, Skjot M. Structural and mutational analyses of the interaction between the barley alpha-amylase/subtilisin inhibitor and the subtilisin savinase reveal a novel mode of inhibition. J Mol Biol 2008;380:681–690.
20. de Vries SJ.How proteins get in touch. Interface prediction and docking of protein complexes. PhD thesis, Utrecht University, 2009.
21. Kufareva I, Budagyan L, Raush E, Totrov M, Abagyan R. PIER: protein interface recognition for structural proteomics. Proteins 2007;67:400–417.
22. Neuvirth H, Raz R, Schreiber G. ProMate: a structure based prediction program to identify the location of protein-protein binding sites. J Mol Biol 2004;338:181–199.
23. Chen H, Zhou HX. Prediction of interface residues in protein-protein complexes by a consensus neural network method: test against NMR data. Proteins 2005;61:21–35.
24. Porollo A, Meller J. Prediction-based fingerprints of protein-protein interactions. Proteins 2007;66:630–645.
25. Liang S, Zhang C, Liu S, Zhou Y. Protein binding site prediction using an empirical scoring function. Nucleic Acids Res 2006;34:3698–3707.
26. Vallee F, Kadziola A, Bourne Y, Juy M, Rodenburg KW, Svensson B, Haser R. Barley alpha-amylase bound to its endogenous protein inhibitor BASI: crystal structure of the complex at 1.9 A resolution. Structure 1998;6:649–659.
27. Tong J, Jiang P, Lu ZH. RISP: a web-based server for prediction of RNA-binding sites in proteins. Comput Methods Prog Biomed 2008;90:148–153.
28. Terribilini M, Sander JD, Lee JH, Zaback P, Jernigan RL, Honavar V, Dobbs D. RNABindR: a server for analyzing and predicting RNA-binding sites in proteins. Nucleic Acids Res 2007;35:W578–W584.
29. Kumar M, Gromiha MM, Raghava GP. Prediction of RNA binding sites in a protein using SVM and PSSM profile. Proteins 2008;71:189–194.
30. Isabet T, Montagnac G, Regazzoni K, Raynal B, El Khadali F, England P, Franco M, Chavrier P, Houdusse A, Menetrey J. The structural basis of Arf effector specificity: the crystal structure of ARF6 in a complex with JIP4. EMBO J 2009;28:2835–2845.
31. O'Shea EK, Klemm JD, Kim PS, Alber T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science 1991;254:539–544.
32. Strelkov SV, Herrmann H, Geisler N, Wedig T, Zimbelmann R, Aebi U, Burkhard P. Conserved segments 1A and 2B of the intermediate filament dimer: their atomic structures and role in filament assembly. EMBO J 2002;21:1255–1266.
33. Shiba T, Koga H, Shin HW, Kawasaki M, Kato R, Nakayama K, Wakatsuki S. Structural basis for Rab11-dependent membrane recruitment of a family of Rab11-interacting protein 3 (FIP3)/Arfophilin-1. Proc Natl Acad Sci USA 2006;103:15416–15421.
34. Li JJ, Williams BL, Haire LF, Goldberg M, Walker E, Durocher D, Yaffe MB, Jackson SP, Smerdon SJ. Structural and functional versa-

tility of the FHA domain in DNA-damage signaling by the tumor suppressor kinase Chk2. Mol Cell 2002;9:1045–1054.
35. Sreerama YN, Gowda LR. Antigenic determinants and reactive sites of a trypsin/chymotrypsin double-headed inhibitor from horse gram (Dolichos biflorus). Biochimica Et Biophysica Acta Prot Struct Mol Enzymology 1997;1343:235–242.
36. Laskowski M, Qasim MA. What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes? Biochim Biophys Acta 2000;1477:324–337.
37. Katz BA, Finer-Moore J, Mortezaei R, Rich DH, Stroud RM. Episelection: novel Ki approximately nanomolar inhibitors of serine proteases selected by binding or chemistry on an enzyme surface. Biochemistry 1995;34:8264–8280.
38. Bao R, Zhou CZ, Jiang C, Lin SX, Chi CW, Chen Y. The ternary structure of the double-headed arrowhead protease inhibitor API-A complexed with two trypsins reveals a novel reactive site conformation. J Biol Chem 2009;284:26676–26684.
39. Meenan NAG, Sharma A, Fleishman SJ, MacDonald C, Morel B, Boetzel R, Moore GR, Baker D, Kleanthous C. The structural and energetic basis for high selectivity in a high affinity protein-protein interaction. Proc Natl Acad Sci USA 2010;107:10080–10085.
40. Montalvao RW, Cavalli A, Salvatella X, Blundell TL, Vendruscolo M. Structure determination of protein-protein complexes using NMR chemical shifts: case of an endonuclease colicin-immunity protein complex. J Am Chem Soc 2008;130:15990–15996.
41. Ko TP, Liao CC, Ku WY, Chak KF, Yuan HS. The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. Structure 1999;7:91–102.
42. Li W, Hamill SJ, Hemmings AM, Moore GR, James R, Kleanthous C. Dual recognition and the role of specificity-determining residues in colicin E9 DNase-immunity protein interactions. Biochemistry (NY) 1998;37:11771–11779.
43. Kuhlmann UC, Pommer AJ, Moore GR, James R, Kleanthous C. Specificity in protein-protein interactions: the structural basis for dual recognition in endonuclease colicin-immunity protein complexes. J Mol Biol 2000;301:1163–1178.
44. Krachler AM, Sharma A, Kleanthous C. Self-association of TPR domains: lessons learned from a designed, consensus-based TPR oligomer. Proteins 2010;78:2131–2143.
45. Main ER, Xiong Y, Cocco MJ, D'Andrea L, Regan L. Design of stable alpha-helical arrays from an idealized TPR motif. Structure 2003;11:497–508.
46. Kajander T, Cortajarena AL, Mochrie S, Regan L. Structure and stability of designed TPR protein superhelices: unusual crystal packing and implications for natural TPR proteins. Acta Crystallogr D Biol Crystallogr 2007;63:800–811.
47. Inbar Y, Benyamini H, Nussinov R, Wolfson HJ. Prediction of multimolecular assemblies by multiple docking. J Mol Biol 2005;349:435–447.
48. Sircar A, Gray JJ. SnugDock: paratope structural optimization during antibody-antigen docking compensates for errors in antibody homology models. PLoS Comput Biol 2010;6:e1000644.
49. Ben-Zeev E, Kowalsman N, Ben-Shimon A, Segal D, Atarot T, Noivirt O, Shay T, Eisenstein M. Docking to single-domain and multiple-domain proteins: old and new challenges. Proteins 2005;60:195–201.
50. Hendrix DK, Klein TE, Kuntz ID. Macromolecular docking of a three-body system: the recognition of human growth hormone by its receptor. Protein Sci 1999;8:1010–1022.
51. DeLano WL. The PyMOL molecular graphics system. http://www.pymol.org, 2002.