

✂ Author's Choice

Building Macromolecular Assemblies by Information-driven Docking

INTRODUCING THE HADDOCK MULTIBODY DOCKING SERVER*

Ezgi Karaca[‡], Adrien S. J. Melquiond[‡], Sjoerd J. de Vries[‡], Panagiotis L. Kastiris, and Alexandre M. J. J. Bonvin[§]

Over the last years, large scale proteomics studies have generated 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 confront. To meet this challenge, complementary modeling techniques such as docking are thus needed. Among the current docking methods, HADDOCK (High Ambiguity-Driven DOCKing) 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 the critical assessment of prediction of interactions (CAPRI), a blind experiment for the prediction of interactions. Although most docking programs are limited to binary complexes, HADDOCK can deal with multiple molecules (up to six), 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 six 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 multibody 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 was able to generate good to high quality solutions and ranked them at the top, demonstrating its ability to model symmetric multicomponent assemblies. Docking methods can thus play an important role in adding the structural dimension to interactomes. However, although the current docking methodologies were 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, cryoelectron microscopy, etc.) will remain highly desirable to obtain reliable results. *Molecular & Cellular Proteomics* 9:1784–1794, 2010.

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 the 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 make purification and crystallization difficult. Finally, there are some very well studied classes of interactions, such as enzyme-inhibitor, antibody-antigen, and GTPase-GAP (GTPase-activating protein) 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 ob-

From the Bijvoet Center for Biomolecular Research, Science Faculty, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

✂ Author's Choice—Final version full access.

Received, February 2, 2010, and in revised form, March 8, 2010

Published, MCP Papers in Press, March 19, 2010, DOI 10.1074/mcp.M000051-MCP201

tain 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),¹ cryoelectron tomography, and single molecule cryoelectron 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 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), although this method is very sensitive to noise caused by conformational changes upon binding. In the same way, radical probe MS (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).

¹ The abbreviations used are: SAXS, small angle x-ray scattering; AIR, ambiguous interaction restraint; CAPRI, critical assessment of prediction of interactions; CNS, Crystallography & NMR System; CPORT, consensus prediction of interface residues in transient complexes; EM, electron microscopy; E2A, glucose-specific enzyme IIA; HADDOCK, high ambiguity-driven docking; HPr, histidine-containing phosphocarrier protein; IM, ion mobility separation; NCS, non-crystallographic symmetry; PPI, protein-protein interactions; r.m.s.d., root mean square deviation; RP, radical probe; i-r.m.s.d., interface root mean square deviation; l-r.m.s.d., ligand r.m.s.d.; fnat, fraction of native contacts.

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 seen in the recent rounds of the critical assessment of prediction of interactions (CAPRI) experiment (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.

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 I. Interface residues are defined as “*active residues*” that are believed to participate in the formation of the interface, and “*passive residues*” are those that are possibly at the interface; other kinds of data can be entered directly. (See the original HADDOCK studies (26–28) and “Materials and Methods” for more details.) HADDOCK has performed very well in translating these data into structures and structural models. More than 60 Protein Data Bank structures calculated using HADDOCK have been deposited to date as experimental structures in the Protein Data Bank (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 (Protein Data Bank code 1J80 (30)) for which RP-MS data are available, PROXIMO was able to closely reproduce the crystal structure (root mean square deviation (r.m.s.d.) of the top scoring model from the reference crystal structure is 1.26 Å); using the same data, HADDOCK could get even closer with an r.m.s.d. of only 0.68 Å from the crystal structure (results not shown).

TABLE I
Various experimental data that can be incorporated into HADDOCK

Experimental data	HADDOCK representation
Mutagenesis data	Active and passive residues
Hydrogen/deuterium exchange data	Active and passive residues
Bioinformatics interface predictions	Active and passive residues
Mass spectrometry data	
Cross-linking data	Custom CNS restraints
Radical probe mass spectrometry	Active and passive residues
Limited proteolysis mass spectrometry	Active and passive residues or directly as an MTMDAT-generated HADDOCK parameter file
NMR data	
Chemical shift perturbation data	Active and passive residues
Cross-saturation experiments	Active and passive residues
Residual dipolar couplings	Directly
Diffusion anisotropy restraints	Directly
NOEs ^a as custom CNS restraints	Custom CNS restraints
Dihedral angles	Directly
Hydrogen bonds	Directly
Paramagnetic restraints	Under development
Shape data	
SAXS	Under development
EM	Under development

^a Nuclear Overhauser effects.

Most docking methods are designed to deal with just two molecules, making 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, whereas 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 perform 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 was developed by the team of SymmDock, can build hetero-oligomer complexes, but it does not 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 multibody docking of complexes. Like the HADDOCK program itself, the server supports the docking of up to six 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 multibody server on a small benchmark comprising complexes of various symmetries and increasing numbers of components (from three to five). To drive the docking, bioinformatics interface predictions and/or available experimental information were used. The HADDOCK server is available on line. <http://haddock.chem.uu.nl>.

MATERIALS AND METHODS

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 neighbors. 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 (Fig. 1). The effective distance between all those atoms, d_{iAB}^{eff} , is calculated as follows.

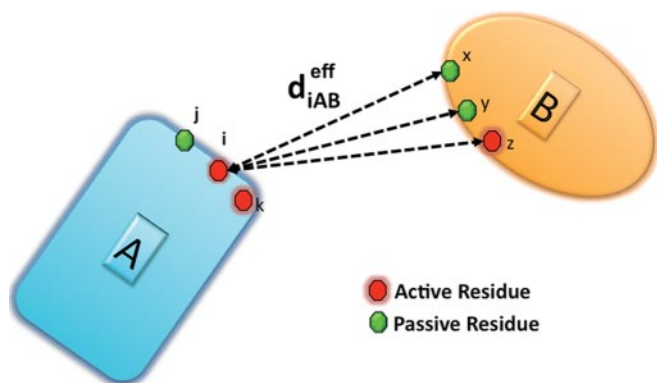


FIG. 1. Illustration of AIRs used in HADDOCK to drive docking. Active residues correspond to residues experimentally identified or predicted to be at the interface. Passive residues are surface neighbors of active residues. AIRs are defined for each active residue with the effective distance being calculated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

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

Here N_{Atom} indicates all atoms of the source residue on molecule A, N_{resB} indicates the residues defined to be at the interface of the target molecule B, and N_{Atom} indicates 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) semiflexible refinement in torsion angle space (it1), and (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.

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

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

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

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 7.5-Å cutoff based on their pairwise r.m.s.d. values, and the cluster ranks

TABLE II
Definition and illustration of 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		$d(AB) = d(BA)$ $d(AC) = d(CA)$ $d(AD) = d(DA)$ $d(BC) = d(CB)$ $d(BD) = d(DB)$ $d(CD) = d(DC)$

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, or C5) symmetries or any combination thereof. This also allows dealing with dihedral symmetries because dihedral symmetry can be interpreted as a combination of cyclic symmetry pairs (e.g. D2 symmetry is a combination of six C2 symmetry pairs (see Table II)). The symmetry restraints can be applied both within and between molecules. Compared with 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: non-crystallographic symmetry (NCS) (40) and distance symmetry restraints (41, 42), both available within CNS.

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.

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

TABLE III
Properties of multimer docking benchmark

Protein Data Bank code	CATH classification	Complex type	Docking type	Symmetry type	Number of amino acids
1QU9 (66)	Mainly β	Homotrimer	Bound	C3	128
1URZ (67)	Mainly α /mainly β	Homotrimer	Unbound	C3	400
1OUS (68)	$\alpha\beta$	Homotetramer	Bound	D2	114
1VIM (69)	$\alpha\beta$	Homotetramer	Bound	D2	200
1VPN (70)	Mainly β	Homopentamer	Bound	C5	289
3CRO (71)	Mainly α	Homodimer-double-stranded DNA	Unbound	C2	71 (protein), 20 (DNA)

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_α 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_α 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 II. 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, or C5 as shown in Table II) 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 multibody web interface of HADDOCK. The procedure followed to dock the protein-DNA complex differs from the generic multibody docking protocol in the sense that two subsequent docking rounds were performed: 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 modeling rather large deformations in the DNA and is explained comprehensively in a recent work of van Dijk *et al.*²

In four of the test cases (Protein Data Bank codes 1QU9, 1OUS, 1VIM, and 1VPN), the interface information was obtained through the consensus interface prediction server CPORT³ using the “very sensitive” option. In the case of Protein Data Bank code 1URZ, a former CAPRI target, we used the same interface definition as was used

previously in CAPRI (43). For the protein-DNA complex, Protein Data Bank code 3CRO, sequence conservation and experimental data (mutagenesis and ethylation interference) were used to define the protein-DNA interaction site.

The interface information was converted into AIRs via the setup page of the HADDOCK web site. The generated AIR files together with the input structures were then supplied to the multibody server as an input for the docking. To favor compactness of the solution, center of mass restraints were enabled. 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 at their default settings.

Evaluation of Docking Models

The models were evaluated according to the CAPRI criteria (13). For a complex to be classified as acceptable (one star), its interface root mean square deviation (i-r.m.s.d.) from the complex had to be lower than 4 Å, or its ligand r.m.s.d. (l-r.m.s.d.) had to be lower than 10 Å. In addition, the fraction of native contacts (fnat) had to be ≥ 0.1 . For good predictions (two stars), the criteria were i-r.m.s.d. ≤ 2 Å or l-r.m.s.d. ≤ 5 Å and fnat ≥ 0.3 . For high quality predictions (three stars), the criteria were i-r.m.s.d. ≤ 1 Å or l-r.m.s.d. ≤ 1 Å and fnat ≥ 0.5 . A cluster was considered one/two/three star(s) if at 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 III). 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 (1URZ), the starting structures correspond to the dimeric form of the complex, whereas 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 four bound cases and two unbound cases.

For modeling of the benchmark complexes, we made use of the new multibody interface of the HADDOCK web server. The web server provides a user-friendly interface that gives full control over the various HADDOCK parameters and supports a wide range of experimental restraints (Fig. 2).

² M. van Dijk and A. M. J. J. Bonvin (2010). Pushing the limits of what is achievable in protein-DNA docking. Benchmarking HADDOCK's performance. *Nucl. Acid Res.*, in press (2010).

³ S. de Vries and A. M. J. J. Bonvin, submitted manuscript.

Home >>

HADDOCK

Software web portal

Home HADDOCK Whispy 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 Whispy DNA Publications Forum Contact

2008 © NMR Department. All rights reserved. Webdesign by Marc van Dijk (MRC) JCS

FIG. 2. View of multibody web interface of HADDOCK for data-driven docking.

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, or C5) and any type of dihedral symmetry that can be expressed as a combination of the available cyclic symmetry pairs (see “Materials and Methods”). 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.

The performance of our multibody docking approach was demonstrated for six complexes (Table III) using a combination of experimental and/or bioinformatics predictions. For four of the complexes (1QU9, 1OUS, 1VIM, and 1VPN), active and passive residues were defined based on consensus bioinformatics interface predictions from CPORT (see “Materials and Methods”). 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 IV. Using the above information, the HADDOCK multibody 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 (two-star) prediction (see Table V). Furthermore, analysis of the results showed that the imposed symmetries are fulfilled.

In four cases, bound docking was performed, including a trimer (1QU9), two tetramers (1OUS and 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 multicomponent systems, this demonstrates an outstanding performance. In the two unbound cases consisting of a CAPRI target and a protein-DNA complex, good results (two-star quality predictions) were also obtained (see Table V). The performance of the protein-DNA docking (3CRO) and the ability of HADDOCK to catch the conformational changes in the DNA demonstrate that the excellent capabilities of HADDOCK are not limited to just protein-protein complexes. An overlay of the top predictions onto their respective reference crystal structures is shown in Fig. 3.

DISCUSSION

Pushing Back 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 multibody 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 that is able to perform simultaneous docking of multibody complexes up to six components. A multibody docking server was recently released on the HADDOCK web portal, allowing the users, through a user-friendly interface, 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 study against a benchmark set of six multimeric complexes,

TABLE IV
List of active and passive residues used in HADDOCK to dock various benchmark complexes

Protein Data Bank code	Active residues	Passive residues
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

^a The active and passive residue information is gathered via CPORT.

^b The active and passive residue information is gathered via CPORT and literature data.

^c The active and passive residue information is gathered via conservation and experimental data (mutagenesis and ethylation interference).

TABLE V
Multibody docking results obtained via using multibody interface of HADDOCK web server

Protein Data Bank code	Quality/rank	Best structure i-r.m.s.d./l-r.m.s.d. ^a		Best cluster i-r.m.s.d./l-r.m.s.d. ^a	
		Å		Å	
1QU9 ^b	★★★/1	0.8/0.7		★★★/1	
1URZ ^c	★★/1	1.7/5.2		★★/1	
1OUS ^b	★★★/1	0.9/1.2		★★★/1	
1VIM ^b	★★★/1	1.0/1.2		★★★/1	
1VPN ^b	★★★/1	0.7/0.7		★★/1	
3CRO ^c	★★/1	1.79/2.2		★★/1	

^a For the definitions of i-r.m.s.d. and l-r.m.s.d. refer to “Materials and Methods.”

^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 “Materials and Methods” for details).

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 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-r.m.s.d. values for the best model ranging between 0.7 and 2.2 Å. Although we could produce excellent results even with the inclusion of bioinformatics predictions, which usually contain a considerable amount of false positives, one should always keep in mind that the information supplied to HADDOCK should be of high quality. This is because the complexity of the interaction space is larger in the case of multibody docking compared with 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 (IM) coupled to 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

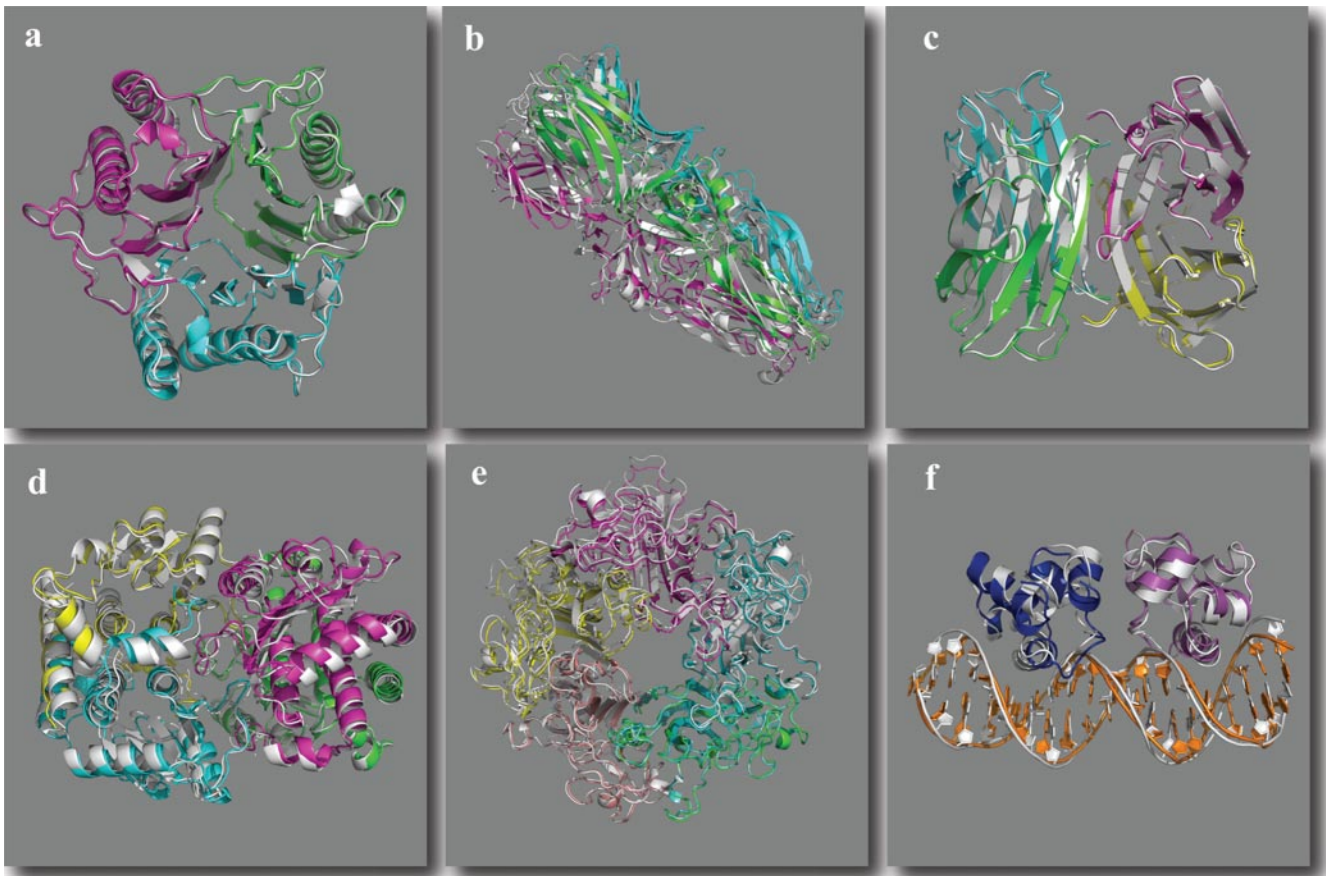


FIG. 3. View of best HADDOCK solutions (having colored monomers) superimposed onto their respective crystal reference structures (shown in light gray). a, 1QU9; b, 1URZ; c, 1OUS; d, 1VIM; e, 1VPN; and f, 3CRO. The figures were generated with Pymol (Delano Scientific LLC).

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 Three-dimensional Structures of Biomolecules?—In today's proteomics era, large scale screening techniques are used to characterize protein-protein interactions (PPIs) *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) or by high throughput mass spectrometry protein complex identification coupled to tandem affinity purification (52), systematic yeast two-hybrid screening (53–55), complementary mapping techniques (e.g. protein fragment complementation assay (56), and *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, proteomics data sets derived to map PPIs, 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 nine of the currently best performing scoring functions against a large set of high quality binding affinity data derived from the 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 PPIs.

Need for Combining Experimental Information and Modeling—By combining a variety of experimental approaches, one can easily increase our knowledge about biologically relevant interaction (64). The experimental information can guide large scale docking studies to upgrade the information contained in interactome maps by adding the three-dimensional structural dimension to the PPIs. Moving toward 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 complexes.⁴ 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 will be developed in the future to open the route to large scale annotation of interactomes.

Acknowledgments—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. We thank Dr. Marc van Dijk (Utrecht University) for providing the data for the protein-DNA complex discussed in this work.

* 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 032220 and FP7 e-Infrastructure “e-NMR” I3 project, Grant 213010).

‡ These authors contributed equally to this work.

§ To whom correspondence should be addressed. Tel.: 31-30-2533859; Fax: 31-30-2537623; E-mail: a.m.j.j.bonvin@uu.nl.

REFERENCES

1. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczy, J., Levine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A.,

Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. K., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubinfeld, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramsay, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kasprzyk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y. J., and Szustakowski, J. (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., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferreria, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M.,

⁴ J. Hennig, S. J. de Vries, K. D. M. Hennig, M. Sunnerghagen, and A. M. J. J. Bonvin, manuscript in preparation.

- Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigó, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejarawal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (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., Förster, 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., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., Punna, T., Peregrín-Alvarez, J. M., Shales, M., Zhang, X., Davey, M., Robinson, M. D., Paccanaro, A., Bray, J. E., Sheung, A., Beattie, B., Richards, D. P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M. M., Vlasblom, J., Wu, S., Orsi, C., Collins, S. R., Chandran, S., Haw, R., Rillstone, J. J., Gandi, K., Thompson, N. J., Musso, G., St Onge, P., Ghanny, S., Lam, M. H., Butland, G., Altaf-Ul, A. M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J. S., Ingles, C. J., Hughes, T. R., Parkinson, J., Gerstein, M., Wodak, S. J., Emili, A., and Greenblatt, J. F. (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., Eyck, L. T., Sternberg, M. J., Vajda, S., Vakser, I., and Wodak, S. J. (2003) CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins* **52**, 2–9
 13. Méndez, R., Lepplae, R., De María, 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., Sharif, 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* **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., Nussinov, R., and Wolfson, H. J. (2003) Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. *Proteins* **52**, 107–112
 22. Grosdidier, S., Pons, C., Solernou, A., and Fernández-Recio, J. (2007) Prediction and scoring of docking poses with pyDock. *Proteins* **69**, 852–858
 23. Cheng, T. M., 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., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, T., and Bonvin, A. M. (2007) HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins* **69**, 726–733
 27. De Vries, S. J., Van Dijk, M., and Bonvin, A. M. (2010) The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **5**, 883–897
 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 Cn or Dn 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. André, 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., and Weng, Z. (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* **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. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921
 39. Fernández-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 225–231
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 138–140
43. van Dijk, A. D., de Vries, S. J., Dominguez, C., Chen, H., Zhou, H. X., and Bonvin, A. M. (2005) Data-driven docking: HADDOCK's adventures in CAPRI. *Proteins* **60**, 232–238
44. Levy, E. D., BoeriErba, E., 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., 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. *Eur. J. Mass Spectrom.* **15**, 113–130
48. Bernstein, S. L., Dupuis, N. F., Lazo, N. D., Wyttenbach, T., Condrón, 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., 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., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreau, M., Muskaf, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A. R., Sassi, H., Nielsen, P. A., Rasmussen, K. J., Andersen, J. R., Johansen, L. E., Hansen, L. H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B. D., Matthiesen, J., Hendrickson, R. C., Gleeson, F., Pawson, T., Moran, M. F., Durocher, D., Mann, M., Hogue, C. W., Figeys, D., and Tyers, M. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183
52. Gavin, A. C., Böschke, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Höfert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M. A., Copley, R. R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G. (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., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamar, G., Yang, M., Johnston, M., Fields, S., and Rothberg, J. M. (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., Braun, P., Yildirim, M. A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., Hao, T., Rual, J. F., Dricot, A., Vazquez, A., Murray, R. R., Simon, C., Tardivo, L., Tam, S., Svrikapa, N., Fan, C., de Smet, A. S., Motyl, A., Hudson, M. E., Park, J., Xin, X., Cusick, M. E., Moore, T., Boone, C., Snyder, M., Roth, F. P., Barabási, A. L., Tavernier, J., Hill, D. E., and Vidal, M. (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., Serna Molina, M. M., Shames, I., Malitskaya, Y., Vogel, J., Bussey, H., and Michnick, S. W. (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., Houfek, T., Mitchell, T., Miller, P., Dean, R. A., Gerstein, M., and Snyder, M. (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 Biol.* **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., Böschke, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dümpelfeld, B., Edelmann, A., Heurtier, M. A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A. M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J. M., Kuster, B., Bork, P., Russell, R. B., and Superti-Furga, G. (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. Kastriitis, P. L., and Bonvin, A. M. J. J. (2010) Are scoring functions in protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. *J. Proteome Res.*, in press
63. Fernandez-Ballester, G., and Serrano, L. (2006) Prediction of protein-protein interaction based on structure. *Methods Mol. Biol.* **340**, 207–234
64. Tong, A. H., Drees, B., Nardelli, G., Bader, G. D., Brannetti, B., Castagnoli, L., Evangelista, M., Ferracuti, S., Nelson, B., Paoluzi, S., Quondam, M., Zucconi, A., Hogue, C. W., Fields, S., Boone, C., and Cesareni, G. (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., 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., Berg-seid, M. G., Buchanan, S. G., Buchanan, M. D., Batiyenko, Y., Christopher, J. A., Emtage, S., Eroshkina, A., Feil, I., Furlong, E. B., Gajiwala, K. S., Gao, X., He, D., Hendle, J., Huber, A., Hoda, K., Kearins, P., Kissinger, C., Laubert, B., Lewis, H. A., Lin, J., Loomis, K., Lorimer, D., Louie, G., Maletic, M., Marsh, C. D., Miller, I., Molinari, J., Muller-Dieckmann, H. J., Newman, J. M., Noland, B. W., Pagarigan, B., Park, F., Peat, T. S., Post, K. W., Radojicic, S., Ramos, A., Romero, R., Rutter, M. E., Sanderson, W. E., Schwinn, K. D., Tresser, J., Winhoven, J., Wright, T. A., Wu, L., Xu, J., and Harris, T. J. (2005) Structural analysis of a set of proteins resulting from a bacterial genomics project. *Proteins* **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. Mondragón, A., and Harrison, S. C. (1991) The phage-434 Cro/Or1 complex at 2.5 Å resolution. *J. Mol. Biol.* **219**, 321–334