This is the pre-peer review version of the following article:

Melquiond, A. S.J., Karaca, E., Kastritis, P. L. and Bonvin, A. M.J.J. (2012). Next challenges in protein–protein docking: from proteome to interactome and beyond. *WIREs Comput Mol Sci*, **2**: 642–651. doi: 10.1002/wcms.91

Which has been published in its full form at: dx.doi.org/doi: 10.1002/wcms.91

Next challenges in protein-protein docking: From proteome to interactome and beyond.

First author: Full name and affiliation; plus email address if corresponding author [Adrien S.J. Melquiond, Bijvoet Center for Biomolecular Research, Faculty of Science, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands]
Second author: Full name and affiliation; plus email address if corresponding author [Ezgi Karaca, Bijvoet Center for Biomolecular Research, Faculty of Science, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands]
Third author: Full name and affiliation; plus email address if corresponding author [Panagiotis L. Kastritis, Bijvoet Center for Biomolecular Research, Faculty of Science, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands]
Fourth author: Full name and affiliation; plus email address if corresponding author [Alexandre M.J.J. Bonvin*, Bijvoet Center for Biomolecular Research, Faculty of Science, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands; a.m.j.j.bonvin@uu.nl]

Keywords

Molecular docking, large-scale, high-throughput docking, macromolecular assemblies, multi-body docking, experimental restraints, binding affinity prediction

Abstract

Advances in biophysics and biochemistry have pushed back the limits for the structural characterization of biomolecular assemblies. Large efforts have been devoted to increase both resolution and accuracy of the methods, probe into the smallest biomolecules as well as the largest macromolecular machineries, unveil transient complexes along with dynamic interaction processes and, lately, dissect whole organism interactomes using high-throughput strategies. The atomic description of such interactions, rarely reached by large-scale project in structural biology, remains however indispensable to fully understand the subtleties of the recognition process, measure the impact of a mutation or predict the effect of a drug binding to a complex. Mixing even a limited amount of experimental and/or bioinformatic data with modeling methods such as macromolecular docking presents a valuable strategy to predict the three-dimensional structures of complexes. Recent developments indicate that the docking community is seething to tackle the greatest challenge of adding the structural dimension to interactomes.

Interactomes are huge, intricate and highly dynamic molecular networks that determine the fate of the cell. They rely on thousands of protein complexes that form the executive machinery underlying biological processes, from DNA replication to protein degradation through metabolism. To catalyze a wide diversity of chemical processes, proteins get in touch with other proteins, nucleic acids, sugars, lipids, and various other molecules. In today's post-genomic era, this versatility has been uncovered by studies that have revealed the number and composition of these macromolecular assemblies while others have illuminated the structural and functional aspects of their interactions. The quest for the elucidation of such networks is usually followed by the thirst of understanding, and lastly by the fantasy of predicting. While systems biology is still coping with the former, molecular docking is already drifting toward large-scale structure prediction.

At the beginning of this century, structural genomics initiatives have opened the door to high-throughput and large-scale three-dimensional (3D) structural determination of biomolecules. Along with the thriving pace of traditional structural biology, this is resulting in tens of thousands of new structures that are likely to feed upcoming projects aiming at adding the structural dimension to interactomes. As a consequence, structural biology is evolving toward systems biology to gain a better understanding of how biomolecular units recognize each other and interact, modulate each other's properties and/or location and work together to fulfill their tasks [1]. Information from either experimentally determined structures or comparative modeling has recently been used to support the structural anatomy of a genome-reduced bacterium [2]; complemented by single-particle electron microscopy and electron tomography, this is providing 3D details on the proteome organization. This outstanding example of the structural mapping of a cell relies on structural

templates to construct models of binary interactions. Unfortunately, there is still a large gap between the number of complexes identified by large-scale proteomics efforts and those for which high-resolution 3D experimental structures are available. At the same time, methods such as cryo-electron microscopy (cryo-EM) or Small angle X-ray scattering (SAXS) are generating a wealth of lower resolution structural information that can be used to complement the prediction efforts. Therefore, the ultimate *tour de force* would be to combine atomic structures or models of individual subunits with low-resolution data and a mixed bag of other experimental or bioinformatics information to add the structural dimension to interactomes. This also means tackling the challenges of predicting large conformational changes potentially occurring upon binding, dealing with heterogeneous multi-component assemblies and predicting binding affinity. These are all required in order to be able to model whole complexes or systems of the utmost importance in biology. The molecular docking community is no longer waiting and is already stepping into those challenges as will be further described in the following.

UNCOVERING PROTEIN-PROTEIN INTERACTIONS

One of the most obvious challenges in the protein interaction world is simply to identify "who interacts with whom". In that regard, mass spectrometry (MS) techniques that determine the composition, stoichiometry and subunit organization of macromolecular complexes is of indispensable value [3]. Determination of whole organism interactomes from high-throughput protein interaction approaches is, however, still non-trivial and far from complete and could, therefore, well be complemented by computational means.

Why predicting interactomes from 3D structures of biomolecules?

High-throughput and large-scale screening techniques are used to characterize protein-protein interactions (PPI's) in vivo. Despite the massive number of interactions detected by these proteomic techniques, interactome coverage remains low, roughly 50% and 10% for the yeast and human interactomes, respectively [4-6]. The proteome coverage is typically limited to about 70% for the best approaches and the inherent fraction of false positives certainly hampers a better coverage of PPI's. Moreover, some methods have difficulties with specific types of interactions, which spur the use of complementary techniques. Finally, PPI proteomic datasets have often a very limited overlap (roughly 20%), even when similar detection methods are used [7].

Beyond their significant contribution to unravel PPI networks, most experimental methods focus on mapping the network topology, i.e. drawing binary interactions between dots. Even if this representation of the "proteins at work" is crucial to characterize the actors of the biological processes, it does not depict the biophysical properties of the underlying biomolecular complexes, such as the dynamics of the interactions, which distinguishes between what is obligate and what is transient, neither does it answer the question whether the interactions occur in cascade or simultaneously… For macromolecular assemblies, for example, PPI networks typically provide little information about who contacts whom within a complex and how mutations might affect function [8]. To start addressing such questions, typically more classical and time-consuming (structural) biology approaches must be followed.

Computational methods to predict protein assemblies may therefore play a role to analyze interactomes, providing additional insights with leverage of the structural models [8]. Approaches using the knowledge of the 3D structure are indeed generally more accurate than those based on sequence only [9]. They are consequently better suited to tackle the issues mentioned above. These "new players" aim at predicting how the proteome is wired and how dynamic changes in the interactome can be induced by environmental factors. However, until methods become available that can reliably estimate binding free energies, predicting interactions from structural models only might remain a hopeless exercise [10].

Binding affinity prediction in protein-protein docking

The ultimate goal in protein-protein docking is to be able to include in the scoring function a thermodynamic description of the system that would allow to describe accurately the affinity or the Gibbs free energy changes of the complex upon binding of the free components. Such a function would actually allow predicting interactomes and/or understanding them, adding the kinetic/thermodynamic dimension to the system under study. It would open the route to structure-based design of protein-protein interactions and their specificity, and would also be relevant for the design of inhibitors of such interactions.

Ideally, a scoring function that can also predict successfully the affinity of any given protein-protein complex should meet the following criteria:

- The scoring function should be based on relevant physicochemical properties or principles of the system that have been determined to play a major role in the energetics of binding [11]
- These physicochemical properties should be combined in a simple and fast model in order to reproduce accurately, within the experimental error, a large number of experimentally determined high-quality binding affinity data
- The algorithm should perform well against other independent datasets
- The algorithm should rank on top the near-native docking poses and should produce reasonably correct binding affinity predictions for these
- Given a set of proteins that are unknown to interact, the algorithm should be able to discriminate binders from non-binders, generating a higher and relevant binding affinity value for the proteins that actually bind.

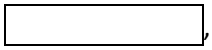
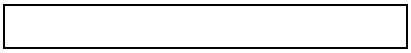
This last point was recently addressed in CAPRI round 21 (see sidebar), where 34 groups were asked to discriminate between native and designed complexes. The results highlighted the clear need for improvements in the field (*Dr. Sarel Fleishman, personal communication*).

The Critical Assessment of PRedicted Interactions (CAPRI) experiment

CAPRI [12] is a "blind" docking experiment in which participants have a limited time to predict the structure of a complex given only the structures, sometimes even only the sequences, of its free constituents. A recent assessment shows that docking algorithms efficiently and accurately predict complexes that undergo small-to-medium size conformational changes [13] while they still encounter major difficulties in case of large conformational changes. Please refer to [14, 15] for

recent reviews about molecular dockings software and [16] for the current challenges and limitations highlighted by the CAPRI experiment.

If all the abovementioned prerequisites can be fulfilled, an accurate binding affinity prediction program should dramatically improve the ranking of the generated docking decoys and might also provide information about the thermodynamics of any macromolecular binders (and non-binders). However, until now, limited efforts have been made towards achieving this goal, basically, due to the absence of a large dataset of experimental binding affinity data. Most of the algorithms that provide an estimate of the binding affinity of the bound partners [17-19] have been parameterized against small datasets of experimental binding affinities [17, 18], or using redundant training/scoring sets [19]. Their general applicability is therefore limited. They are also often restricted to systems that theoretically bind as rigid bodies, neglecting any conformational change that could take place upon binding. Note that such algorithms, even if they have been parameterized and validated, have not yet been incorporated into protein-protein docking to test whether affinity predictors are valuable in ranking the list of decoys generated by the search algorithm.

One large-scale benchmark of protein-protein binding affinities has recently been compiled [20], incorporating experimental data for 144 protein-protein complexes for which the 3D structures of both the free components and their complex are known. Testing a number of widely used scoring functions against a previous, more limited, version of this dataset clearly highlighted the need for improvement in scoring functions used in docking in order for them to correlate with affinity [21]. None of the current scoring functions was able to reproduce accurately the experimentally determined binding affinity data although some weak correlations were observed. Note that scoring functions have often not been designed to deal with such problems, since scoring in docking and binding affinity prediction are two different problems. Various reasons could explain these results [21]: quality of the experimental data, conformational changes taking place upon binding, co-factors that might be needed for binding, allosteric regulation, effect of the solvent, etc. However, a main reason could well be that the current scoring functions do not account for the underlying energetics of the free components. Figure 1 illustrates this point: While calculating the energetics of the docking poses, classical scoring functions will incorrectly predict that  , although the corresponding free energies of binding  are different because of the differences in the free state. The latter are typically neglected in docking, which can result in poor correlations if any with binding affinity, even if scoring functions do show a strong performance in ranking and selecting high quality models in the CAPRI competition [13]. An ideal scoring function that could also predict binding affinity should, in principle, be able to (indirectly or directly) account for the free energy of the unbound partners. Determining the energy landscape of the unbound partners is, however, still very difficult to achieve experimentally [22-24].

The recently published protein-protein binding affinity benchmark should act as a catalyzer for the development and improvement of binding affinity prediction algorithms and, hopefully, scoring and binding affinity prediction will start to converge in a not too far future.

Unraveling the specificity of interactions in a combinatorial system: the ubiquitome

Predicting the binding affinity of a possible complex based on the structure of its free components and answering the question “is it biologically relevant?” might be challenging. But things become even tougher when trying to address specificity, considering for example the E2/E3-RING ubiquitome with its ~10000 putative interactions.

The ubiquitination pathway is controlled by an E1 activating enzyme, an E2 conjugating enzyme, and an E3 ligase, which work in a cascade. E3 proteins containing a a RING domain, corresponding to a non-enzymatic recognition factor, accounts for the largest E3 subfamily. E2 and E3 proteins are working in a combinatorial manner to generate different forms of substrate modification, essential for the stability and function of many proteins in eukaryotic cells. However, very little is known about the mechanisms that govern the specificity of their interaction. Recently, two proteomics studies shed light on human E2/E3-RING interactions [25, 26]. The resulting protein interactions charts, combined with point mutations to perturb native E2-E3 interactions and with bioinformatics analyses, have revealed proteins with clear preferences for specific partners while other are very versatile recognizing many partners – the so-called hub proteins [25, 26]. These studies also confirmed the presence of common structural elements that govern the E2/E3-RING recognition. Such a system constitutes a formidable challenge for the field of computational structural biology when aiming at distinguishing the biologically relevant from the biophysically plausible “look-alike” complexes. Predicting the specificity of interactions in such systems will require a deeper knowledge about the mechanisms that underlie functional from non-functional pairing.

SHAPING THE STRUCTURES OF MACROMOLECULAR MACHINES

Current docking techniques perform quite well in the modeling of binary complexes that experience small conformational change upon binding up to about 2Å [27, 28]. This suggests that, with the present state of the docking approaches, we are able to model efficiently and accurately dimeric complexes that follow either a *lock-and-key* or *induced fit* mechanism, typically enzyme-inhibitor complexes. Complexes following the *conformational selection* mechanism can also be modeled provided a suitable ensemble of starting conformations is used for docking. In reality a combination of methods is expected to give the best results as biomolecular recognition typically occurs following a consensus model reconciliating all the above mechanisms [29]. Adding the structural dimension to interactomes will, however, involve the modeling of large and dynamic macromolecular assemblies, potentially involving nucleic acids or small ligands, which constitutes a paramount challenge. The size of the systems might present a first challenge, which might be overcome by using coarse-grained representations and possibly restraining the search for correct poses to relevant regions of the interaction space. Modeling of large macromolecular assemblies will also require being able to deal with multiple molecules, of possibly different natures. Currently there are only few docking programs that can address this challenge [30-37]. Finally, the utmost level of complexity is reached when faced with large conformational changes of biomolecules upon binding, or even binding-induced folding events. Taken together, all these aspects imply a wide, complex and jagged conformational space to be sampled, which becomes impossible with standard docking approaches [38]. To surmount this barrier, additional data and/or new methodologies must be incorporated into the docking process in order to narrow the conformational space and thus focus on the main interest area of the interaction space.

End of the binary era, welcome into multi-body docking!

Our in-house docking software HADDOCK [39] distinguishes itself from others by the use of experimental and/or bioinformatics information to drive the docking process. HADDOCK is not restricted to proteins, but also allows the incorporation of small molecules, sugar, RNA, DNA or peptides into the docking procedure [40, 41] and can deal with up to six molecules simultaneously. So far, it is one of the very few molecular docking programs able to perform direct docking of generic multi-body complexes consisting of up to six components of various natures, with the potential use of cyclic and/or dihedral symmetries [37]. We have recently implemented these features into the HADDOCK web server [41] through the user-friendly Multi-body interface: http://haddock.chem.uu.nl/services/HADDOCK/haddockserver-multi.html. Our server was tested on a benchmark of six complexes: five symmetric protein homo-oligomers, including two unbound cases, and one symmetric protein-DNA complex. Using bioinformatic predictions of the interfaces, native to near-native predictions were obtained for all cases, irrespective of the difficulty, with interface-RMSD values for the best model ranging between 0.7 and 2.2 Å [37]. Considering the current state of the art in the docking field and the difficulty of our benchmark, this sets a milestone for the modeling of macromolecular assemblies.

Another noteworthy example is the Integrative Modeling Platform (IMP) [42] which was designed for the modeling of macromolecular assemblies and, like HADDOCK, combines spatial restraints derived from experimental data and an optimization method that generates structures consistent with these data. It was successfully applied to unravel the architecture of large assemblies (several MDa), such as the Nuclear Pore Complex [43, 44] or the AAA-ATPase ring of the 26S proteasome [45].

Predicting the unpredictable: how to cope with large conformational changes?

Next to modeling macromolecular assemblies, dealing with large conformational changes upon binding still represents one of the major challenges in docking. First of all, we are still not able to predict *a priori* the type and extent of the conformational changes, and, consequently, we cannot choose the best-suited method in order to model these changes. Recently, Dobbins and coworkers revealed by means of Normal Mode Analysis (NMA) that the slowest mode of proteins experiencing significant conformational changes (C_α-RMSD>2.0Å) has a 2.5 times lower frequency than for proteins undergoing limited conformational change (C_α-RMSD<1.0Å) [28]. Therefore, the evaluation of NMA eigenvalues might be deterministic for predicting the amplitude of the conformational changes. The second limitation that hampers the prediction of large conformational changes is that classical docking approaches might not provide accurate conformational sampling. Hybrid methodologies that combine different approaches are therefore needed. Until now, six hybrid methods have been proposed to model large-scale conformational changes: ATTRACT [46], SwarmDock [47], the Multibody multistage docking procedure of MolFit [48], FlexDock [49], the fold-tree representation of Rosetta [50] and, recently, our flexible multi-domain docking approach [51]. ATTRACT can incorporate soft harmonic low frequency modes into the docking procedure. SwarmDock also uses normal modes within their docking algorithm, but, in contrast to ATTRACT, it takes a linear combination of both low and high frequency normal modes into account. MolFit treats the individual domains of the molecules as soft rigid objects and then docks them in a sequential

multi-stage docking protocol. The method of FlexDock follows a similar methodology, dissecting the flexible protein into rigid domains and performing a pairwise docking in the direction of the conformational change predicted by an Elastic Network Model. In Rosetta, the protein is represented as a fold-tree, allowing defining flexible regions between centers of the rigid molecules and thus making the domains moving with respect to each other during rigid-body optimization. All these approaches are quite promising, but have limitations in some cases in treating simultaneously induced backbone and side-chains conformational changes occurring at the interface.

We recently proposed a *Flexible Multi-domain Docking* (FMD) docking approach [51] for dealing with large-domain motions and small- to medium-scale interface changes at the same time, making use for this of the multi-body docking option of HADDOCK. The FMD protocol follows a divide-and-conquer approach: It treats the flexible binding partner as a collection of sub-domains with connectivity restraints between them and uses the multi-body docking ability of HADDOCK to dock the separated domains simultaneously (Figure 2). The FMD procedure allows the modeling of large-scale domain motions at the rigid body docking stage, followed by HADDOCK's regular flexible refinement for dealing with limited induced side-chain and backbone conformational changes. The performance of the FMD protocol was demonstrated on a benchmark of eleven protein-protein complexes, covering a vast range of conformational change from 1.5Å to as much as 19.5Å. The results indicated that FMD outperforms standard two-body docking and could generate at least a top ranking near-native solution for each case (having an interface-RMSD ranging from 1.1 to 4.6Å). This novel docking methodology, FMD, is thus able to model conformational changes as large as 19.5Å (Figure 2), something that has never been shown before [51]. We should note that the FMD protocol was developed assuming the knowledge of the interface regions and that the performance will strongly depend on the quality and amount of the information available to define the interaction regions. It should be quite clear that modeling large and complex macromolecular assemblies will only lead to reliable results provided some information is introduced into the process to limit the interaction space to be searched and thereby the complexity of the problem.

United we stand, divided we fall

All the efforts for addressing the previous challenges will serve the ultimate goal of linking high-resolution structural data to high-throughput protein interaction information. As discussed before, macromolecular assemblies are highly dynamic as their organization and composition may vary in time and space. Furthermore, their precise quaternary structure is not always conserved across species. All these aspects render experimental structural studies of such assemblies very challenging. The result is an information gap between low-resolution information and atomic models of large macromolecular assemblies. To bypass these limitations, the modeling of biologically relevant complexes must incorporate all types of possible experimental information. By combining interaction and shape 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 mutagenesis, chemical cross-linking, fluorescence resonance energy transfer, mass spectrometry of intact complexes, Cryo-EM, SAXS and analytical ultracentrifugation [52].

Docking-based computational approaches have been recently developed for fitting the atomic structures of the components of a macromolecular complex into low-resolution (Cryo-EM density

maps [53], SAXS profiles [54, 55]) representations of the whole assembly, leading to medium- to high-resolution models. These promising approaches rely typically on some kind of geometric algorithm, maximizing the surface overlap between the high-resolution structure of the components and the low-resolution density map on one hand or filtering out rigid docking models that do not agree with the radius of gyration predicted from the SAXS profile of the complex on the other hand. Their overall results show significant improvement of their performance (by a two fold or more) compared to the docking software used alone. They however face some limitations: swap between components, orientation errors, wrong interfaces between molecules, incomplete, and/or inaccurate or missing starting structures of the components. The conformational changes occurring upon binding are, until now, not taken into account and the scoring of the generated models relies mainly on the maximization of the geometric criterion, failing at ranking correctly the energetics of the poses. Finally, most methods focus on one particular type of experimental information, e.g. cryo-EM maps or SAXS profiles, and are yet not able to deal in a versatile way with various sources of information, except for a few like HADDOCK and IMP.

High-Throughput docking, the problem of computational resources

Looking toward large-scale docking, the computational structural biology community must also develop tools for high-throughput predictions in order to tackle large biological systems or even full interactomes. GRID computing might offer a solution, providing shared large computational resources (CPUs, storage, network interfaces, etc.) and a secure protocol for access. HADDOCK is one of the few pioneers in this domain, being integrated into the European e-Infrastructure WeNMR platform (http://www.wenmr.eu). WeNMR gathers researchers from around the world into a Virtual Research Community (VRC) giving them access to more than 16000 CPU spread across Europe and beyond. It represents a growing community in the life science area with more than 270 users to date and, as such, is one of the largest e-infrastructure dedicated to structural biology.

Conclusion

We have reviewed here the future challenges in protein-protein docking, focusing mainly on the recent trend toward large-scale predictions. These include the structural modeling of biomolecular complexes of varying size and complexity, from small peptides to large macromolecular assemblies, likely heterogeneous in nature; the prediction of large conformational changes that might arise upon binding; the integration of diverse experimental information, from low- to high-resolution, into the modeling process; the prediction of the free energy of binding. The outstanding examples of the Nuclear Pore Complex [43, 44] or the AAA-ATPase [45] show that high-resolution structural modeling of macromolecular assemblies comes within reach. The *"one for all, and all for one"* motto of integrated approaches should allow overcoming the limitations of individual methods when applied to large assemblies. By combining various types of information, one can fill the gap in resolution, accuracy and coverage. This, together with the recent and ongoing developments in the molecular docking field discussed here should pave the route for the prediction of challenging macromolecular complexes.

Molecular docking is also stepping into high-throughput structural modeling of complexes as attested by two recent studies [56, 57]. The screening of thousands of generated 3D models of complexes for the yeast interactome (*Saccharomyces cerevisiae*) lead to two main results: the determination of potential new interactions that will seed future experimental studies and the structural observation of concurrency or exclusiveness of multiple interactions that play an important role in some biological processes. These studies together with other on-going and future high-throughput docking experiments are likely to complement experimental mapping of interactomes in a near future, highlighting how molecular docking should be able to add a new layer of knowledge onto PPI's networks.

Acknowledgments

This work was supported by the Netherlands Organization for Scientific Research (VICI grant #700.56.442 to AMJJB) and by the European Community FP7 e-Infrastructure WeNMR project (grant number 261572).

References

1. Beltrao P, Kiel C, Serrano L: Structures in systems biology. Curr Opin Struct Biol 2007, 17:378-384.
2. Kuhner S, van Noort V, Betts MJ, Leo-Macias A, Batisse C, Rode M, Yamada T, Maier T, Bader S, Beltran-Alvarez P, et al: Proteome organization in a genome-reduced bacterium. Science 2009, 326:1235-1240.
3. Synowsky SA, van den Heuvel RH, Mohammed S, Pijnappel PW, Heck AJ: Probing genuine strong interactions and post-translational modifications in the heterogeneous yeast exosome protein complex. Mol Cell Proteomics 2006, 5:1581-1592.
4. Hart GT, Ramani AK, Marcotte EM: How complete are current yeast and human protein-interaction networks? Genome Biology 2006, 7.
5. Gordon PM, Soliman MA, Bose P, Trinh Q, Sensen CW, Riabowol K: Interspecies data mining to predict novel ING-protein interactions in human. BMC Genomics 2008, 9:426.
6. Venkatesan K, Rual JF, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, Hao T, Zenkner M, Xin X, Goh KI, et al: An empirical framework for binary interactome mapping. Nat Methods 2009, 6:83-90.
7. Goll J, Uetz P: The elusive yeast interactome. Genome Biology 2006, 7.
8. Kiel C, Beltrao P, Serrano L: Analyzing protein interaction networks using structural information. Annu Rev Biochem 2008, 77:415-441.
9. Hue M, Riffle M, Vert JP, Noble WS: Large-scale prediction of protein-protein interactions from structures. BMC Bioinformatics 2010, 11:144.
10. Aloy P, Russell RB: Structural systems biology: modelling protein interactions. Nat Rev Mol Cell Biol 2006, 7:188-197.
11. Jelesarov I: Energetics of Protein-Protein Interactions. In Protein-Protein Complexes: Analysis, Modeling and Drug Design. Edited by Zacharias M. London: Imperial College Press; 2010: 46-88

12. Janin J, Henrick K, Moult J, Ten Eyck L, Sternberg MJE, Vajda S, Vasker I, Wodak SJ: CAPRI: A Critical Assessment of PRedicted Interactions. Proteins-Structure Function and Bioinformatics 2003, 52:2-9.
13. Lensink MF, Wodak SJ: Docking and scoring protein interactions: CAPRI 2009. Proteins 2010, 78:3073-3084.
14. Vajda S, Kozakov D: Convergence and combination of methods in protein-protein docking. Curr Opin Struct Biol 2009, 19:164-170.
15. Moreira IS, Fernandes PA, Ramos MJ: Protein-protein docking dealing with the unknown. J Comput Chem 2010, 31:317-342.
16. Pons C, Grosdidier S, Solernou A, Perez-Cano L, Fernandez-Recio J: Present and future challenges and limitations in protein-protein docking. Proteins 2010, 78:95-108.
17. Audie J, Scarlata S: A novel empirical free energy function that explains and predicts protein-protein binding affinities. Biophys Chem 2007, 129:198-211.
18. Su Y, Zhou A, Xia X, Li W, Sun Z: Quantitative prediction of protein-protein binding affinity with a potential of mean force considering volume correction. Protein Sci 2009, 18:2550-2558.
19. Liu S, Zhang C, Zhou H, Zhou Y: A physical reference state unifies the structure-derived potential of mean force for protein folding and binding. Proteins 2004, 56:93-101.
20. Kastritis PL, Moal IH, Hwang H, Weng Z, Bates PA, Bonvin AMJJ, Janin J: A structure-based benchmark for protein-protein binding affinity. Protein Sci 2011.
21. Kastritis PL, Bonvin AMJJ: Are scoring functions in protein-protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. J Proteome Res 2010, 9:2216-2225.
22. Papoian GA: Proteins with weakly funneled energy landscapes challenge the classical structure-function paradigm. Proc Natl Acad Sci U S A 2008, 105:14237-14238.
23. Mello CC, Barrick D: An experimentally determined protein folding energy landscape. Proc Natl Acad Sci U S A 2004, 101:14102-14107.
24. Street TO, Barrick D: Predicting repeat protein folding kinetics from an experimentally determined folding energy landscape. Protein Sci 2009, 18:58-68.
25. Markson G, Kiel C, Hyde R, Brown S, Charalabous P, Bremm A, Semple J, Woodsmith J, Duley S, Salehi-Ashtiani K, et al: Analysis of the human E2 ubiquitin conjugating enzyme protein interaction network. Genome Res 2009, 19:1905-1911.
26. van Wijk SJ, de Vries SJ, Kemmeren P, Huang A, Boelens R, Bonvin AM, Timmers HT: A comprehensive framework of E2-RING E3 interactions of the human ubiquitin-proteasome system. Mol Syst Biol 2009, 5:295.
27. Bonvin AMJJ: Flexible protein-protein docking. Current opinion in structural biology 2006, 16:194-200.
28. Dobbins SE, Lesk VI, Sternberg MJ: Insights into protein flexibility: The relationship between normal modes and conformational change upon protein-protein docking. Proceedings of the National Academy of Sciences of the United States of America 2008, 105:10390-10395.
29. Grünberg R, Nilges M, Leckner J: Flexibility and conformational entropy in protein-protein binding. Structure 2006, 14:683-693.
30. Berchanski A, Eisenstein M: Construction of molecular assemblies via docking: modeling of tetramers with D2 symmetry. Proteins 2003, 53:817-829.
31. Berchanski A, Segal D, Eisenstein M: Modeling oligomers with Cn or Dn symmetry: application to CAPRI target 10. Proteins 2005, 60:202-206.

32. Comeau SR, Camacho CJ: Predicting oligomeric assemblies: N-mers a primer. Journal of Structural Biology 2005, 150:233-244.
33. Inbar Y, Benyamini H, Nussinov R, Wolfson HJ: Prediction of multimolecular assemblies by multiple docking. Journal of Molecular Biology 2005, 349:435-447.
34. Pierce B, Tong WW, Weng ZP: M-ZDOCK: a grid-based approach for C-n symmetric multimer docking. Bioinformatics 2005, 21:1472-1478.
35. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ: Geometry-based flexible and symmetric protein docking. Proteins, Structure Function and Bioinformatics 2005, 60:224-231.
36. Andre I, Bradley P, Wang C, Baker D: Prediction of the structure of symmetrical protein assemblies. Proceedings of the National Academy of Sciences of the United States of America 2007, 104:17656-17661.
37. Karaca E, Melquiond ASJ, de Vries SJ, Kastritis PL, Bonvin AMJJ: Building macromolecular assemblies by information-driven docking: introducing the HADDOCK multi-body docking server. Molecular & cellular proteomics : MCP 2010, 9:1784-1794.
38. Zacharias M: Accounting for conformational changes during protein-protein docking. Current opinion in structural biology 2010, 20:180-186.
39. Dominguez C, Boelens R, Bonvin AMJJ: HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. Journal of the American Chemical Society 2003, 125:1731-1737.
40. de Vries SJ, Melquiond AS, Kastritis PL, Karaca E, Bordogna A, van Dijk M, Rodrigues JP, Bonvin AM: Strengths and weaknesses of data-driven docking in critical assessment of prediction of interactions. Proteins 2010, 78:3242-3249.
41. de Vries SJ, van Dijk M, Bonvin AMJJ: The HADDOCK web server for data-driven biomolecular docking. Nature protocols 2010, 5:883-897.
42. Alber F, Forster F, Korkin D, Topf M, Sali A: Integrating diverse data for structure determination of macromolecular assemblies. Annu Rev Biochem 2008, 77:443-477.
43. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprapto A, Karni-Schmidt O, Williams R, Chait BT, et al: Determining the architectures of macromolecular assemblies. Nature 2007, 450:683-694.
44. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprapto A, Karni-Schmidt O, Williams R, Chait BT, et al: The molecular architecture of the nuclear pore complex. Nature 2007, 450:695-701.
45. Forster F, Lasker K, Nickell S, Sali A, Baumeister W: Toward an integrated structural model of the 26S proteasome. Mol Cell Proteomics 2010, 9:1666-1677.
46. May A, Zacharias M: Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein-protein docking. Proteins 2008, 70:794-809.
47. Moal IH, Bates PA: SwarmDock and the Use of Normal Modes in Protein-Protein Docking. International journal of molecular sciences 2010, 11:3623-3648.
48. Ben-Zeev E, Kowalsman N, Ben-Shimon A, Segal D, Atarot T, Noivirt O, Shay T, Eisenstein M: Docking to single-domain and multiple-domain proteins: old and new challenges. Proteins 2005, 60:195-201.
49. Schneidman-Duhovny D, Nussinov R, Wolfson HJ: Automatic prediction of protein interactions with large scale motion. Proteins 2007, 69:764-773.
50. Wang C, Bradley P, Baker D: Protein-protein docking with backbone flexibility. Journal of Molecular Biology 2007, 373:503-519.

51. Karaca E, Bonvin AMJJ: A multi-domain flexible docking approach to deal with large conformational changes in the modeling of biomolecular complexes. Structure 2011, In press.
52. Melquiond ASJ, Bonvin AMJJ: Data-driven docking: Using external information to spark the biomolecular rendez-vous. In Protein-protein complexes: analysis, modelling and drug design. Edited by Zacharias M. London: Imperial College Press; 2010: 182-208
53. Lasker K, Topf M, Sali A, Wolfson HJ: Inferential optimization for simultaneous fitting of multiple components into a CryoEM map of their assembly. J Mol Biol 2009, 388:180-194.
54. Schneidman-Duhovny D, Hammel M, Sali A: Macromolecular docking restrained by a small angle X-ray scattering profile. J Struct Biol 2010.
55. Pons C, D'Abramo M, Svergun DI, Orozco M, Bernado P, Fernandez-Recio J: Structural characterization of protein-protein complexes by integrating computational docking with small-angle scattering data. J Mol Biol 2010, 403:217-230.
56. Mosca R, Pons C, Fernandez-Recio J, Aloy P: Pushing structural information into the yeast interactome by high-throughput protein docking experiments. PLoS Comput Biol 2009, 5:e1000490.
57. Katebi AR, Kloczkowski A, Jernigan RL: Structural interpretation of protein-protein interaction network. BMC Struct Biol 2010, 10 Suppl 1:S4.
58. Schreuder H, Tardif C, Trump-Kallmeyer S, Soffientini A, Sarubbi E, Akeson A, Bowlin T, Yanofsky S, Barrett RW: A new cytokine-receptor binding mode revealed by the crystal structure of the IL-1 receptor with an antagonist. Nature 1997, 386:194-200.

Figure captions

Figure 1: Schematic representation of the energy landscape of two different protein-protein complexes.

Figure 2: Flexible Multi-domain HADDOCKing illustrated with the interleukin-1 receptor and its antagonist IL1RA, PDBid: 1IRA, experiencing a conformational change of 19.5Å upon binding [58]. The receptor (orange) is cut into domains at its' hinge region. The docking is then performed on the artificially generated multi-body system: two sub-domains of the receptor and the ligand (blue).

Figure 1

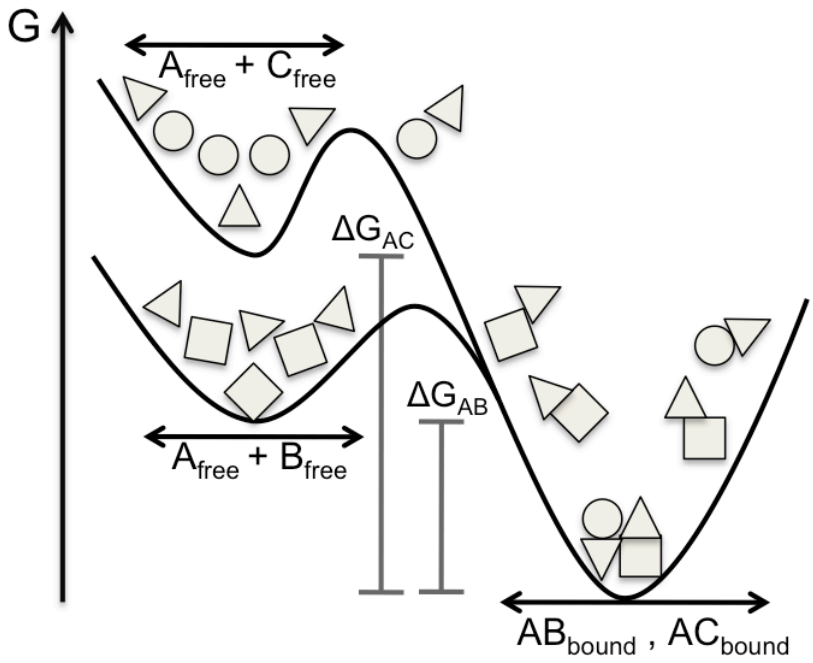
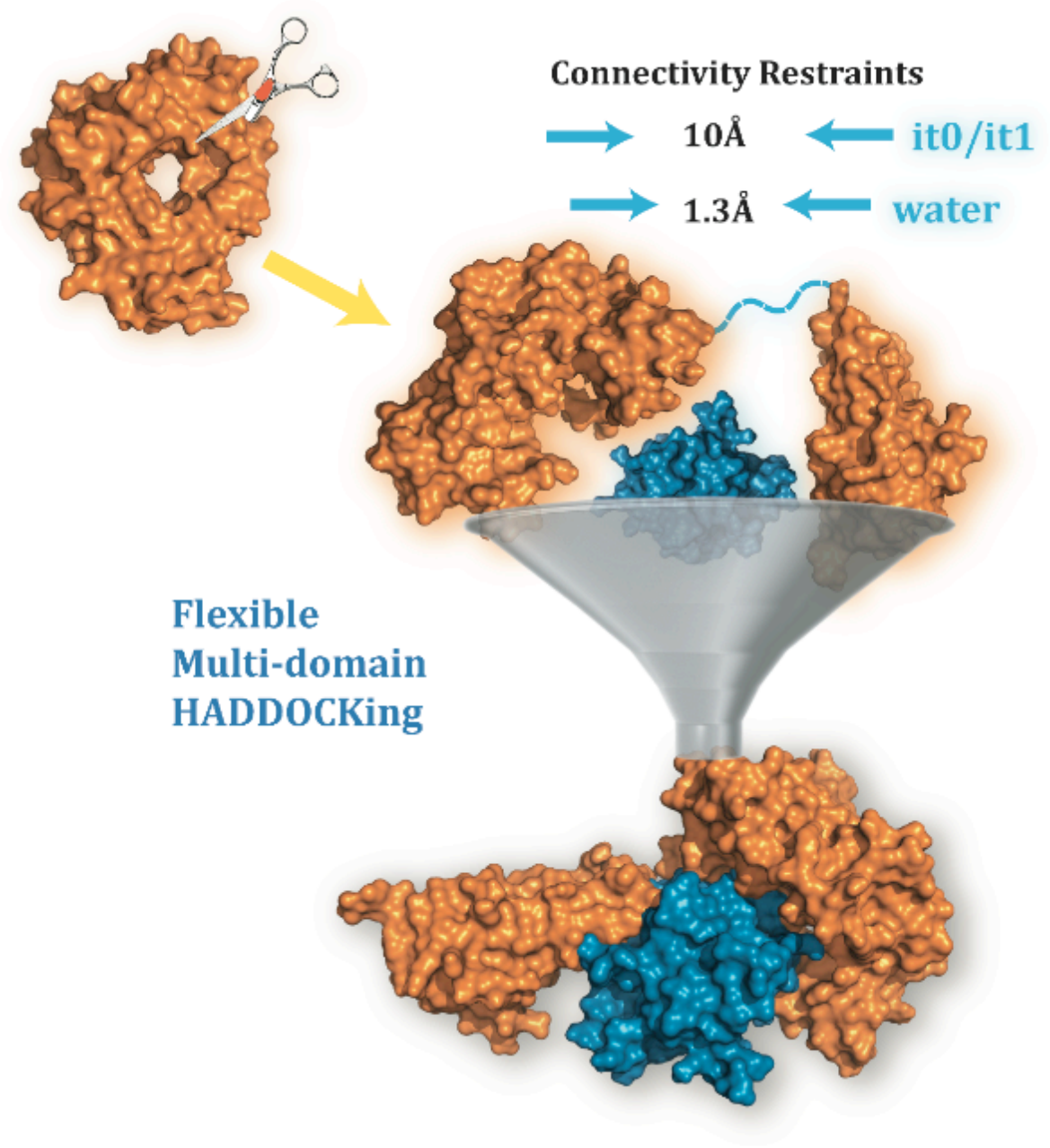

G
A_{free} + C_{free}
ΔG_{AC}
ΔG_{AB}
A_{free} + B_{free}
AB_{bound} , AC_{bound}

Figure 2


Connectivity Restraints
10Å
it0/it1
1.3Å
water
Flexible
Multi-domain
HADDOCKing