Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Structural Biology

Three-dimensional modeling of protein interactions and complexes is going 'omics

Amelie Stein[1,*], Roberto Mosca[1,*] and Patrick Aloy[1,2]

High-throughput interaction discovery initiatives have revealed the existence of hundreds of multiprotein complexes whose functions are regulated through thousands of protein–protein interactions (PPIs). However, the structural details of these interactions, often necessary to understand their function, are only available for a tiny fraction, and the experimental difficulties surrounding complex structure determination make computational modeling techniques paramount. In this manuscript, we critically review some of the most recent developments in the field of structural bioinformatics applied to the modeling of protein interactions and complexes, from large macromolecular machines to domain–domain and peptide-mediated interactions. In particular, we place a special emphasis on those methods that can be applied in a proteome-wide manner, and discuss how they will help in the ultimate objective of building 3D interactome networks.

Addresses
[1] Institute for Research in Biomedicine (IRB Barcelona), Joint IRB-BSC Program in Computational Biology, c/Baldiri i Reixac 10-12, 08028 Barcelona, Spain
[2] Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain

Corresponding author: Aloy, Patrick (patrick.aloy@irbbarcelona.org)
[*] These authors contributed equally to this work.

Current Opinion in Structural Biology 2011, **21**:1–9

This review comes from a themed issue on
Theory and Simulation
Edited by Jeffrey Skolnick and Richard Friesner

0959-440X/$ – see front matter
Published by Elsevier Ltd.

DOI 10.1016/j.sbi.2011.01.005

Introduction

Proteins are key players in virtually all biological events that take place within and between cells. Yet they are social molecules and often accomplish their function as part of large molecular machines, whose action is coordinated through intricate regulatory networks of transient protein–protein interactions. It is thus the inter-relationships between molecules, rather than the individual components, that will ultimately determine the behavior of a biological system. Consequently, large resources have been devoted to unveiling protein inter-relationships in a high-throughput manner in several model organisms, including human [1–3]. However, interaction discovery experiments can only indicate that two proteins interact, but do not reveal the molecular details or the mechanism of binding. This atomic level of detail, often needed to understand function, is only captured in high-resolution three-dimensional (3D) structures, in which individual residue contacts are resolved and the interaction interfaces characterized. Unfortunately, although the number of 3D structures of proteins stored in the Protein Data Bank (PDB) [4] is exponentially increasing, structural data are very scarce in comparison to PPI data, and a considerable fraction of the proteomes of model organisms are not yet covered. For protein interactions the coverage is even lower (Box 1) [5,6], despite the recent implementation of second generation structural genomics (SG) projects working toward solving more structures of interactions and complexes. However, to get the most out of these community efforts, we should couple the outcome of SG pipelines to novel and improved computational modeling techniques so that each new structure solved experimentally can be used to provide details for many more related interactions. Here we review some recent efforts and strategies to model protein interactions and complexes, with a particular emphasis on those that can be applied in a high-throughput manner with the ultimate goal of building 3D interactome networks.

3D modeling of protein interactions and complexes

Classification of interfaces in 3D structures

Since not all computational modeling techniques are applicable to every type of interaction, we first need to divide and classify them. By visually analyzing all PPIs of known 3D structure we identified three main types of interfaces, differing in size, topology, and stability [7] (Figure 1). The first type comprises the interfaces between members of multiprotein complexes, which are often large (>2000 Å^2 [8]) and may have interlocking elements that participate in cooperative effects. In many cases, the individual complex components are not stable and/or functional on their own, but only in interaction with their partners. The two remaining types of interfaces both involve globular domains, which can bind either other domains or short linear peptides. Domain–domain interactions (DDIs) are crucial for stable functional associations of proteins, bury less surface area than complexes (1200–2000 Å^2 [8]) and the interfaces tend to be more flat than those found in complexes. In contrast, peptide-mediated interactions (PMIs) have much smaller interfaces, 350 Å^2 on average [9], and are often

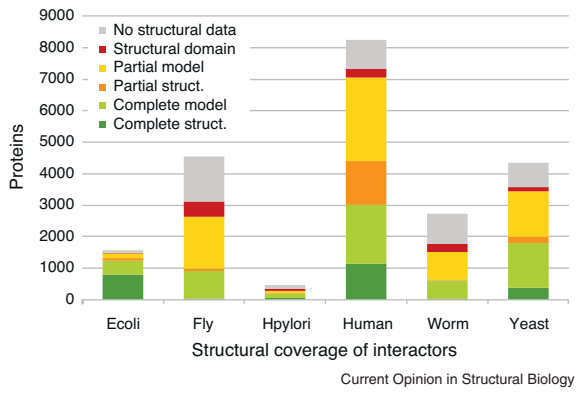
Please cite this article in press as: Stein A, et al. Three-dimensional modeling of protein interactions and complexes is going 'omics, Curr Opin Struct Biol (2011), doi:10.1016/j.sbi.2011.01.005

Box 1 Structural coverage of PPI networks

The structural coverage of PPI networks is high for single proteins but rather limited for interactions. Here we collected experimental PPI data for six model organisms from the major interaction databases [87–92] and verified how many of the entities in the interactomes (both single proteins and interactions) have 3D structures deposited in the PDB. To reduce false positives in the interactomes we considered only interactions that are confirmed by at least two experiments, that belong to the core dataset of a large-scale experiment, or are found in crystallographic structures. For single proteins we considered complete and partial experimental structures from the PDB, complete and partial models from ModBase [93] and representative structures for domain families from Pfam [94] (Chart 1). For interactions we took into account experimental structures and homology templates with >30% sequence identity taken from the biological units in the PDB and structural DDI templates from 3did [35] (Chart 2). While there is structural data for a large part of the interactors (60–80%), the portion of interactions having an experimental structure or a template for comparative modeling is limited (less than 30%).

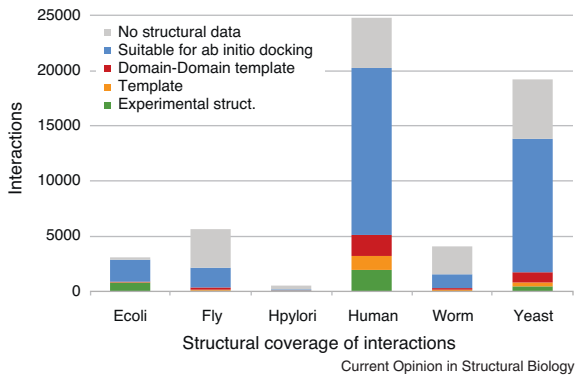
For the interactions lacking both an experimental structure and a template we highlight those that have structural data for both interactors (and are thus suitable for *ab initio* docking methods) and those that cannot be modeled ('no structural data'). *Ab initio* docking can significantly extend the structural coverage of interactomes, particularly in yeast and human.

Chart 1



Structural coverage of interactors.

Chart 2



Structural coverage of interactions.

involved in transient interactions. Below we will discuss recent advances in modeling all three types of interactions.

Protein complexes

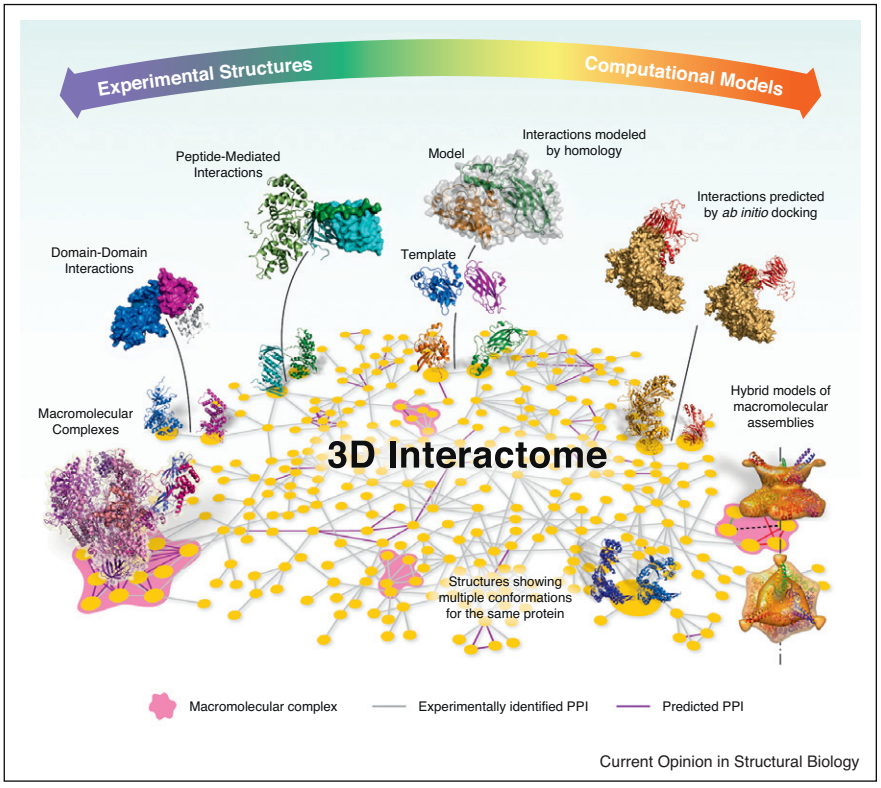
The number of experimental structures of heteromeric complexes deposited in the PDB has seen an almost exponential growth during the last 20 years (Figure 2). While most initial SG efforts were aimed at fully covering the fold space for single proteins, second generation initiatives like SPINE2Complexes (http://www.spine2.eu/SPINE2/) or 3D Repertoire (http://www.3drepertoire.org) mainly focus on solving macromolecular complexes. Nevertheless, and despite several successful stories (e.g. [10,11]), the structural coverage of protein complexes is still very low. As in first generation SG projects, bioinformatics is playing a key role first selecting and prioritizing complex targets [12–14] and then building models of the macromolecular assemblies by combining the atomic structures of the single components with experimental data from different sources (e.g. [15,16]).

It is precisely in the modeling aspects that most exciting developments are taking place. Besides the remarkable accomplishment that represented the structural description of the nuclear pore complex [16], the last two years have seen the creation of models for several large macromolecular structures including the eukaryotic ribosome [17], human RNA polymerase II [18], the AAA-ATPase/20S core particle subcomplex of the 26S proteasome [19], and the histone methyltransferase complex Set1C from yeast [20]. All these works refer to a common emerging methodological paradigm [16,18] that combines experimental interaction data coming from several sources (e.g. mass spectrometry), low and high-resolution structures of individual components of the complex, comparative modeling and *ab initio* protein–protein docking. In particular, structural constraints coming from experimental evidence are often translated into terms of a scoring function which is optimized during the modeling process. Optimization of the scoring function is achieved by placing the structures and models of the single components of the complex in a way that maximizes the number of satisfied constraints.

In the specific field of *ab initio* docking, several multi-protein docking strategies have been introduced to help in the assemblage of large homomeric and heteromeric complexes [21–28], and most of them are also able to take

Please cite this article in press as: Stein A, et al. Three-dimensional modeling of protein interactions and complexes is going 'omics, Curr Opin Struct Biol (2011), doi:10.1016/j.sbi.2011.01.005

Figure 1



Building 3D interactomes. Structural annotation of interactomes requires not only the mapping of structures for the single proteins (and their conformations) but also the interactions between proteins and inside macromolecular complexes. Interactions can be of different types (stable interactions inside complexes, domain–domain interactions and peptide-mediated interactions, see main text) and their structure can be solved experimentally or modeled using computational techniques (homology modeling, *ab initio* docking, or hybrid methods for the reconstruction of complex macromolecular assemblies). Furthermore, many interactions have not been experimentally detected so far, but may be predicted and modeled. The image of the modeled macromolecular assembly is adapted from [19], with kind permission from the authors.

into account symmetry between the individual components. One of the most interesting developments of the last year has certainly been the possibility of performing simultaneous folding and docking of homo-oligomeric complexes. This method is based on the Rosetta framework for molecular modeling and is, to our knowledge, the first method to tackle the problem of coupled folding and docking of proteins in a systematic way, representing a first step toward the reconstruction of those complexes whose structure is influenced by cooperative effects between the subunits [29•].

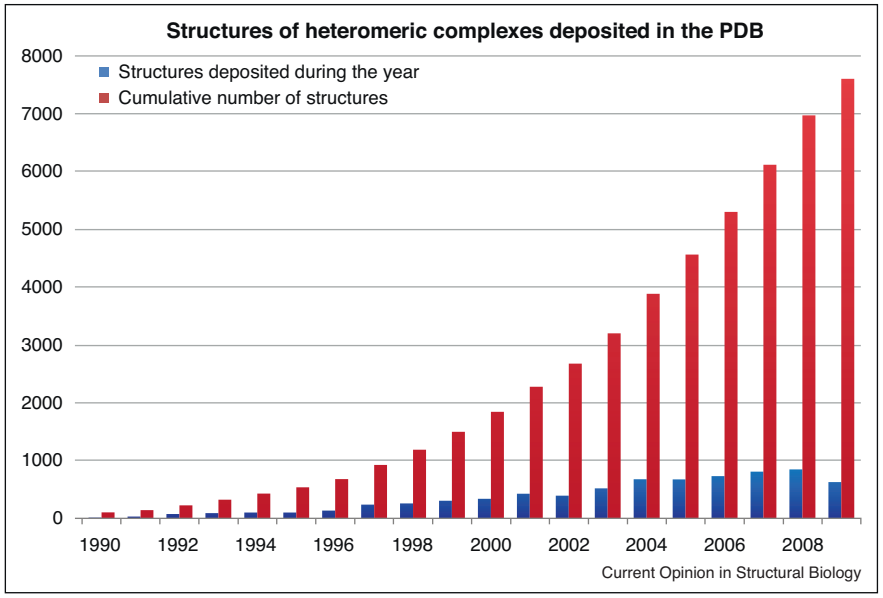
However, the main limitation to scale up these approaches is the necessity of manual collection and curation of the experimental data used as constraints in the complex reconstruction procedure, which has limited their application to a small set of well-studied cases. We hope that the coupling of improved versions of these methods to the outcome of SG initiatives will boost the development of automated pipelines for high-throughput modeling of macromolecular complexes [28].

Domain-based interactions

As outlined above, many protein interactions are mediated by globular binding domains. In domain–domain interactions (DDIs), two globular domains bind to form a stable interface (Figure 1). Because of their stability, DDIs are well amenable to current structure detection techniques; indeed high-resolution 3D structures of over 6000 different pairs of interacting domains are already available in the PDB. These interactions have been collected and catalogued in several databases [30–32] (see [6] for a review), providing a valuable repository of structural templates, including multiple for the same DDI. Recently, studies on this large set of available templates have revealed that some interactions have multiple possible topologies or orientations in which the two domains bind [33,34], although many other domain pairs only show one interaction orientation [35]. In a complementary work, Gao and Skolnick have recently studied local geometries of protein interfaces and observed that ~90% of them have close structural neighbors for the interface, although the monomers are not related [36]. These hidden similarities, only detectable by

Please cite this article in press as: Stein A, et al. Three-dimensional modeling of protein interactions and complexes is going 'omics, Curr Opin Struct Biol (2011), doi:10.1016/j.sbi.2011.01.005

Figure 2



Distribution of the structures of heteromeric complexes deposited in the PDB per year. We considered all those PDB structures with at least two chains of different sequence in the biological unit. Chains with less than 20 residues were not considered. Note that the numbers are not filtered for redundancy.

analyzing 3D structures of the interactions, could also be exploited for the prediction of PPI interfaces. Now, it remains to be analyzed whether all the alternative topologies are functionally relevant, as well as how to select the appropriate template in such cases. Analogously, for modeling PPIs where multiple domains may mediate the interaction, the right domain pair needs to be identified. To address this issue, on the experimental side, there is an increasing tendency for high-throughput interaction discovery screens to set up the gene cloning in a way that they can provide higher-resolution information as, for instance, the specific protein segments that are involved in the interaction [37]. From the computational prediction perspective, this important issue has recently been addressed by Margalit and coworkers, who showed that, although the distribution of domains in the PDB and the proteome are likely to differ, some DDI pairs mediate interactions much more frequently than others [38•], and thus, in the absence of other evidence, these should be taken as the first template option.

Additionally, around 200 globular domains are known to bind short peptide stretches to form domain-motif or peptide-mediated interactions (PMIs) [39,40], and their number increases rapidly [41]. PMIs form much smaller interfaces than DDIs and thus often mediate transient interactions, making their structure determination very challenging. Furthermore, because of the shortness and degeneracy of the motifs, which are usually characterized by only a few key residues [40], they cannot be easily identified automatically. Several collections of PMIs in 3D structures have been released in the last few years, often relying on manual curation efforts [9,42,43]. These now enable detailed analyses of PMIs as well as the development of predictive methods, like the flexible docking and modeling of such interactions recently presented by Raveh *et al.* [44•]. The method allows changes in the peptide backbone, which is crucial because motifs often occur in unstructured protein regions [45] and adapt their conformation upon binding [46•]. Stein and Aloy observed that bound peptides often assume a particular linear and elongated structure, and developed a method to automatically identify 'hidden' PMIs in 3D structures, revealing 32 novel PMIs that were significantly enriched in the interactomes of several model organisms [46•]. Complementarily, a recently developed method by Russell and coworkers does not require structures of the PMI, but searches the surfaces of binding domain structures for regions that could accommodate a given peptide [47•]. While this approach provides less detail of the putative interaction interface, it is applicable to a larger set of recognition domains, namely all for which 3D structures are available independent of peptide ligands. Several PMIs require post-translational modifications (PTMs), such as phosphorylation or methylation, in order to be recognized by their binding domain. This enables the dynamic activation of motif-mediated interactions depending on a particular state of the cell. Furthermore, it allows for switching between different binding partners in the case of neighboring motifs in the same sequence that depend on different states of PTM — if for example one motif is activated by phosphorylation while another motif nearby

Please cite this article in press as: Stein A, et al. Three-dimensional modeling of protein interactions and complexes is going 'omics, Curr Opin Struct Biol (2011), doi:10.1016/j.sbi.2011.01.005

is switched off by phosphorylation at the same residue, then depending on whether the residue is phosphorylated or not different motifs are recognized by their respective interaction partners, thus dynamically rewiring the interaction network (E Akiva, H Margalit, unpublished results). Although this information is still scarce, mass spectrometry-based studies are expected to shortly provide data on which residues are modified in which cellular state, which is especially important for studying the dynamic processes in cellular networks [48,49].

Protein interaction prediction and the challenge of specificity

Although high-throughput interaction discovery experiments have provided tens of thousands of PPIs, the coverage of the interaction space, even for the tackled model organisms, is still quite low [50,51] with many organisms that have not yet been studied or are not amenable to current methods of protein interaction detection. In this context, *in silico* methods to predict PPIs will certainly help to extend the current knowledge of the interaction space. Indeed, many computational methods to predict PPIs, mainly based on transferring the interactions from one organism to another via orthology mapping, have been reported in the last years (see [52,53] for reviews). Structure-based methods to predict interactions between proteins were introduced several years ago [54–56], and they mainly work by modeling or threading the sequences of the two potential interacting proteins onto a structural template and measuring several energy scores on the model, although slight variations have been introduced recently [57,58].

Another possibility that has recently emerged is the use of classical protein–protein docking methods to predict potential interaction partners for a given protein. In a recent study, Valencia and coworkers show that by looking at the entire distribution of a docking scoring function based mainly on surface complementarity they can separate true interactors from a large background of nonredundant potential interactors in a significantly large number of cases [59•]. The authors suggest that 'although protein surface morphology may not be enough to find the native interface it at least contain sufficient information to identify a *bona fide* interactor when a general binding profile is compared to those of many noninteracting proteins'. These results can be put in the context of the theory of encounter complexes [60] and show that a potential exist for the development of new approaches for predicting interaction partners using docking. However, it seems clear that the scoring functions used by several *ab initio* docking methods are not yet sensitive enough to predict experimentally determined binding constants (K_d), even if weak correlations emerge when binding affinities are grouped according to the methods by which they were determined [61]. Furthermore, given the computational complexity of such methods, application to large-scale experiments involving thousands of interactions is, at the moment, only feasible on dedicated high-performance computation facilities [57,62•]. Overall, there are good perspectives ahead, but the adaptation and optimization of docking methods to predict PPIs is certainly a challenge [28].

However, arguably, the most pressing challenge that PPI prediction methods will have to face is related to identifying the subtle changes in the interfaces that determine interaction specificity, a subject that is often neglected during the development and benchmark of the different strategies. Particularly given the modular nature of many PPIs, the question of specificity arises as to how can proteins distinguish their dedicated partners from close homologs, which is particularly important for protein families with many members or widely used binding domains. For DDIs, classical methods to score putative interaction pairs [54] have recently been refined to also consider the conservation of positions within the family. In [63], the authors exploit the observation that, for some families, the conserved residues in the interface are crucial for binding, while less conserved ones are important for specificity, to derive statistical family-specific potentials that yield higher accuracy in specificity determination. In the case of PMIs, most current specificity predictions are family-based, although King and Bradley recently presented a general framework for the prediction of specificity in PMIs [64]. However, the method requires detailed information from the user and is thus currently not applicable for high-throughput studies [64]. Recent advances in PMI specificity prediction benefit from two sides, namely large-scale peptide binding screens [65–67] and structure-based methods for the identification of PMI interfaces [43,46•,47•] as described above. The large-scale phage-display studies provide data on positive as well as negative interactions, which are ideal for training and/or evaluating predictive methods that build on the atomic details of 3D structures. Structural analyses of domain mutations have also shown that positions not directly contacting the peptide nevertheless contribute to binding specificity [66], implying that predictive methods need to move beyond immediate contacts, for example, by considering protein dynamics or by integrating results from atomic interfaces and large-scale screens, as the latter can provide data on indirect effects in the interaction that is not obvious from atomic data and/or to expensive to predict computationally [68,69•]. For instance, Smith and Kortemme incorporated backbone flexibility to improve the prediction of binding preferences of native, mutated and engineered PDZ domains [69•], and Keating and colleagues exploited flexibility in the design of peptides binding the Bcl-2 domain [70], which is crucial in the regulation of apoptosis and thus an interesting cancer target [71]. Competition is another important consideration in interaction partner design. Grigoryan *et al.* recently presented a computational framework that considers both positive and negative

Please cite this article in press as: Stein A, et al. Three-dimensional modeling of protein interactions and complexes is going 'omics, Curr Opin Struct Biol (2011), doi:10.1016/j.sbi.2011.01.005

constraints (desired and undesired binding partners), and experimentally validated their approach [72••]. It is expected that novel computational methods will incorporate these findings, leading to a much more accurate prediction of interaction specificities as well as to the possibility of designing custom binding partners.

Assembling 3D interactomes

To get the most out of interaction data and 3D structures it is fundamental to merge these complementary types of biological data into a systematic view that allows us to move between resolution limits: from the physical wiring of a eukaryotic cell to the molecular bases of each individual interaction. This integration is what we call *3D interactome*, that is, an interactome fully annotated with structural information (Figure 1). However, what was an old dream for most structural and cell biologists seems much more feasible now and, for some specific cases, will soon become a reality. Indeed, the structural mapping of the central metabolic network of *Thermotoga maritima* has recently been presented [73••]. Although simpler than most protein interactome networks, this is a perfect example of how the inclusion of structural details into large complex systems can provide valuable information on, for instance, the evolutionary mechanisms that have most likely shaped the current versions of metabolic networks. In addition, from a more functional perspective, having the structural picture has also helped to confirm or identify novel substrates for a significant number of component proteins in the metabolic network.

In more complex PPI networks, the inclusion of structural details has also been key to gain insights on the role of protein hubs in network evolution [74] and to clarify the role of disorder in interaction networks [75]. Using the same rationale, Nussinov and coworkers built the p53 pathway along with structure-based predictions of which interactions in the pathway can and cannot coexist [76]. The possibility of applying and extending these methods will largely depend on the availability of structural data for the experimentally identified PPIs. Most of what is known, or can be predicted, at structural level for PPIs has been recently collected by Vakser and coworkers in a database that contains homology models for over 25,000 interactions in 771 organisms spanning the entire tree of life, from viruses to human [77]. Furthermore, in the lack of a structural template for comparative modeling, which is the case for more than 70% of the interactomes of several model organisms (Box 1), one can still count on the observation that in a large number of cases structural data are available for the single interactors, making it possible to apply *ab initio* docking methods as it has been done for yeast [62•].

However, 3D interactomes are likely to have their major impact on the quest for therapeutics, since PPIs have emerged as novel and attractive pharmacological targets. Targeting interactions has certain advantages over more traditional targets since it may offer a more subtle, specific form of regulation that can avoid side-effects owing to off-target effects, or to total ablation of normal enzyme activity [78]. During the last year, genetic methods have been developed that allow testing edge-specific or 'edgetic' effects, indicating that it is important to differentiate the individual interactions a protein is involved in, which is not possible by simply removing nodes from the network [79,80]. Performing such studies on the structural level, for example, by mapping results from edgetic perturbations onto 3D interfaces, or by analyzing 3D interactomes for residues involved in multiple interfaces, could pinpoint positions responsible for particular interactions but not relevant for others, such that mutations in those regions would specifically affect individual interactions. This could greatly increase our understanding of mutation-based diseases and might even allow for the design of small molecules to interfere with particular interactions of a protein while leaving others intact [81]. PMIs are considered easier to target than the larger DDI or complex interfaces because of their relatively small interaction interface [81,82]. However, not many structures of PPI interfaces with drug-like ligands are available [83], making the identification of targetable sites challenging. Davis and Sali recently presented a method that uses homology transfer to identify sites on protein surfaces at which both small molecules and other proteins bind [84•]. This consolidation of data from multiple structures may allow identification of interesting sites for drug targeting as well as small molecules to base drug design on. Alternatively, peptides derived from interaction interfaces with native binding partners may serve as templates for inhibitors [85,86]. Such interaction-based techniques will allow the development of novel drugs tailored to act on specific protein interactions and thus redirect the flow of information in PPI networks.

Concluding remarks

Protein complexes and interaction networks represent the cornerstone of much of modern biology, and are crucial to bridge the gaps between molecule-based disciplines and systemic approaches. However, the level of detail required is often beyond the resolution of most 'omics techniques, and can only be provided through the knowledge of the corresponding 3D structures. Bottlenecks surrounding structure determination have highlighted the need for new computational methods able to complement experimental pipelines, and the last years have seen the first steps toward the exciting quest of populating interactome networks with high-resolution structural data. The goal is indeed daunting but, if we succeed in our attempts, the creation of 3D interactomes may provide crucial insights for the understanding of the complex genome-to-phenome relationship.

Please cite this article in press as: Stein A, et al. Three-dimensional modeling of protein interactions and complexes is going 'omics, Curr Opin Struct Biol (2011), doi:10.1016/j.sbi.2011.01.005

Acknowledgements

We would like to thank Marina Mapelli (IEO Milan) and Juan Fernandez-Recio (BSC) for the helpful suggestions. This work was partially supported by the grants BIO2010-22073 and PSS-010000-2009 from the Spanish Ministerio de Educación y Ciencia and the European Commission under FP7 Grant Agreement 223101(AntiPathoGN).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Rual J-F, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N *et al.*: **Towards a proteome-scale map of the human protein–protein interaction network**. *Nature* 2005, **437**: 1173-1178.

2. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S *et al.*: **A human protein–protein interaction network: a resource for annotating the proteome**. *Cell* 2005, **122**:957-968.

3. Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, McBroom-Cerajewski L, Robinson MD, O'Connor L, Li M *et al.*: **Large-scale mapping of human protein–protein interactions by mass spectrometry**. *Mol Syst Biol* 2007, **3**:89.

4. Berman H, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov I, Bourne P: **The Protein Data Bank**. *Nucleic Acids Res* 2000, **28**:235-242.

5. Schuster-Böckler B, Bateman A: **Reuse of structural domain–domain interactions in protein networks**. *BMC Bioinformatics* 2007, **8**:259.

6. Prieto C, De Las Rivas J: **Structural domain–domain interactions: assessment and comparison with protein–protein interaction data to improve the interactome**. *Proteins* 2010, **78**:109-117.

7. Aloy P, Russell RB: **Structural systems biology: modelling protein interactions**. *Nat Rev Mol Cell Biol* 2006, **7**: 188-197.

8. Chakrabarti P, Janin J: **Dissecting protein–protein recognition sites**. *Proteins* 2002, **47**:334-343.

9. Stein A, Aloy P: **Contextual specificity in peptide-mediated protein interactions**. *PLoS ONE* 2008, **3**:e2524.

10. Chandran V, Fronzes R, Duquerroy S, Cronin N, Navaza J, Waksman G: **Structure of the outer membrane complex of a type IV secretion system**. *Nature* 2009, **462**:1011-1015.

11. Makde RD, England JR, Yennawar HP, Tan S: **Structure of RCC1 chromatin factor bound to the nucleosome core particle**. *Nature* 2010, **467**:562-566.

12. Bravo J, Aloy P: **Target selection for complex structural genomics**. *Curr Opin Struct Biol* 2006, **16**:385-392.

13. Pache RA, Aloy P: **Incorporating high-throughput proteomics experiments into structural biology pipelines: identification of the low-hanging fruits**. *Proteomics* 2008, **8**:1959-1964.

14. Brooks MA, Gewartowski K, Mitsiki E, Létoquart J, Pache RA, Billier Y, Bertero M, Corréa M, Czarnocki-Cieciura M, Dadlez M *et al.*: **Systematic bioinformatics and experimental validation of yeast complexes reduces the rate of attrition during structural investigations**. *Structure* 2010, **18**:1075-1082.

15. Aloy P, Böttcher B, Ceulemans H, Leutwein C, Mellwig C, Fischer S, Gavin A-C, Bork P, Superti-Furga G, Serrano L *et al.*: **Structure-based assembly of protein complexes in yeast**. *Science* 2004, **303**:2026-2029.

16. 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.

17. Taylor DJ, Devkota B, Huang AD, Topf M, Narayanan E, Sali A, Harvey SC, Frank J: **Comprehensive molecular structure of the eukaryotic ribosome**. *Structure* 2009, **17**:1591-1604.

18. Lasker K, Phillips JL, Russel D, Velazquez-Muriel J, Schneidman-Duhovny D, Tjioe E, Webb B, Schlessinger A, Sali A: **Integrative structure modeling of macromolecular assemblies from proteomics data**. *Mol Cell Proteomics* 2010, **9**:1689-1702.

19. Förster F, Lasker K, Beck F, Nickell S, Sali A, Baumeister W: **An atomic model AAA-ATPase/20S core particle sub-complex of the 26S proteasome**. *Biochem Biophys Res Commun* 2009, **388**:228-233.

20. Tuukkanen A, Huang B, Henschel A, Stewart F, Schroeder M: **Structural modeling of histone methyltransferase complex Set1C from *Saccharomyces cerevisiae* using constraint-based docking**. *Proteomics* 2010, **10**:4186-4195.

21. Berchanski A, Shapira B, Eisenstein M: **Hydrophobic complementarity in protein–protein docking**. *Proteins* 2004, **56**:130-142.

22. Berchanski A, Segal D, Eisenstein M: **Modeling oligomers with Cn or Dn symmetry: application to CAPRI target 10**. *Proteins* 2005, **60**:202-206.

23. Comeau SR, Camacho CJ: **Predicting oligomeric assemblies: N-mers a primer**. *J Struct Biol* 2005, **150**:233-244.

24. André I, Bradley P, Wang C, Baker D: **Prediction of the structure of symmetrical protein assemblies**. *Proc Natl Acad Sci U S A* 2007, **104**:17656-17661.

25. Pierce B, Tong W, Weng Z: **M-ZDOCK: a grid-based approach for Cn symmetric multimer docking**. *Bioinformatics* 2005, **21**:1472-1478.

26. Inbar Y, Benyamini H, Nussinov R, Wolfson HJ: **Prediction of multimolecular assemblies by multiple docking**. *J Mol Biol* 2005, **349**:435-447.

27. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ: **Geometry-based flexible and symmetric protein docking**. *Proteins* 2005, **60**:224-231.

28. Karaca E, Melquiond ASJ, de Vries SJ, Kastritis PL, Bonvin AMJJ: **Building macromolecular assemblies by information-driven docking: introducing the HADDOCK multibody docking server**. *Mol Cell Proteomics* 2010, **9**:1784-1794.

29. • Das R, André I, Shen Y, Wu Y, Lemak A, Bansal S, Arrowsmith CH, Szyperski T, Baker D: **Simultaneous prediction of protein folding and docking at high resolution**. *Proc Natl Acad Sci U S A* 2009, **106**:18978-18983.

This paper presents a method for the simultaneous folding and docking of multiple-chain interleaved homo-oligomers, which are very common across all organisms but challenging to model. For 11 out of 27 test cases a backbone accuracy of better than 3 Å with respect to the native oligomer could be achieved. Furthermore the authors demonstrate that the models produced by the method can be used for crystallographic phasing, or in combination with NMR data to rapidly determine the structures of homo-oligomers.

30. Stein A, Russell RB, Aloy P: **3did: interacting protein domains of known three-dimensional structure**. *Nucleic Acids Res* 2005, **33**:D413-D417.

31. Finn RD, Marshall M, Bateman A: **iPfam: visualization of protein–protein interactions in PDB at domain and amino acid resolutions**. *Bioinformatics* 2005, **21**:410-412.

32. Winter C, Henschel A, Kim WK, Schroeder M: **SCOPPI: a structural classification of protein–protein interfaces**. *Nucleic Acids Res* 2006, **34**:D310-D314.

33. Kim WK, Henschel A, Winter C, Schroeder M: **The many faces of protein–protein interactions: a compendium of interface geometry**. *PLoS Comput Biol* 2006, **2**:e124.

34. Teyra J, Paszkowski-Rogacz M, Anders G, Pisabarro MT: **SCOWLP classification: structural comparison and analysis of protein binding regions**. *BMC Bioinformatics* 2008, **9**:9.

35. Stein A, Céol A, Aloy P: **3did: identification and classification of domain-based interactions of known three-dimensional structure**. *Nucleic Acids Res* 2011 doi: 10.1093/nar/gkq1962.

Please cite this article in press as: Stein A, et al. Three-dimensional modeling of protein interactions and complexes is going 'omics, Curr Opin Struct Biol (2011), doi:10.1016/j.sbi.2011.01.005

36. Gao M, Skolnick J: **Structural space of protein–protein interfaces is degenerate, close to complete, and highly connected**. *Proc Natl Acad Sci U S A* 2010, **107**:22517-22522.

37. Boxem M, Maliga Z, Klitgord N, Li N, Lemmens I, Mana M, de Lichtervelde L, Mul JD, van de Peut D, Devos M *et al.*: **A protein domain-based interactome network for *C. elegans* early embryogenesis**. *Cell* 2008, **134**:534-545.

38. • Itzhaki Z, Akiva E, Margalit H: **Preferential use of protein domain-pairs as interaction mediators: order and transitivity**. *Bioinformatics* 2010, **26**:2564-2570.
In this paper, the authors analyze the structures of domain-mediated interactions and determine which domain pairs are preferentially used in interactions and which are rarely used. In particular, they find that homotypic DDIs are much more common than heterotypic interactions. The preference relations among domain pairs are described through a network that can be used to predict interaction-mediating domains.

39. Ceol A, Chatr-aryamontri A, Santonico E, Sacco R, Castagnoli L, Cesareni G: **DOMINO: a database of domain–peptide interactions**. *Nucleic Acids Res* 2007, **35**:D557-D560.

40. Gould CM, Diella F, Via A, Puntervoll P, Gemünd C, Chabanis-Davidson S, Michael S, Sayadi A, Bryne JC, Chica C *et al.*: **ELM: the status of the 2010 eukaryotic linear motif resource**. *Nucleic Acids Res* 2010, **38**:D167-D180.

41. Petsalaki E, Russell RB: **Peptide-mediated interactions in biological systems: new discoveries and applications**. *Curr Opin Biotechnol* 2008, **19**:344-350.

42. Encinar JA, Fernandez-Ballester G, Sánchez IE, Hurtado-Gomez E, Stricher F, Beltrao P, Serrano L: **ADAN: a database for prediction of protein–protein interaction of modular domains mediated by linear motifs**. *Bioinformatics* 2009, **25**:2418-2424.

43. London N, Movshovitz-Attias D, Schueler-Furman O: **The structural basis of peptide–protein binding strategies**. *Structure* 2010, **18**:188-199.

44. • Raveh B, London N, Schueler-Furman O: **Sub-angstrom modeling of complexes between flexible peptides and globular proteins**. *Proteins* 2010, **78**:2029-2040.
The authors present a method for docking and refining PMIs featuring backbone relaxation for the peptide to account for conformational changes upon binding. The method predicts high-scoring near-native models for almost all of the benchmark PMIs when starting from bound structures, and for 59% of the cases starting from apo domain structures.

45. Fuxreiter M, Tompa P, Simon I: **Local structural disorder imparts plasticity on linear motifs**. *Bioinformatics* 2007, **23**:950-956.

46. • Stein A, Aloy P: **Novel peptide-mediated interactions derived from high-resolution 3-dimensional structures**. *PLoS Comput Biol* 2010, **6**:e1000789.
On the basis of the observed linear and elongated structures of peptides in PMIs, the authors have developed a method for the pattern-independent recognition of such interactions in 3D structures. Around 50% of the derived consensus motifs show significant enrichment in current model organism interactomes, further supporting their functional relevance.

47. • Petsalaki E, Stark A, García-Urdiales E, Russell RB: **Accurate prediction of peptide binding sites on protein surfaces**. *PLoS Comput Biol* 2009, **5**:e1000335.
This paper presents a method that predicts peptide binding sites on the surfaces of globular domains, given a regular expression for the peptide. Due to the fact that structures of domains are much more frequently available than those of PMIs, this method may predict interface predictions for many interactions that are currently not directly available in 3D structures.

48. Alexander RP, Kim PM, Emonet T, Gerstein MB: **Understanding modularity in molecular networks requires dynamics**. *Sci Signal* 2009, **2**:pe44.

49. Heck AJ: **Native mass spectrometry: a bridge between interactomics and structural biology**. *Nat Methods* 2008, **5**:927-933.

50. Huang H, Jedynak BM, Bader JS: **Where have all the interactions gone? Estimating the coverage of two-hybrid protein interaction maps**. *PLoS Comput Biol* 2007, **3**:e214.

51. Stumpf MPH, Thorne T, de Silva E, Stewart R, An HJ, Lappe M, Wiuf C: **Estimating the size of the human interactome**. *Proc Natl Acad Sci U S A* 2008, **105**:6959-6964.

52. Shoemaker BA, Panchenko AR: **Deciphering protein–protein interactions. Part II. Computational methods to predict protein and domain interaction partners**. *PLoS Comput Biol* 2007, **3**:e43.

53. Lewis ACF, Saeed R, Deane CM: **Predicting protein–protein interactions in the context of protein evolution**. *Mol Biosyst* 2010, **6**:55-64.

54. Aloy P, Russell RB: **Interrogating protein interaction networks through structural biology**. *Proc Natl Acad Sci U S A* 2002, **99**:5896-5901.

55. Aloy P, Ceulemans H, Stark A, Russell RB: **The relationship between sequence and interaction divergence in proteins**. *J Mol Biol* 2003, **332**:989-998.

56. Lu L, Lu H, Skolnick J: **MULTIPROSPECTOR: an algorithm for the prediction of protein–protein interactions by multimeric threading**. *Proteins* 2002, **49**:350-364.

57. Hue M, Riffle M, Vert J-P, Noble WS: **Large-scale prediction of protein–protein interactions from structures**. *BMC Bioinformatics* 2010, **11**:144.

58. Singh R, Park D, Xu J, Hosur R, Berger B: **Struct2Net: a web service to predict protein–protein interactions using a structure-based approach**. *Nucleic Acids Res* 2010, **38**:W508-W515.

59. • Wass MN, Fuentes G, Pons C, Pazos F, Valencia A: **Towards the prediction of protein interaction partners using physical docking.** *Mol Syst Biol* 2010, **7**:469. doi:10.1038/msb.2011.3.
This paper presents a large scale docking experiment of 56 proteins against a library of 922 randomly chosen protein structures including one real interactor. The authors show that by using a rigid body docking program with a scoring function based mainly on surface complementarity the score distribution of the real interactor can be distinguished from the distributions of the background dataset in a surprisingly high number of cases.

60. Blundell TL, Fernandez-Recio J: **Cell biology: brief encounters bolster contacts**. *Nature* 2006, **444**:279-280.

61. 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.

62. • Mosca R, Pons C, Fernández-Recio J, Aloy P: **Pushing structural information into the yeast interactome by high-throughput protein docking experiments**. *PLoS Comput Biol* 2009, **5**:e1000490.
This paper presents the first large-scale *ab initio* protein–protein docking study on the yeast interactome. Rigid body docking methods are first benchmarked and then applied to more than 3000 protein pairs. A threshold on the score is used to identify cases of higher confidence.

63. Panjkovich A, Aloy P: **Predicting protein–protein interaction specificity through the integration of three-dimensional structural information and the evolutionary record of protein domains**. *Mol Biosyst* 2010, **6**:741-749.

64. King CA, Bradley P: **Structure-based prediction of protein–peptide specificity in rosetta**. *Proteins* 2010, **78**:3437-3449.

65. Stiffler MA, Chen JR, Grantcharova VP, Lei Y, Fuchs D, Allen JE, Zaslavskaia LA, MacBeath G: **PDZ domain binding selectivity is optimized across the mouse proteome**. *Science* 2007, **317**:364-369.

66. Tonikian R, Zhang Y, Sazinsky SL, Currell B, Yeh J-H, Reva B, Held HA, Appleton BA, Evangelista M, Wu Y *et al.*: **A specificity map for the PDZ domain family**. *PLoS Biol* 2008, **6**:e239.

67. Tonikian R, Xin X, Toret CP, Gfeller D, Landgraf C, Panni S, Paoluzi S, Castagnoli L, Currell B, Seshagiri S *et al.*: **Bayesian modeling of the yeast SH3 domain interactome predicts spatiotemporal dynamics of endocytosis proteins**. *PLoS Biol* 2009, **7**:e1000218.

68. Chen JR, Chang BH, Allen JE, Stiffler MA, MacBeath G: **Predicting PDZ domain–peptide interactions from primary sequences**. *Nat Biotechnol* 2008, **26**:1041-1045.

Please cite this article in press as: Stein A, et al. Three-dimensional modeling of protein interactions and complexes is going 'omics, Curr Opin Struct Biol (2011), doi:10.1016/j.sbi.2011.01.005

69. Smith CA, Kortemme T: **Structure-based prediction of the**
• **peptide sequence space recognized by natural and synthetic PDZ domains**. *J Mol Biol* 2010, **402**:460-474.
By incorporating backbone simulations, the authors accurately predict the peptide binding preferences for a number of PDZ domains, including those that deviate from the peptide bound in the structure (as shown by phage display screens). Furthermore, the method is able to predict preferences of mutant domains, which is promising for protein interaction prediction and design.

70. Fu X, Apgar JR, Keating AE: **Modeling backbone flexibility to achieve sequence diversity: the design of novel alpha-helical ligands for Bcl-xL**. *J Mol Biol* 2007, **371**:1099-1117.

71. Lessene G, Czabotar PE, Colman PM: **BCL-2 family antagonists for cancer therapy**. *Nat Rev Drug Discov* 2008, **7**:989-1000.

72. Grigoryan G, Reinke AW, Keating AE: **Design of protein-**
•• **interaction specificity gives selective bZIP-binding peptides**. *Nature* 2009, **458**:859-864.
The authors present a computational framework for designing protein interaction specificity that considers both positive and negative constraints in order to achieve affinity as well as binding specificity. They apply the method to bZIP transcription factors, a family of very similar proteins, and experimentally validate the high selectivity of the designed peptides.

73. Zhang Y, Thiele I, Weekes D, Li Z, Jaroszewski L, Ginalski K,
•• Deacon AM, Wooley J, Lesley SA, Wilson IA *et al.*: **Three-dimensional structural view of the central metabolic network of *Thermotoga maritima***. *Science* 2009, **325**:1544-1549.
This paper presents the first 3D reconstruction of the central metabolic network of the bacterium *Thermotoga maritima*, using experimentally determined and modeled structures. On the basis of this network the authors show that relatively few folds, which are all present in a 30% core of the network, are used to accomplish the metabolic functions.

74. Kim PM, Lu LJ, Xia Y, Gerstein MB: **Relating three-dimensional structures to protein networks provides evolutionary insights**. *Science* 2006, **314**:1938-1941.

75. Kim PM, Sboner A, Xia Y, Gerstein M: **The role of disorder in interaction networks: a structural analysis**. *Mol Syst Biol* 2008, **4**:179.

76. Tuncbag N, Kar G, Gursoy A, Keskin O, Nussinov R: **Towards inferring time dimensionality in protein–protein interaction networks by integrating structures: the p53 example**. *Mol Biosyst* 2009, **5**:1770-1778.

77. Kundrotas PJ, Zhu Z, Vakser IA: **GWIDD: genome-wide protein docking database**. *Nucleic Acids Res* 2010, **38**:D513-D517.

78. Aloy P, Russell R: **Targeting and tinkering with interaction networks**. *FEBS Lett* 2008, **582**:1219.

79. Zhong Q, Simonis N, Li Q-R, Charloteaux B, Heuze F, Klitgord N, Tam S, Yu H, Venkatesan K, Mou D *et al.*: **Edgetic perturbation models of human inherited disorders**. *Mol Syst Biol* 2009, **5**:321.

80. Dreze M, Charloteaux B, Milstein S, Vidalain P-O, Yildirim MA, Zhong Q, Svrzikapa N, Romero V, Laloux G, Brasseur R *et al.*: **'Edgetic' perturbation of a *C. elegans* BCL2 ortholog**. *Nat Methods* 2009, **6**:843-849.

81. Wells JA, McClendon CL: **Reaching for high-hanging fruit in drug discovery at protein–protein interfaces**. *Nature* 2007, **450**:1001-1009.

82. Russell RB, Aloy P: **Targeting and tinkering with interaction networks**. *Nat Chem Biol* 2008, **4**:666-673.

83. Arkin MR, Wells JA: **Small-molecule inhibitors of protein–protein interactions: progressing towards the dream**. *Nat Rev Drug Discov* 2004, **3**:301-317.

84. Davis FP, Sali A: **The overlap of small molecule and protein**
• **binding sites within families of protein structures**. *PLoS Comput Biol* 2010, **6**:e1000668.
Using structures of protein–protein and protein–ligand interactions, the authors identify 'bi-functional positions' that are involved in both types of interactions, via homology transfer. They show that bi-functional positions are significantly less conserved than those in other interfaces, and discuss different classes of protein–protein interfaces in which such positions are found.

85. London N, Raveh B, Movshovitz-Attias D, Schueler-Furman O: **Can self-inhibitory peptides be derived from the interfaces of globular protein–protein interactions?** *Proteins* 2010, **78**:3140-3149.

86. Rubinstein M, Niv MY: **Peptidic modulators of protein–protein interactions: progress and challenges in computational design**. *Biopolymers* 2009, **91**:505-513.

87. Aranda B, Achuthan P, Alam-Faruque Y, Armean I, Bridge A, Derow C, Feuermann M, Ghanbarian AT, Kerrien S, Khadake J *et al.*: **The IntAct molecular interaction database in 2010**. *Nucleic Acids Res* 2010, **38**:D525-D531.

88. Breitkreutz B-J, Stark C, Reguly T, Boucher L, Breitkreutz A, Livstone M, Oughtred R, Lackner DH, Bähler J, Wood V *et al.*: **The BioGRID Interaction Database: 2008 update**. *Nucleic Acids Res* 2008, **36**:D637-D640.

89. Ceol A, Aryamontri AC, Licata L, Peluso D, Briganti L, Perfetto L, Castagnoli L, Cesareni G: **MINT, the molecular interaction database: 2009 update**. *Nucleic Acids Res* 2010, **38**:D532-D539.

90. Chautard E, Ballut L, Thierry-Mieg N, Ricard-Blum S: **MatrixDB, a database focused on extracellular protein–protein and protein–carbohydrate interactions**. *Bioinformatics* 2009, **25**:690-691.

91. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A *et al.*: **Human Protein Reference Database—2009 update**. *Nucleic Acids Res* 2009, **37**:D767-D772.

92. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D: **The Database of Interacting Proteins: 2004 update**. *Nucleic Acids Res* 2004, **32**:D449-D451.

93. Pieper U, Eswar N, Webb BM, Eramian D, Kelly L, Barkan DT, Carter H, Mankoo P, Karchin R, Marti-Renom MA *et al.*: **MODBASE, a database of annotated comparative protein structure models and associated resources**. *Nucleic Acids Res* 2009, **37**:D347-D354.

94. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K *et al.*: **The Pfam protein families database**. *Nucleic Acids Res* 2010, **38**:D211-D222.

Please cite this article in press as: Stein A, et al. Three-dimensional modeling of protein interactions and complexes is going 'omics, Curr Opin Struct Biol (2011), doi:10.1016/j.sbi.2011.01.005