

**Building macromolecular assemblies by information-driven docking:
Introducing the HADDOCK multi-body docking server.**

**Ezgi Karaca[#], Adrien S.J. Melquiond[#], Sjoerd J. de Vries[#], Panagiotis L. Kastritis and
Alexandre M.J.J. Bonvin***

Bijvoet Center for Biomolecular Research, Science Faculty, Utrecht University

Padualaan 8, 3584 CH Utrecht, The Netherlands

These authors contributed equally to this work.

*To whom correspondence should be addressed. Tel: +31 30 2533859; Fax: +31 30 2537623;

Email: a.m.j.j.bonvin@uu.nl

Keywords: 3D structure prediction; multi-body docking; symmetrical homo-oligomers;
biomolecular complexes; symmetry restraints; bioinformatic predictions.

Running Title: Multi-component Docking

ABBREVIATIONS

AIR	ambiguous interaction restraint
CAPRI	critical assessment of prediction of interactions
CNS	crystallography & NMR system
cryo-ET	cryo electron tomography
cryo-EM	cryo electron microscopy
E2A	glucose-specific enzyme IIA
GAP	GTPase activating protein
GTPase	GTP hydrolase
HMS-PCI	high-throughput mass-spectrometry protein complex identification
HPr	histidine-containing phosphocarrier protein
IM-MS	ion mobility separation coupled to mass spectrometry
MS	mass spectroscopy
NCS	non-crystallographic symmetry
NMR	nuclear magnetic resonance
PCA	protein fragment complementation assay
PDB	protein data bank
PPI	protein-protein interactions
RMSD	root mean square deviation
RP-MS	radical probe mass spectrometry
SAXS	small angle x-ray scattering
SY2H	systematic yeast two-hybrid screening
TAP-MS	tandem affinity purification coupled to mass spectrometry

SUMMARY

Over the last years, large-scale proteomic studies are generating a wealth of information of biomolecular complexes. Adding the structural dimension to the resulting interactomes represents a major challenge that classical structural experimental methods alone will have difficulties to cope with. To meet this challenge, complementary modeling techniques such as docking are thus needed. Among the current docking methods, HADDOCK distinguishes itself from others by the use of experimental and/or bioinformatics data to drive the modeling process and has shown a strong performance in CAPRI, a blind experiment for the prediction of interactions.

While most docking programs are limited to binary complexes, HADDOCK can deal with multiple molecules (up to 6), a capability that will be required to build large macromolecular assemblies. We present here a novel web interface of HADDOCK that allows the user to dock up to 6 biomolecules simultaneously. This interface allows the inclusion of a large variety of both experimental and/or bioinformatics data and supports several types of cyclic and dihedral symmetries in the docking of multi-body assemblies. The server was tested on a benchmark of six cases, containing five symmetric homo-oligomeric protein complexes and one symmetric protein-DNA complex. Our results reveal that, in the presence of either bioinformatics and/or experimental data, HADDOCK shows an excellent performance: in all cases HADDOCK is able to generate good to high-quality solutions and ranks them at the top, demonstrating its ability to model symmetric multi-component assemblies.

Docking methods can thus play an important role in adding the structural dimension to interactomes. However, although the current docking methodologies are successful for a vast range of cases, considering the variety and complexity of macromolecular assemblies, inclusion

of some kind of experimental information (e.g. from mass spectrometry, nuclear magnetic resonance, cryo-electron microscopy, etc) will remain highly desirable in order to obtain reliable results.

INTRODUCTION

Proteins are the wheels and millstones of the complex machinery that underlies human life. Catalyzing a huge diversity of chemical processes, proteins work in close association with other biomolecules: nucleic acids, sugars, lipids, and other proteins. This huge network of protein interactions enables the cell to respond quickly to changes in the environment, such as temperature, oxygen or nutrient concentration. However, to fully understand this network, insights at atomic level are needed.

In the wake of the elucidation of the human genome (1, 2), many structural genomics projects are solving the structures of what is now becoming a considerable fraction of the human proteome (3). These projects are now moving to the next level, which is solving the atomic-resolution structures of protein complexes. However, this is a challenge that is considerably greater than obtaining the structures of single proteins. First of all, a protein can take part in 10 interactions on average, thus the number of complexes is expected to be at least an order of magnitude larger than the proteome and their composition can even vary over time. Second, associations between subunits in protein complexes are often weak and reversible, which makes purification and crystallization difficult. Finally, there are some very well studied classes of interactions, such as enzyme-inhibitors, antibody-antigens and GTPase-GAP interactions, but these classes represent binary interactions between proteins. In contrast, many of the most important functions in the cell are carried out by large, dynamic molecular assemblies, such as the ribosome, the proteasome, the spliceosome, RNA polymerases and the nuclear pore complex (4, 5). For such assemblies, high-resolution methods such as X-ray crystallography and NMR spectroscopy often provide atomic-level information at the level of individual subunits or

subcomplexes, but they typically encounter difficulties at the level of the full complex.

Fortunately, low-resolution information about protein complexes can often be obtained. Affinity purification (6, 7) followed by mass spectrometry is a high-throughput technique to study the composition of a complex. However, dissociation inside the mass spectrometer can be a problem for transient or unstable complexes, in which case chemical cross-linking can help. Once the composition of the complex is known, there is a variety of experimental techniques available to obtain structural information on the complex. The most detailed information can be gathered by using data obtained from various NMR experiments, for example chemical shift perturbations (8) or residual dipolar couplings (9); unfortunately, NMR is limited to complexes that are fairly small in size, making its applicability in the context of large assemblies less suited. Techniques that provide information about the shape of a protein complex, such as small angle X-ray scattering (SAXS), cryo-electron tomography (cryo-ET), and single-molecule cryo-electron microscopy (cryo-EM) are more suited to characterize large complexes. Unfortunately, all of these techniques suffer from limitations in resolution that are either fundamental or caused by structural heterogeneities of the complex.

A well-known approach to obtain information on residues at an interface is site-directed mutagenesis (10). In principle, a loss of binding affinity indicates that the mutated residue mediates the interaction, although the reverse is not true. Also, one must take care of secondary effects, such as unfolding or conformational change caused by the mutation. Apart from that, very detailed information about interface residues can be obtained by extensive mutagenesis experiments, such as alanine scanning and double mutant cycles. Mass spectrometry offers the opportunity to get peptide-level or residue-level information about protein interfaces by accurate mass measurements of peptides from the protein complex, generated either *a priori* through

proteolytic cleavage, or inside the mass spectrometer (MS/MS). For example, interface residues can be identified as residues that undergo slower Hydrogen/Deuterium (H/D) exchange upon complex formation. This process can be monitored at the peptide level by mass spectrometry (or, in smaller complexes, at the residue level by NMR), though this method is very sensitive to noise caused by conformational changes upon binding. In the same way, Radical Probe Mass Spectrometry (RP-MS) uses differences in oxidation of residues by hydroxyl radicals generated in the mass spectrometer to identify interface residues. Finally, chemical cross-linking followed by MS can provide direct information about residue contact sites between different binding partners of the complex. Several cross-linking reagents can provide complementary information. However, it has been reported that the cross-linkers may disrupt the structure of the protein complex and that care should therefore be taken to interpret the results (11).

There is a need for computational approaches to translate this low-resolution information into atomic-resolution models that can provide functional and mechanistic insights. One of the most promising approaches is docking, the prediction of the structure of a complex starting from the free, unbound structures of its constituents. In recent years, docking methods have made much progress in the blind prediction of the structure of protein complexes, as is seen in the recent rounds of the CAPRI (Critical Assessment of PRediction of Interactions) experiment (<http://www.ebi.ac.uk/msd-srv/capri>) (12, 13). Most docking methods are *ab initio*, which means that experimental data are not required. However, it is possible in several *ab initio* methods to use experimentally determined interface residues in the docking: in MolFit (14, 15) and ATTRACT (16, 17), it is possible to up weight the interaction scores of interface residues, in ZDOCK(18, 19) it is possible to block non-interface residues and in PatchDock (20, 21), ZDOCK, pyDock (22, 23) and several other methods, it is possible to filter the docking results

based on experimental information. Next to purely *ab initio* approaches, there are also methods that make use of different types experimental information, for example PROXIMO (24), based on RP-MS data, and MultiFit (25), a hybrid fitting/docking approach based on electron microscopy data.

(Table 1)

A method that distinguishes itself from the variety of above-mentioned docking approaches is HADDOCK (26-28). In HADDOCK, the docking can be driven by a variety of experimental data, using information about interface, contacts and relative orientations inside a complex simultaneously. Originally developed for NMR data, HADDOCK is able to deal with a large variety of experimental data, as shown in Table 1. Interface residues are defined as *active residues* that are believed to participate in the formation of the interface, and *passive residues* that are possibly at the interface; other kinds of data can be entered directly. (See the original HADDOCK papers (26-28) and Methodology for more details). HADDOCK has performed very well in translating these data into structures and structural models. More than 60 PDB structures calculated using HADDOCK have been deposited to date as experimental structures in the Protein Data Bank (<http://www.rcsb.org>)(29). Moreover, HADDOCK has shown a strong performance in CAPRI. Finally, HADDOCK is a general-purpose program that can integrate many kinds of data, but even with a single source of data it is able to perform as well as more specialized programs: For example, HADDOCK was able to closely reproduce the NMR-calculated E2A-HPr complex using only chemical shift perturbation data. For the ribonuclease S-protein-peptide complex (PDB code 1J80 (30)), for which RP-MS data are available, PROXIMO

was able to closely reproduce the crystal structure (RMSD of the top-scoring model from the reference crystal structure is 1.26 Å); using the same data, HADDOCK could get even closer with a RMSD of only 0.68Å from the crystal structure (Results not shown).

Most docking methods are designed to deal with just two molecules, which makes their application limited with regard to large macromolecular assemblies. In most programs, multi-component complexes can be assembled by adding each component one at a time while simultaneous docking of the whole complex is typically not possible. Recently five *ab initio* docking programs (Molfit (31, 32), ClusPro (33), Rosetta (34), M-ZDOCK (35) and SymmDock (36)) gave birth to specific versions for the prediction of the symmetric multimers. Among these programs, Molfit, ClusPro and Rosetta are performing a rotational/translational search about the proper symmetry axes. These programs can deal with different types of cyclic and dihedral symmetries. Different than the other two, Rosetta is able to assemble complexes having helical and icosahedral symmetries. M-ZDOCK and SymmDock are suited for the prediction of macromolecules with cyclic symmetries. However, the ability to deal with arbitrary large molecular assemblies is currently rare. CombDock (37), which is developed by the team of SymmDock, can build hetero-oligomer complexes but it doesn't have a symmetry option. Only HADDOCK can deal with molecular complexes that are hetero-oligomers or homo-oligomers with arbitrary symmetry operators between and within each component.

The flexibility of HADDOCK comes at a price: it requires the user to have the structure calculation program CNS(38) installed and a considerable degree of expertise in its usage and molecular modeling in general, and it requires a cluster of computers. To alleviate this problem and to open up HADDOCK for a wide community we have recently developed the HADDOCK web server (27). The server offers multiple web interfaces, ranging from very simple and user-

friendly to very powerful and flexible, exposing the full range of HADDOCK options to the expert user. However, up until now, the HADDOCK server was unable to deal with more than two molecules. Here we present a novel web interface for multi-body docking of complexes. Like the HADDOCK program itself, the server supports the docking of up to 6 molecules simultaneously; all HADDOCK options, including symmetry restraints, are made available to the user. Even larger assemblies can in principle be modeled if the docking is performed in an incremental way. Here we demonstrate the performance of the multi-body server on a small benchmark comprising complexes of various symmetries and increasing numbers of components (from 3 to 5). To drive the docking, bioinformatic interface predictions and/or available experimental information were used. The HADDOCK server is available at <http://haddock.chem.uu.nl/servicesdevel/HADDOCK2/haddock.php>.

METHODOLOGY

Ambiguous interaction restraints and docking protocol

HADDOCK uses experimental and/or bioinformatics data to drive the complex formation *in silico*. The experimental and/or prediction data are used to define *active* and *passive residues*. *Active residues* are described as the identified interface residues and *passive residues* correspond to their solvent accessible neighbours. These are used to define a network of ambiguous interaction restraints (AIRs) between the molecules to be docked. An AIR defines that a residue on the surface of a biomolecule should be in close vicinity to another residue or group of residues on the partner biomolecule when they form the complex. By default this is described as an ambiguous distance restraint between all atoms of the source residue to all atoms of all target residue(s) that are assumed to be in the interface of the complex (Figure 1). The effective distance between all those atoms, d_{iAB}^{eff} is calculated as follows:

$$d_{iAB}^{eff} = \left(\sum_{m_{iA}=1}^{N_{Atom}} \sum_{k=1}^{N_{resB}} \sum_{n_{kB}=1}^{N_{Atom}} \frac{1}{d_{m_{iA}n_{kB}}^6} \right)^{-1/6} \quad (\text{Eq. 1})$$

(Figure 1)

Here N_{Atom} indicates all atoms of the source residue on molecule A, N_{resB} the residues defined to be at the interface of the target molecule B and N_{Atom} all atoms of a residue on molecule B. The $1/r^6$ summation somewhat mimics the attractive part of a Lennard-Jones potential and ensures that the AIRs are satisfied as soon as any two atoms of the biomolecules are in contact. The AIRs are incorporated as an additional energy term to the energy function that is

minimized during the docking. The ambiguous nature of these restraints easily allows experimental data that often provide evidence for a residue making contacts to be used as a driving force for the docking. As such the AIRs define a network of restraints between the possible interaction interface(s) of the molecules to be docked without defining the relative orientation of the molecules, minimizing the necessary search through conformational space needed to assemble the interfaces.

The docking protocol in HADDOCK consists of three stages:

- i) rigid body energy minimization (*it0*)
- ii) semi-flexible refinement in torsion angle space (*it1*)
- iii) a final explicit solvent refinement (*water*).

In the last two stages, flexible segments are typically defined automatically based on the identified intermolecular contacts. The solutions are ranked at the end of each docking stage based on the following HADDOCK scoring functions:

- $it0: 0.01 \cdot E_{vdW} + 1.0 \cdot E_{Elec} + 0.01 \cdot E_{AIR} - 0.01 \cdot BSA + 1.0 \cdot E_{Desolv} + 0.1 \cdot E_{Sym}$ (Eq. 2)

- $it1: 1.0 \cdot E_{vdW} + 1.0 \cdot E_{Elec} + 0.1 \cdot E_{AIR} - 0.01 \cdot BSA + 1.0 \cdot E_{Desolv} + 0.1 \cdot E_{Sym}$ (Eq. 3)

- $water: 1.0 \cdot E_{vdW} + 0.2 \cdot E_{Elec} + 0.1 \cdot E_{AIR} + 1.0 \cdot E_{Desolv} + 0.1 \cdot E_{Sym}$ (Eq. 4)

The weighted parameters that are used in different stages of the scoring are: van der Waals (E_{vdW}), electrostatics (E_{Elec}), restraint violation (E_{AIR}), desolvation (E_{Desolv}) (39) symmetry restraint energies, and buried surface area (BSA). The solutions are clustered using a 7.5Å cut-off based on their pairwise RMSD values and the cluster ranks are determined according to the average energy of the four best structures of each cluster.

Dealing with symmetry

HADDOCK can deal with biomolecules having cyclic (C2, C3, C5) symmetries or any combination thereof. This also allows dealing with dihedral symmetries since dihedral symmetry can be interpreted as a combination of cyclic symmetry pairs (e.g. D2 symmetry is a combination of 6 C2 symmetry pairs (see Table 2.)). The symmetry restraints can be applied both within and between molecules. Compared to other docking programs supporting symmetric molecules, the unique characteristic of HADDOCK is that it applies symmetry on the molecules while docking them simultaneously. For the generation of symmetric complexes, two types of restraints should be used in combination: NCS (non-crystallographic symmetry) (40) and distance symmetry restraints (41, 42), both available within CNS.

(Table 2)

Non-crystallographic symmetry. NCS restraints force two (or more) monomers to be identical without defining any symmetry operation between them. This is achieved through minimization of the following potential energy function:

$$E_{NCS} = k_{NCS} \sum_{a=1}^A \sum_{m=1}^M (x'_{am} - \bar{x}_a)^2 + (y'_{am} - \bar{y}_a)^2 + (z'_{am} - \bar{z}_a)^2 \quad (\text{Eq. 5})$$

This energy term is calculated after superposition of the monomers onto the first monomer. In the potential expression, A is the number of atoms, M is the number of monomers, k_{NCS} is a constant, $(x'_{am}, y'_{am}, z'_{am})$ are the Cartesian coordinates of the a^{th} atom on the m^{th} monomer and

$(\bar{x}_a, \bar{y}_a, \bar{z}_a)$ corresponds to the average position of a^{th} atom with respect to the superimposed coordinates (41). Using NCS restraints in HADDOCK only requires the user to define pairs of segments on which the NCS restraints will be applied. These can belong either to the same molecule or to separate molecules allowing to define both intra- and intermolecular symmetries. The only requirement is that the number and type of atoms should be identical in both segments.

Cyclic and Dihedral Symmetry. The implementation of this type of symmetry in HADDOCK is based on the symmetry distance restraints defined by Nilges (41, 42) for the NMR structure calculation of symmetrical dimers. The symmetry is imposed by requiring that pairs of intermolecular distances between all symmetric C_n atoms should have identical values. In the case of a dimer composed of molecules A and B, this condition can be illustrated as follows:

$$\Delta = d(A_i, B_j) - d(B_i, A_j) \quad (\text{Eq. 6})$$

Δ is summed over all distances between C_n atoms of the defined segments. Here the idea is to minimize Δ , so that the symmetric distances between the monomers are equal to each other. This is illustrated in Table 2. The major advantage of this approach is that it does not require knowledge of the position of the symmetry axis and it can be applied to different symmetries (C2, C3, C5 as shown in Table 2) and oligomeric proteins. The symmetric pairs should be defined as explained above for NCS restraints.

Docking procedure for symmetric complexes

All test cases, except for the protein-DNA complex were docked using the Multi-body web interface of HADDOCK (<http://haddock.chem.uu.nl/servicesdevel/HADDOCK2/haddock.php>). The procedure followed to dock the protein-DNA complex differs from the generic multi-body docking protocol in the sense that two subsequent docking rounds were performed, with, in the second round custom-built DNA models that captured the conformational changes in the DNA from the first docking are used as starting structures. This approach allows to model rather large deformations in the DNA and is explained comprehensively in a recent work of van Dijk *et al*¹.

In four of the test cases (1QU9, 1OUS, 1VIM, 1VPN), the interface information was obtained through the consensus interface prediction server CPORT² (<http://haddock.chem.uu.nl/servicesdevel/CPORT/cport.html>) using the “very sensitive” option. In the case of 1URZ, a former CAPRI target, we used the same interface definition as was used previously in CAPRI (43) For the protein-DNA complex, 3CRO, sequence conservation and experimental data (mutagenesis, ethylation interference) were used to define the protein-DNA interaction site.

The interface information was converted into Ambiguous Interaction Restraints (AIR_s) via the setup page of the HADDOCK website (<http://www.nmr.chem.uu.nl/haddock>). The generated AIR files, together with the input structures were then supplied to the multi-body server as an input for the docking. To favor compactness of the solution, center of mass restraints were turned on. For each complex the proper combination of NCS and symmetry restraints were defined. Sampling of 180° rotated solutions was disabled. The number of structures was increased to 5000, 400 and 400 for *it0*, *it1* and *water*, respectively. All other parameters were left to their default settings.

Evaluation of the docking models

The models were evaluated according to the CAPRI criteria (13) For a complex to be classified as acceptable (1*), its interface root mean square deviation (i-RMSD) from the complex has to be lower than 4 Å or its ligand RMSD (l-RMSD) lower than 10 Å. In addition, the fraction of native contacts (fnat) had to be ≥ 0.1 . For good predictions (2*), the criteria were $i\text{-RMSD} < 2$ or $l\text{-RMSD} < 5$, and $fnat \geq 0.3$. For high quality predictions (3*), the criteria were $i\text{-RMSD} < 1$ or $l\text{-RMSD} < 1$, and $fnat \geq 0.5$. A cluster was considered one-/two-/three-star if it least one of its top four members was of one-/two-/three-star quality or better.

RESULTS

We have compiled a benchmark of six multimer assemblies. The complexes are homomeric with different numbers of components and various symmetries (see Table 3). One of them corresponds to a dimeric protein – DNA complex. In four cases the docking was performed starting from the separated components of the crystal structure (“bound docking”). In one case (PDB ID: 1URZ), the starting structures correspond to the dimeric form of the complex while the trimeric form had to be predicted; this complex corresponds to a viral envelope protein that was a target in CAPRI (target 10). For the protein-DNA complex (3CRO) the docking was performed from the unbound conformation of the monomers and a canonical B-DNA model. In summary, our benchmark consists of 4 bound cases and 2 unbound cases.

(Table 3)

For modeling of the benchmark complexes, we made use of the new multi-body interface of the HADDOCK web server (<http://haddock.chem.uu.nl/servicesdevel/HADDOCK2/haddock.php>). The web server provides a user-friendly interface that gives full control over the various HADDOCK parameters and supports wide range of experimental restraints (Figure 2). This interface is freely accessible to non-profit users requiring “guru” access. It allows the simultaneous docking of up to six molecules and supports several types of cyclic symmetries (C2, C3, C5) and any type of dihedral symmetry that can be expressed as a combination of the available cyclic symmetry pairs (see

Methodology). Our server is the first to support cyclic and dihedral symmetries at the same time and to allow simultaneous docking of up to six molecules.

(Figure 2)

The performance of our multi-body docking approach was demonstrated for six complexes (Table 3) using a combination of experimental and/or bioinformatic predictions. For four of the complexes (1QU9, 1OUS, 1VIM, 1VPN), *active* and *passive residues* were defined based on consensus bioinformatics interface predictions from CPORT (see Methodology). For the other two complexes, a combination of experimental and predicted information was used. The list of *active* and *passive residues* for each complex is given in Table 4. Using the above information, the HADDOCK multi-body server produced and ranked at the top high quality models, demonstrating the excellent performance of our approach. Both the top ranked models and the top ranked clusters according to the HADDOCK score contained at least a medium quality (2*) prediction (See Table 5). Furthermore, analysis of the results showed that the imposed symmetries are fulfilled.

(Tables 4 and 5)

In four cases, bound docking was performed, including a trimer (1QU9), two tetramers (1OUS, 1VIM) and a pentamer (1VPN). For each of them the first ranked HADDOCK model corresponds to a high-quality prediction. Considering the increased docking complexity due to the large interaction space to be sampled in the case of multi-component systems, this

demonstrates an outstanding performance. In the two unbound cases consisting of a CAPRI target and a protein-DNA complex, good results (2* quality predictions) were also obtained (See Table 5). The performance of the protein-DNA docking (3CRO) and the ability of HADDOCK to catch the conformational changes in the DNA demonstrate that HADDOCK's excellent capabilities are not limited to just protein-protein complexes. An overlay of the top predictions onto their respective reference crystal structures is shown in Figure 3.

(Figure 3)

DISCUSSION/PERSPECTIVES

Pushing back the limits of structural prediction of macromolecular assemblies

In the structural characterization of biomolecules, most of the developments on the modeling side are limited to rather “small” binary systems, often only applicable to proteins. Just a few molecular docking programs can deal with multi-body assemblies and they are generally restricted to the prediction of symmetric homomeric complexes (32-37). So far, HADDOCK (26, 28, 43) is the only molecular docking program which is able to perform simultaneous docking of multi-body complexes, up to six components. A multi-body docking server was recently released on the HADDOCK web portal, allowing through an user-friendly interface the users to exploit the full range of experimental data supported by HADDOCK and to fully customize the docking process. The performance of this web server was evaluated in this paper against a benchmark set of six multimeric complexes, including a protein-DNA complex. Cyclic or dihedral symmetries, which are present in the large majority of homomers (44), were defined, combined with the interface information derived either from experimental evidences and/or predictions made by our consensus interface prediction program, CPORT. The results show that HADDOCK is able to generate native to near-native predictions for all cases with i-RMSD values for the best model ranging between 0.7 and 2.2 Å. Although we could produce excellent results even with the inclusion of bioinformatic predictions, which usually contain considerable amount of false positives, one should always keep in mind that the information supplied to HADDOCK should be of high quality. This is due to the complexity of the interaction space that is larger in the case of multi-body docking compared to the two-body docking.

The vast quantity of low-resolution experimental data that could further be used in

HADDOCK paves the route for the prediction of large macromolecular complexes. By combining distance and interaction restraints from low-resolution methods with molecular docking, architectural or even atomic models might be generated. These restraints can be derived from a variety of experimental measurements including MS of intact complexes, chemical cross-linking, cryo-EM, SAXS, fluorescence resonance energy transfer and analytical ultracentrifugation. One very recent addition to this series of biophysical tools is ion mobility separation coupled to mass spectrometry (IM-MS) (45). IM is an established technique for studying shape and conformation of small molecules and individual proteins. When coupled with MS, mass and subunit composition of a protein complex can be determined simultaneously with its overall topology and shape (46, 47). The cross-sections of amyloid oligomers formed in the early steps of amyloid fibril formation calculated by IM-MS (48) could be used as a restraint in data-driven docking to discriminate between quaternary topologies for a specific oligomeric state. This can be done for example by inferring a radius of gyration restraint from this cross-section measurement or by predicting the cross-sections from the docking models and using the experimental data as a filter.

Are we ready to predict interactomes from the 3D structures of biomolecules?

In today's proteomic era, large-scale screening techniques are used to characterize protein-protein interactions (PPI's) in vivo (49, 50). Despite the massive number of interactions detected by protein complex purification techniques using MS (originally either by high-throughput MS protein complex identification (51), HMS-PCI, or coupled to Tandem Affinity Purification (52), TAP-MS); Systematic Yeast Two-Hybrid screening (53-55), SY2H; and complementary mapping techniques (e.g. Protein Fragment Complementation Assay (56), PCA;

in vitro proteome chip screening (57)), the interactome coverage remains low, roughly 50% and 10% for the yeast and human interactomes, respectively (58). This becomes apparent in the rather limited overlap between various data sets obtained with different approaches (59). This can be explained by a limit in proteome coverage (up to 70% for the best approaches) and by the inherent high fraction of false positives (previous estimations mention that more than half of all current high-throughput data are spurious (59)). It also highlights the difficulties encountered by some methods for certain types of interactions, strengthening the complementarities between the different techniques. Finally, proteomic datasets derived to map PPI's, even when a similar detection method is used (7, 60), have a limited overlap (only 18%) (61).

Computational methods to predict protein assemblies could in principle play a complementary role in the study of interactomes, providing additional insights with leverage of the structural models. But can present scoring functions used in protein-protein docking methods characterize the binding affinity of a macromolecular complex, a requisite to predict interactomes? To answer this question, we have tested 9 of the currently best performing scoring functions against a large set of high-quality binding affinity data derived from literature(62). The results (data not shown) reveal that scoring is orthogonal to binding affinity prediction (the highest calculated r^2 was 0.09!), even though scoring functions are successfully being used in discriminating native structure from decoys. Hence, even if structural modeling tools and molecular docking approaches can significantly improve the selection accuracy of PPI networks (63), these computational methods need to be optimized for both purposes, e.g. annotation and prediction of PPI's.

The need for combining experimental information and modeling

By combining a variety of experimental approaches one can easily increase our knowledge about biologically relevant interaction (62). The experimental information can guide large-scale docking studies to upgrade the information contained in interactome maps by adding the 3D structural dimension to the PPI's. Moving towards systems biology, computational methods could aim at predicting how the proteome is wired and how dynamic changes in the interactome occur in response to different environmental factors. In that regard, mass spectrometry techniques that determine the composition and stoichiometry of macromolecular complexes will be of indispensable value.

But how far are we from a high-throughput method to screen for protein complex structures? Recently, we have linked HADDOCK to MTMDAT, an automated software for the analysis of mass spectrometry data (65), creating effectively a pipeline for high-throughput, MS-based structural modeling of complexes³. This pipeline allows feeding automatically into HADDOCK the interface information identified by MS from digestion experiments. This is only one example of how experiments and modeling can be coupled and we expect that many other related applications would be developed in the future to open the route to large-scale annotation of interactomes.

ACKNOWLEDGMENTS

This work was supported by the Netherlands Organization for Scientific Research [VICI grant #700.56.442 to A.B.] and the European Community [FP6 integrated Project SPINE2-COMPLEX contract no. 032220, and FP7 e-Infrastructure “e-NMR” I3 project, grant number 213010]. The Dutch BiG Grid project with financial support from the Netherlands Organization for Scientific Research (NWO) is acknowledged for the use of the computing and storage facilities.

REFERENCES

1. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.
2. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., et al. (2001) The sequence of the human genome. *Science* 291, 1304-1351.
3. Xie, L., and Bourne, P. E. (2005) Functional coverage of the human genome by existing structures, structural genomics targets, and homology models. *PLoS Comput. Biol.* 1, e31.
4. Alber, F., Forster, F., Korkin, D., Topf, M., and Sali, A. (2008) Integrating diverse data for structure determination of macromolecular assemblies. *Annu. Rev. Biochem.* 77, 443-477.
5. Mueller, M., Jenni, S., and Ban, N. (2007) Strategies for crystallization and structure determination of very large macromolecular assemblies. *Curr. Opin. Struct. Biol.* 17, 572-579.
6. Collins, S. R., Kemmeren, P., Zhao, X. C., Greenblatt, J. F., Spencer, F., Holstege, F. C., Weissman, J. S., and Krogan, N. J. (2007) Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics* 6, 439-450.
7. Krogan, N. J., Cagney, G., Yu, H. Y., Zhong, G. Q., Guo, X. H., Ignatchenko, A., Li, J., Pu, S. Y., Datta, N., et al. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440, 637-643.
8. van Dijk, A. D., Kaptein, R., Boelens, R., and Bonvin, A. M. (2006) Combining NMR relaxation with chemical shift perturbation data to drive protein-protein docking. *J. Biomol. NMR* 34, 237-244.
9. van Dijk, A. D., Fushman, D., and Bonvin, A. M. (2005) Various strategies of using residual dipolar couplings in NMR-driven protein docking: application to Lys48-linked di-ubiquitin and validation against ¹⁵N-relaxation data. *Proteins* 60, 367-381.
10. Cunningham, B. C., Jhurani, P., Ng, P., and Wells, J. A. (1989) Receptor and antibody epitopes in human growth hormone identified by homolog-scanning mutagenesis. *Science* 243, 1330-1336.
11. Peters, K., and Richards, F. M. (1977) Chemical cross-linking: reagents and problems in studies of membrane structure. *Annu. Rev. Biochem.* 46, 523-551.
12. Janin, J., Henrick, K., Moulton, J., Ten Eyck, L., Sternberg, M. J. E., Vajda, S., Vasker, I., and Wodak, S. J. (2003) CAPRI: A Critical Assessment of PRedicted Interactions. *Proteins* 52, 2-9.
13. Mendez, R., Leplae, R., De Maria, L., and Wodak, S. J. (2003) Assessment of blind predictions of protein-protein interactions: current status of docking methods. *Proteins* 52, 51-67.
14. Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C., and Vakser, I. A. (1992) Molecular-Surface Recognition - Determination of Geometric Fit between Proteins and Their Ligands by Correlation Techniques. *Proc. Natl. Acad. Sci. U. S. A.* 89, 2195-2199.
15. Ben-Zeev, E., and Eisenstein, M. (2003) Weighted geometric docking: Incorporating external information in the rotation-translation scan. *Proteins* 52, 24-27.
16. Zacharias, M. (2003) Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci.* 12, 1271-1282.
17. Zacharias, M. (2004) Rapid protein-ligand docking using soft modes from molecular

- dynamics simulations to account for protein deformability: binding of FK506 to FKBP. *Proteins: Struct., Funct., Bioinf.* 54, 759-767.
18. Chen, R., Li, L., and Weng, Z. (2003) ZDOCK: an initial-stage protein-docking algorithm. *Proteins* 52, 80-87.
 19. Pierce, B., and Weng, Z. (2007) ZRANK: reranking protein docking predictions with an optimized energy function. *Proteins* 67, 1078-1086.
 20. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H. J. (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* 33, W363-W367.
 21. Schneidman-Duhovny, D., Inbar, Y., Polak, V., Shatsky, M., Halperin, I., Benyamini, H., Barzilai, A., Dror, O., Haspel, N., et al. (2003) Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. *Proteins: Struct., Funct., Bioinf.* 52, 107-112.
 22. Grosdidier, S., Pons, C., Solernou, A., and Fernandez-Recio, J. (2007) Prediction and scoring of docking poses with pyDock. *Proteins* 69, 852-858.
 23. Man-Kuang Cheng, T., Blundell, T. L., and Fernandez-Recio, J. (2007) pyDock: electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. *Proteins* 68, 503-515.
 24. Gerega, S. K., and Downard, K. M. (2006) PROXIMO--a new docking algorithm to model protein complexes using data from radical probe mass spectrometry (RP-MS). *Bioinformatics* 22, 1702-1709.
 25. Lasker, K., Topf, M., Sali, A., and Wolfson, H. J. (2009) Inferential optimization for simultaneous fitting of multiple components into a CryoEM map of their assembly. *J. Mol. Biol.* 388, 180-194.
 26. De Vries, S. J., van Dijk, A. D. J., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, T., and Bonvin, A. M. J. J. (2007) HADDOCK versus HADDOCK: New features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins: Struct., Funct., Bioinfo.* 69, 726-733.
 27. De Vries, S. J., Van Dijk, M., and Bonvin, A. M. (in Press) The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.*
 28. Dominguez, C., Boelens, R., and Bonvin, A. M. J. J. (2003) HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* 125, 1731-1737.
 29. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235-242.
 30. Ratnaparkhi, G. S., and Varadarajan, R. (2001) Osmolytes stabilize ribonuclease S by stabilizing its fragments S protein and S peptide to compact folding-competent states. *J. Biol. Chem.* 276, 28789-28798.
 31. Berchanski, A., and Eisenstein, M. (2003) Construction of molecular assemblies via docking: modeling of tetramers with D2 symmetry. *Proteins* 53, 817-829.
 32. Berchanski, A., Segal, D., and Eisenstein, M. (2005) Modeling oligomers with C_n or D_n symmetry: application to CAPRI target 10. *Proteins* 60, 202-206.
 33. Comeau, S. R., and Camacho, C. J. (2005) Predicting oligomeric assemblies: N-mers a primer. *J. Struct. Biol.* 150, 233-244.
 34. Andre, I., Bradley, P., Wang, C., and Baker, D. (2007) Prediction of the structure of symmetrical protein assemblies. *Proc. Natl. Acad. Sci. U. S. A.* 104, 17656-17661.
 35. Pierce, B., Tong, W. W., and Weng, Z. P. (2005) M-ZDOCK: a grid-based approach for C_n-symmetric multimer docking. *Bioinformatics* 21, 1472-1478.

36. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H. J. (2005) Geometry-based flexible and symmetric protein docking. *Proteins: Struct., Funct., Bioinf.* 60, 224-231.
37. Inbar, Y., Benyamini, H., Nussinov, R., and Wolfson, H. J. (2005) Prediction of multimolecular assemblies by multiple docking. *J. Mol. Biol.* 349, 435-447.
38. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 54 (Pt 5), 905-921.
39. Fernandez-Recio, J., Totrov, M., and Abagyan, R. (2004) Identification of protein-protein interaction sites from docking energy landscapes. *J. Mol. Biol.* 335, 843-865.
40. Brunger, A. T. (1992) *X-PLOR. A System for X-Ray Crystallography and NMR*, Yale University Press, New Haven, CT.
41. Nilges, M. (1993) A Calculation Strategy for the Structure Determination of Symmetrical Dimers by H-1-Nmr. *Proteins* 17, 297-309.
42. O'Donoghue, S. I., and Nilges, M. (1999) *Structure Computation and Dynamics in Protein NMR*, Kluwer Academic/Plenum Publishers, New York.
43. van Dijk, A. D. J., de Vries, S. J., Dominguez, C., Chen, H., Zhou, H. X., and Bonvin, A. M. J. J. (2005) Data-driven docking: HADDOCK's adventures in CAPRI. *Proteins: Struct., Funct., Bioinf.* 60, 232-238.
44. Levy, E. D., Erba, E. B., Robinson, C. V., and Teichmann, S. A. (2008) Assembly reflects evolution of protein complexes. *Nature* 453, 1262-1265.
45. Ruotolo, B. T., Giles, K., Campuzano, I., Sandercock, A. M., Bateman, R. H., and Robinson, C. V. (2005) Evidence for macromolecular protein rings in the absence of bulk water. *Science* 310, 1658-1661.
46. Ruotolo, B. T., Benesch, J. L. P., Sandercock, A. M., Hyung, S. J., and Robinson, C. V. (2008) Ion mobility-mass spectrometry analysis of large protein complexes. *Nat. Protoc.* 3, 1139-1152.
47. Smith, D. P., Knapman, T. W., Campuzano, I., Malham, R. W., Berryman, J. T., Radford, S. E., and Ashcroft, A. E. (2009) Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies. *European Journal of Mass Spectrometry* 15, 113-130.
48. Bernstein, S. L., Dupuis, N. F., Lazo, N. D., Wyttenbach, T., Condron, M. M., Bitan, G., Teplow, D. B., Shea, J. E., Ruotolo, B. T., et al. (2009) Amyloid-beta protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nat. Chem.* 1, 326-331.
49. Walhout, A. J. M., and Vidal, M. (2001) Protein interaction maps for model organisms. *Nat. Rev. Mol. Cell Biol.* 2, 55-62.
50. Auerbach, D., Fetchko, M., and Stagljar, I. (2003) Proteomic approaches for generating comprehensive protein interaction maps. *TARGETS* 2, 85-92.
51. Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., et al. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415, 180-183.
52. Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141-147.
53. Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D.,

- Narayan, V., Srinivasan, M., et al. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623-627.
54. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4569-4574.
 55. Yu, H. Y., Braun, P., Yildirim, M. A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., et al. (2008) High-quality binary protein interaction map of the yeast interactome network. *Science* 322, 104-110.
 56. Tarassov, K., Messier, V., Landry, C. R., Radinovic, S., Molina, M. M. S., Shames, I., Malitskaya, Y., Vogel, J., Bussey, H., et al. (2008) An in vivo map of the yeast protein interactome. *Science* 320, 1465-1470.
 57. Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., et al. (2001) Global analysis of protein activities using proteome chips. *Science* 293, 2101-2105.
 58. Hart, G. T., Ramani, A. K., and Marcotte, E. M. (2006) How complete are current yeast and human protein-interaction networks? *Genome Biology* 7, 120.1-120.9.
 59. von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., and Bork, P. (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417, 399-403.
 60. Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., et al. (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440, 631-636.
 61. Goll, J., and Uetz, P. (2006) The elusive yeast interactome. *Genome Biol.* 7, 223.1-223.6.
 62. Kastritis, P. L., and Bonvin, A. M. J. J. (in Press) Are scoring functions in protein protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. *J. Proteome Res.*
 63. Fernandez-Ballester, G., and Serrano, L. (2006) Prediction of protein-protein interaction based on structure. *Methods in Molecular Biology* 340, 207-234.
 64. Tong, A. H. Y., Drees, B., Nardelli, G., Bader, G. D., Brannetti, B., Castagnoli, L., Evangelista, M., Ferracuti, S., Nelson, B., et al. (2002) A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science* 295, 321-324.
 65. Hennig, J., Hennig, K. D. M., and Sunnerhagen, M. (2008) MTMDAT: Automated analysis and visualization of mass spectrometry data for tertiary and quaternary structure probing of proteins. *Bioinformatics* 24, 1310-1312.
 66. Volz, K. (1999) A test case for structure-based functional assignment: The 1.2 angstrom crystal structure of the yjgF gene product from *Escherichia coli*. *Protein Sci.* 8, 2428-2437.
 67. Bressanelli, S., Stiasny, K., Allison, S. L., Stura, E. A., Duquerroy, S., Lescar, J., Heinz, F. X., and Rey, F. A. (2004) Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. *EMBO J.* 23, 728-738.
 68. Loris, R., Tielker, D., Jaeger, K. E., and Wyns, L. (2003) Structural basis of carbohydrate recognition by the lectin LecB from *Pseudomonas aeruginosa*. *J. Mol. Biol.* 331, 861-870.
 69. Badger, J., Sauder, J. M., Adams, J. M., Antonysamy, S., Bain, K., Bergseid, M. G., Buchanan, S. G., Buchanan, M. D., Batiyenko, Y., et al. (2005) Structural analysis of a set of proteins resulting from a bacterial genomics project. *Proteins: Struct., Funct., Bioinf.* 60, 787-796.

70. Stehle, T., and Harrison, S. C. (1997) High-resolution structure of a polyomavirus VP1-oligosaccharide complex: implications for assembly and receptor binding. *Embo J.* 16, 5139-5148.
71. Mondragon, A., and Harrison, S. C. (1991) The Phage-434 Cro/Or1 Complex at 2.5Å Resolution. *J. Mol. Biol.* 219, 321-334.

FOOTNOTES

¹ van Dijk, M. and Bonvin, A. M. J. J. (2009) submitted

² de Vries, S. and Bonvin, A. M. J. J. in preparation

³ Hennig J., de Vries, S.J., Hennig, K.D.M., S. Sunnerghagen, S. and Bonvin, A. M. J. J. in preparation

FIGURE LEGENDS

Figure 1: Illustration of the ambiguous interaction restraints (AIRs) used in HADDOCK to drive the docking. *Active residues* correspond to residue experimentally identified or predicted to be at the interface. *Passive residues* are surface neighbors of *active residues*. AIRs are defined for each *active residue* with the effective distance being calculated from the sum of all individual distance between any atom of an *active residue* and any atom of all *active* and *passive residues* on the partner molecule (Eq. 1).

Figure 2: View of the HADDOCK's multi-body web interface for data-driven docking. (<http://haddock.chem.uu.nl/services/HADDOCK>).

Figure 3: View of the best HADDOCK solutions (having colored monomers) superimposed onto their respective crystal reference structures (shown in light grey) a. 1QU9 b. 1URZ c. 1OUS d. 1VIM e. 1VPN and f. 3CRO. The figures were generated with Pymol (Delano Scientific LLC, <http://www.pymol.org>).

TABLES

Table 1. The various experimental data that can be incorporated in to HADDOCK

Experimental data	HADDOCK representation
Mutagenesis data	<i>Active and passive residues</i>
H/D exchange data	<i>Active and passive residues</i>
Bioinformatic interface predictions	<i>Active and passive residues</i>
<i>Mass spectrometry data</i>	
Cross-linking data	Custom CNS restraints
Radical probe mass spectrometry	<i>Active and passive residues</i>
Limited proteolysis mass spectrometry	<i>Active and passive residues</i> or: directly, as an MTMDAT-generated HADDOCK parameter file
<i>NMR data</i>	
Chemical shift perturbation data	<i>Active and passive residues</i>
Cross-saturation experiments	<i>Active and passive residues</i>
Residual Dipolar Couplings	Directly
Diffusion anisotropy restraints	Directly
NOEs: as custom CNS restraints	Custom CNS restraints
Dihedral angles	Directly
Hydrogen bonds	Directly
Para-magnetic restraints	Under development
<i>Shape data</i>	
SAXS	Under development
EM	Under development

Table 2. Definition and illustration of the symmetry restraining options in HADDOCK.

C2		$d(A_i B_j) = d(B_i A_j)$
C3		$d(AB) = d(BC)$ $d(BC) = d(CA)$ $d(CA) = d(AB)$
C5		$d(AC) = d(AD)$ $d(BD) = d(BE)$ $d(CE) = d(CA)$ $d(DA) = d(DB)$ $d(EB) = d(EC)$
D2		$d(AB) = d(BA)$ $d(AC) = d(CA)$ $d(AD) = d(DA)$ $d(BC) = d(CB)$ $d(BD) = d(DB)$ $d(CD) = d(DC)$

Table 3. Properties of the Multimer Docking Benchmark

PDB ID	CATH Classification	Complex Type	Docking Type	Symmetry Type	Number of Amino acids
1QU9 (66)	Mainly Beta	Homotrimer	Bound	C3	128
1URZ (67)	Mainly Alpha / Mainly Beta	Homotrimer	Unbound	C3	400
1IOUS (68)	Alpha Beta	Homotetramer	Bound	D2	114
1VIM (69)	Alpha Beta	Homotetramer	Bound	D2	200
1VPN (70)	Mainly Beta	Homopentamer	Bound	C5	289
3CRO (71)	Mainly Alpha	Homodimer- Double stranded DNA	Unbound	C2	71 (Protein) 20 (DNA)

Table 4. List of *active* and *passive residues* used in HADDOCK in order to dock the various benchmark complexes.

PDB ID	<i>Active Residues</i>	<i>Passive Residues</i>
1QU9 ^a	3,4,6,7,8,11-18,21,28-31,33,69,72,73,75,77,81,82, 85,88,92,100-114,120,122,124	2,9,23,24,26,36,37,38,42,52,58,63,64,67,70,79,80, 83,86,89,90,93,96,97,98,99,115,116,118,126-128
1URZ ^b	5,8,9,10,11,13,54,71,73,75,76,78,79,87,93,98,110, 118,193,196,219,222,244,248,251,267,269,270	4,7,12,15,21,22,24,26,28,34,36,56,57,64,66-70, 72,77,81,83,86,92,94,95,96,107,108,120,131,150, 152-154,192,194,195,216-218,224,243,246,250, 253-263,266,271,272,273
1OUS ^a	3,15,17,19,41,42,47-52,71,76-87,89,91,93,98,99, 100-103,106,108,110,112-114	1,2,5,7,9,12,13,21,24,25,27,39,43,45,46,53,54,64, 66,68,69,70,72,73-75,96,97
1VIM ^a	-2-10,16,41-47,50,51,54,55,57,63-73,138,140- 142,144,145,147,150,151,154,155,158,159,162, 163,176-185	12-19,35,60-62,74,75,89,91,95,102,129,133-137, 146,165-168,170-174
1VPN ^a	32-38,52,71,74,75,78,79,107,111-119,123,127, 130-137,139,142,152,160,162,225,228,229,239, 240-245,250,252-260,264-269,274,275,288,289, 291,296,299,300,303,314,316	39-41,50,51,54,56,58,60,63-68,72,73,77,80,81,88, 93,101,102,104-106,108-110,117,124,126,128,138, 140,141,143-146,150,151,153-156,158,170,177, 179,183,185,231-236,238,244,246-249,251,261,

			262,276,290,292-295,297,305,307,309, 311,312
3CRO ^c	Protein: 29,31,32,42-44	DNA: 4-7,13-18,22-25,32-36	Protein: 9,18-20,27,28,30,34,36,37,40,41,45,46

The *active* and *passive residue* information is gathered via ^a CPORT, ^b CPORT and literature data, ^c conservation and experimental data (mutagenesis, ethylation interference).

Table 5. Multi-body docking results obtained via using the multi-body interface of HADDOCK web-server^a.

	Quality/ Rank	Best structure i-RMSD / l-RMSD (Å)	Best cluster Quality/Rank	Best cluster i-RMSD / l-RMSD (Å)
1QU9 ^b	★★★★ / 1	0.8 / 0.7	★★★★/1	0.8±0.1 / 0.7±0.1
1URZ ^{ub}	★★ / 1	1.7 / 5.2	★★/1	1.8±0.1 / 5.3±0.1
1OUS ^b	★★★★/1	0.9 / 1.2	★★★★/1	0.8±0.1 / 1.3±0.6
1VIM ^b	★★★★/1	1.0 / 1.2	★★★★/1	1.2±0.2 / 1.3±0.2
1VPN ^b	★★★★/1	0.7 / 0.7	★★/1	4.1±0.1 / 4.0±0.1
3CRO ^{ub}	★★ / 1	1.79 / 2.2	★★/1	2.12±0.3 / 2.8±0.6

^a For the definition of i-RMSD and l-RMSD refer to the Methodology section

^b Bound docking – the docking was performed with the separated monomers taken from the reference crystal structure

^c Unbound docking – the docking was performed with the free form of the monomers (see Methodology for details)

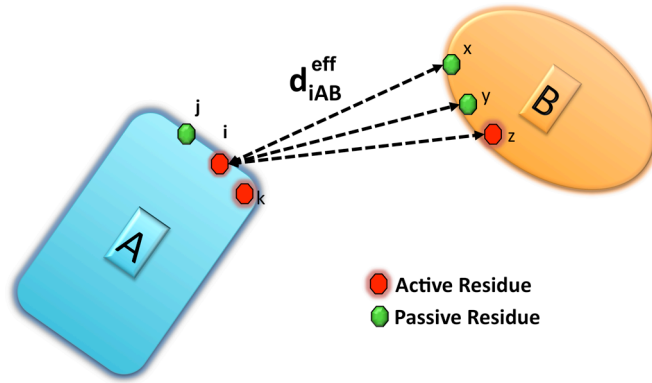


Figure 1

home >>

HADDOCK

Software web portal

Home **HADDOCK** Whisky DNA Publications Forum Contact

WELCOME TO THE UTRECHT BIOMOLECULAR INTERACTION WEB PORTAL >>

This is the multi-body interface to the HADDOCK docking program. This interface provides full control over HADDOCK parameters, including multi-body docking, and supports a wide range of experimental restraints. Unfold the menus by clicking on the double arrows. Submit your job by providing your username and password and press submit.

You may supply a name for your docking run (one word)

Name

Partner molecule 1 ▾

Structure definition

Where is the structure provided?

Which chain of the structure must be used?

PDB structure to submit

or: PDB code to download

Segment ID to use during the docking

What kind of molecule are you docking?

Histidine protonation states ⬆

Semi-flexible segments ⬆

Fully flexible segments ⬆

The N-terminus of your protein is positively charged

The C-terminus of your protein is negatively charged

Partner molecule 2 ⬆

Partner molecule 3 ⬆

Partner molecule 4 ⬆

Partner molecule 5 ⬆

Partner molecule 6 ⬆

Distance restraints ⬆

Sampling parameters ⬆

Parameters for clustering ⬆

Dihedral and hydrogen bond restraints ⬆

Noncrystallographic symmetry restraints ⬆

Symmetry restraints ⬆

Restraints energy constants ⬆

Residual dipolar couplings ⬆

Relaxation anisotropy restraints ⬆

Energy and interaction parameters ⬆

Scoring parameters ⬆

Advanced sampling parameters ⬆

Solvated docking parameters ⬆

Analysis parameters ⬆

Username and password

Username

Password

Home **HADDOCK** Whisky DNA Publications Forum Contact

2008 © NMR Department. All rights reserved. Webdesign by Marc van Dijk
XHTML | CSS

Figure 2

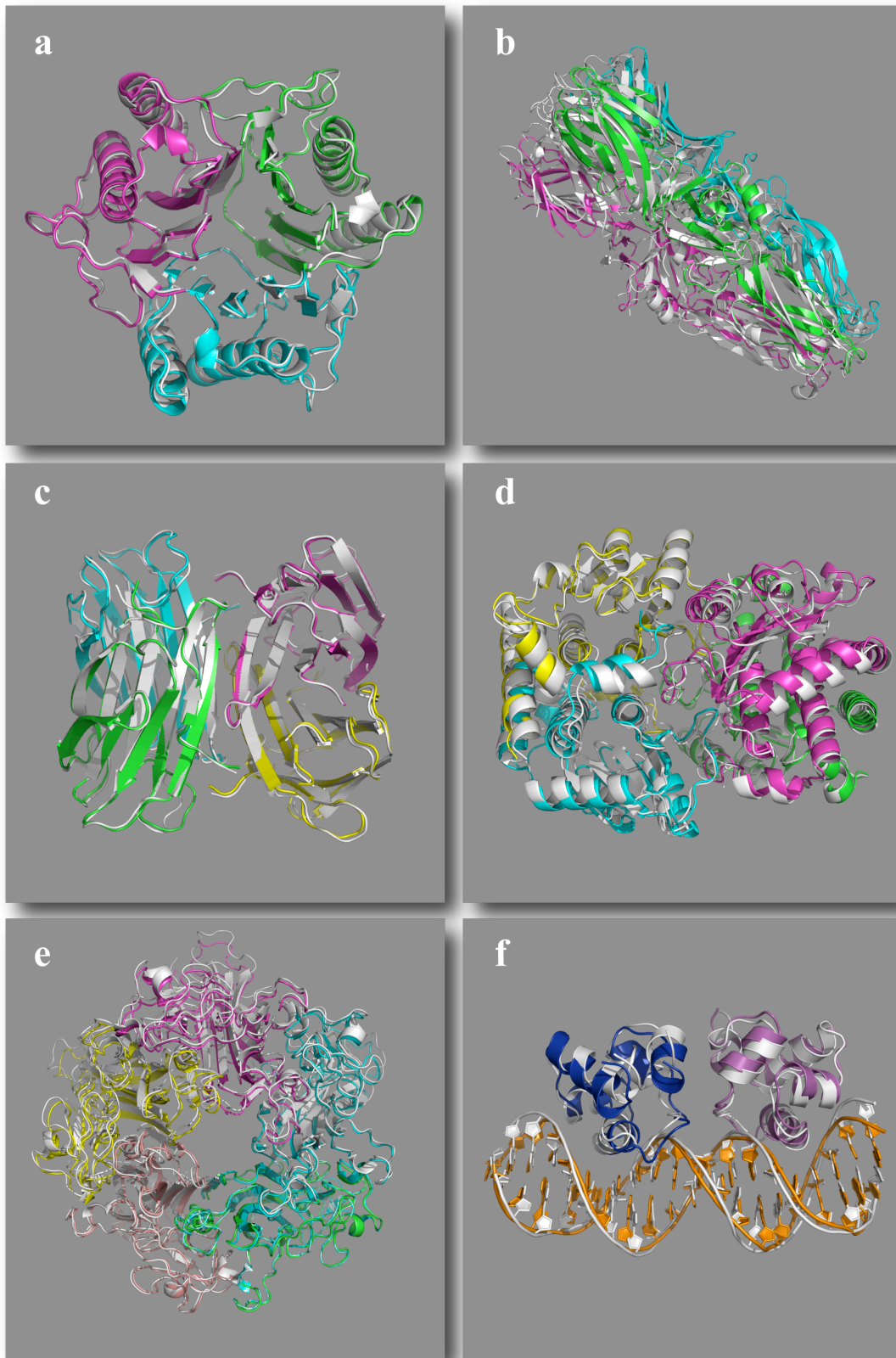


Figure 3