A structural perspective on protein–protein interactions

Robert B Russell[1,2], Frank Alber[3], Patrick Aloy[1], Fred P Davis[3], Dmitry Korkin[3], Matthieu Pichaud[1], Maya Topf[3] and Andrej Sali[3]*

Structures of macromolecular complexes are necessary for a mechanistic description of biochemical and cellular processes. They can be solved by experimental methods, such as X-ray crystallography, NMR spectroscopy and electron microscopy, as well as by computational protein structure prediction, docking and bioinformatics. Recent advances and applications of these methods emphasize the need for hybrid approaches that combine a variety of data to achieve better efficiency, accuracy, resolution and completeness.

Addresses
[1]EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany
[2]EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, UK
[3]Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, and California Institute for Quantitative Biomedical Research, Mission Bay Genentech Hall, Suite N472D, 600 16th Street, University of California at San Francisco, San Francisco, California 94143-2240, USA
*e-mail: sali@salilab.org

Current Opinion in Structural Biology 2004, **14**:313–324

This review comes from a themed issue on
Sequences and topology
Edited by Peer Bork and Christine A Orengo

Available online 18th May 2004

0959-440X/$ – see front matter

DOI 10.1016/j.sbi.2004.04.006

Introduction

Genome sequencing has provided nearly complete lists of the macromolecules present in many organisms (e.g. [1,2]). However, these lists reveal comparatively little about the function of biological systems because the functional units of cells are often complex assemblies of several macromolecules [3]. Such complexes vary widely in their activity and size [3–7], and play crucial roles in most cellular processes. They are often depicted as molecular machines [3], a metaphor that accurately captures many of their characteristic features, such as modularity, complexity, cyclic functions and energy consumption [8]. For instance, the nuclear pore complex, a 50–100 MDa protein assembly, regulates and controls the trafficking of macromolecules through the nuclear envelope [9]; the ribosome is responsible for protein biosynthesis; RNA polymerase catalyzes the formation of RNA [10]; and ATP synthase catalyzes the formation of ATP [7]. Macromolecular assemblies are also involved in transcription control (e.g. the IFNβ enhanceosome) [6,11] and the regulation of cellular transport (e.g. microtubulins in complex with the molecular motors myosin or kinesin) [12–14], and are crucial components in neuronal signaling (e.g. the post-synaptic density complexes) [15]. A structural description of the protein interactions within such complexes is an important step toward a mechanistic understanding of biochemical, cellular and higher order biological processes [16–18,19•].

There are currently about 12 000 known structures, from a variety of organisms, of assemblies involving two or more protein chains (http://pqs.ebi.ac.uk/pqs-doc.shtml) (April 2004) [20]; these complexes can be organized into about 3500 groups based on sequence similarity [19•]. Just how many complexes exist in a particular proteome is not easy to deduce because of the different component types (e.g. proteins, nucleic acids, nucleotides, metal ions) and the varying life span of the complexes (e.g. transient complexes, such as those involved in signaling, and stable complexes, such as the ribosome). Until recently, the most comprehensive information about protein–protein interactions was available for the *Saccharomyces cerevisiae* proteome, consisting of approximately 6200 proteins. This information has been provided by methods such as the yeast two-hybrid system and affinity purification followed by mass spectrometry [21–27,28•,29]. The lower bound on binary protein interactions and functional links in yeast has been estimated to be in the range of approximately 30 000 [30,31]; this number corresponds to about nine protein partners per protein, although not necessarily all direct or interacting at the same time. The human proteome may have an order of magnitude more complexes than the yeast cell and the number of different complexes across all relevant genomes may be several times larger still. Therefore, there may be thousands of biologically relevant macromolecular complexes whose structures are yet to be characterized [32].

We review here recent developments in the experimental and computational techniques that have allowed structural biology to shift its focus from the structures of individual proteins to the structures of large assemblies [19•,33,34]. We also illustrate these developments by listing their application to the determination of the structure of specific assemblies of biological importance. In contrast to structure determination of individual proteins, structural characterization of macromolecular assemblies usually poses a more difficult challenge. We stress that a comprehensive structural description of large complexes generally requires the use of several experimental methods, underpinned by a variety of theoretical approaches to maximize efficiency, completeness, accuracy and resolution [19•,35].

X-ray crystallography and NMR spectroscopy

X-ray crystallography has been the most prolific technique for the structural analysis of proteins and protein complexes, and is still the 'gold standard' in terms of accuracy and resolution (Figure 1a). Structures of several macromolecular assemblies have recently been solved by X-ray crystallography: RNA polymerase [36], the ribosomal subunits [37–41], the complete ribosome and its functional complexes [42], the proteasome [43], the GroEL chaperonin [44], various complexes of the cellular transport machinery [12,13], the Arp2/3 complex [45], photosystem I and the light-harvesting complex of photosystem II [46,47], the SRP (signal recognition particle) complex involved in nascent protein targeting [48•], and various viral capsid and virion structures [49–51]. However, the number of structures of macromolecular assemblies solved by X-ray crystallography is still quite small compared to that of the individual proteins and it will probably be many years before we have a complete repertoire of high-resolution structures for the hundreds of complexes in a typical cell. This discrepancy is due mainly to the difficult production of sufficient quantities of the sample and its crystallization.

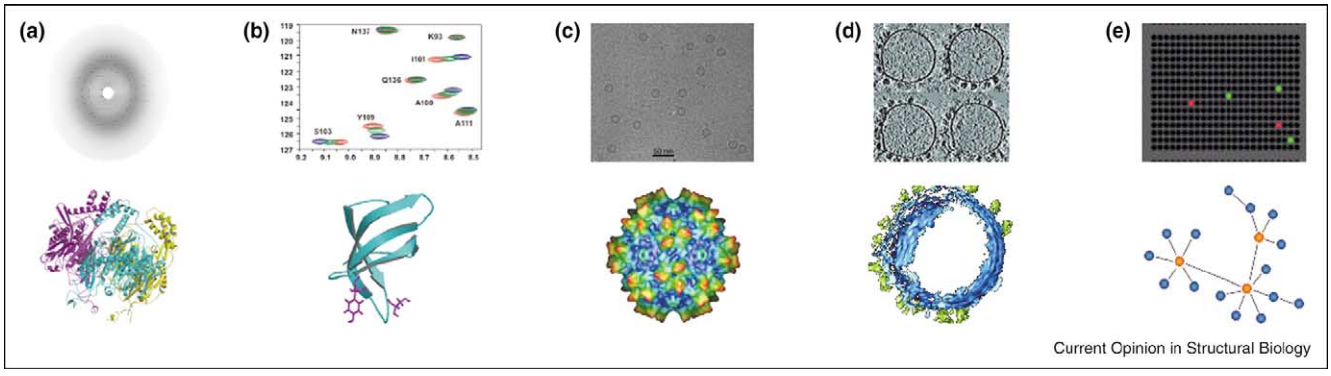
NMR spectroscopy allows the determination of atomic structures of ever-larger subunits and even their complexes [52–54]. Increasingly, it is used to identify residues involved in protein interactions (Figure 1b) [55–58]. Recent technical advances have allowed its application to systems as large as the 900 kDa GroEL–GroES complex [52]. Also, it was recently used to describe structural differences between interactions among different LIM and SH3 domains [59].

Electron microscopy and electron tomography

There are several variants of electron microscopy (EM), including single-particle EM (Figure 1c) [60], electron tomography (Figure 1d) [61] and electron crystallography of regular two-dimensional arrays of the sample [62].

For particles with molecular weights greater than 200–500 kDa, single-particle cryo-EM can determine the

Figure 1



Methods for the structural characterization of macromolecular assemblies. **(a)** Electron diffraction map and three-dimensional structure of the bacterial degradosome component PNPase (polyribonucleotide phosphorylase) determined by X-ray crystallography [183]. X-ray crystallography integrates the diffraction patterns collected after bombarding a crystallized protein or complex with X-rays to construct its three-dimensional structure. In principle, there is no size limit on the structures studied using this technique, although it is often difficult to obtain sufficient material for crystallization. This technique provides atomic-resolution structures and thus molecular details of how the interactions between the different components occur. **(b)** Three-dimensional structure and plot showing chemical shifts upon association of the human survival motor neuron (SMN) tudor domain solved by NMR spectroscopy [184]. NMR spectroscopy extracts distances between atoms by measuring transitions between different nuclear spin states within a magnetic field. These distances are then used as restraints to build three-dimensional structures. NMR spectroscopy also provides atomic-resolution structures, but is generally limited to proteins of about 300 residues. It plays an increasingly important role in studying interaction interfaces between structures determined independently. **(c)** EM micrograph and three-dimensional reconstruction of adeno-associated virus type 2 empty capsids [185]. EM is based on the analysis of images of stained particles. Different views and conformations of the complexes are trapped and thus thousands of images have to be averaged to reconstruct the three-dimensional structure. Classical implementations were limited to a resolution of 20 Å. More recently, single-particle cryo techniques, whereby samples are fast frozen before study, have reached resolutions as high as approximately 6 Å. EM provides information about the overall shape and symmetry of macromolecules. **(d)** Slice images and rendered surface of a ribosome-decorated portion of endoplasmic reticulum [73]. In electron tomography, the specimen studied is progressively tilted upon an axis perpendicular to the electron beam. A set of projection images is then recorded and used to build a three-dimensional model. This technique can tackle large organelles or even complete cells without perturbing their physiological environment. It provides shape information at resolutions of approximately 30 Å and promises to reach higher resolutions soon. **(e)** Yeast two-hybrid array screen and small network of interacting proteins [124••,186]. Interaction discovery comprises many different methods whose objective is to determine spatial proximity between proteins. These include techniques such as the two-hybrid system, affinity purification, FRET, chemical cross-linking, footprinting and protein arrays. These methods provide very limited structural information and no molecular details. Their strength is that they often give a quasi-comprehensive list of protein interactions and the networks they form.

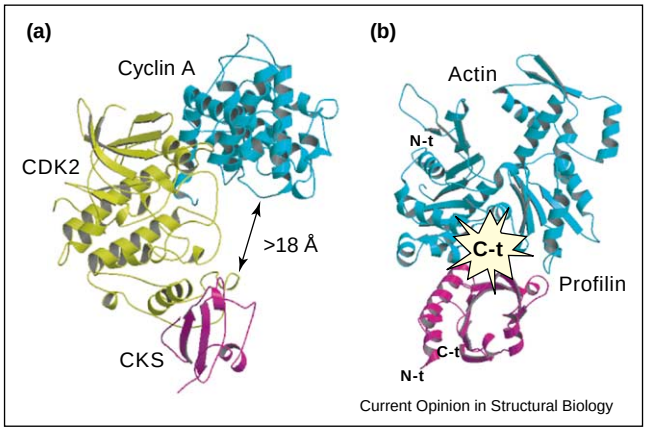
electron density of an assembly at resolutions as high as approximately 5 Å [63•,64–66,67•,68•,69,70•]. The full three-dimensional structure of the particle is reconstructed from many two-dimensional projections of the specimen, each showing the object from a different angle. Imaging by cryo-EM requires neither large quantities of the sample nor the sample in a crystalline form. Therefore, single-particle cryo-EM is a powerful tool to investigate the structure and dynamics of macromolecular assemblies for which X-ray structure determination is very difficult. Although it is generally impossible to build atomic models solely from cryo-EM density maps, the maps give valuable insights into the structure and mechanism of large complexes (e.g. [63•]). They are particularly useful when combined with atomic-resolution structures of the subunits, as reviewed in the section on hybrid methods below.

One of the most exciting developments in structural biology is the new generation of tomography methods based on multiple tilted views of the same object [33,71]. Although electron tomography can be used to study the structures of isolated macromolecular assemblies at a relatively low resolution of a few nanometers, its true potential lies in visualizing the assemblies in an unperturbed cellular context [72]. These data sets provide fascinating three-dimensional images of entities as large as a small cell at approximately 5 nm resolution [73]. To widen the scope of cellular tomography, it is necessary to improve the resolution of the tomographic images, as well as identify the structures in these images [73–75]. Theoretical considerations [76] and ongoing improvements in the instrumentation make a resolution as high as 2 nm a realistic goal [77].

Low-resolution experimental methods

Several experimental techniques can provide structural information about protein interactions at low resolution (Figure 1e). This information may be used to infer the configuration of the proteins in a complex. Methods for the mapping of protein interactions may provide contact or proximity restraints for pairs of proteins that are useful in the modeling of higher order complexes. Such methods include new implementations of the two-hybrid system [78–81], tagged affinity chromatography [82•,83] and the combination of phage display with other techniques [84], such as synthesis of peptides on cellulose membranes (SPOT) [85••]. Because of the low-resolution nature of these biochemical characterizations, care is needed in their interpretation. For example, comparing biochemically derived interaction sets against known three-dimensional structures of complexes revealed potential sources of systematic errors in interaction discovery, such as indirect interactions in two-hybrid systems, the obstruction of interfaces by molecular labels and artificial promiscuity in the detected interactions (Figure 2) [86].

Figure 2



Examples of potential errors in biochemical interaction discovery techniques, as revealed by a structure-based analysis [17]. **(a)** Indirect interactions between cyclin-dependent kinase regulatory subunit (CKS) and cyclin A detected by the yeast two-hybrid system. Several interactions between CKS domains and cyclins were reported in genome-scale two-hybrid studies [21,187]. However, analysis of three-dimensional structures suggests that the endogenous cyclin-dependent kinase 2 (CDK2) probably mediates the interaction, as combining the CDK2–CKS and CDK2–cyclin A structures places the CKS and cyclin domains 18 Å apart [86]. **(b)** An example of an interaction that is not detected by any screen, possibly because molecular labels (e.g. affinity purification tags, or two-hybrid DNA binding or activation domains) are interfering with the interaction. The X-ray structure of the actin–profilin complex reveals that the actin C terminus (C-t) lies at the interaction interface (the other N and C termini are also labeled).

Biophysical, biochemical and molecular biology methods can also be used to derive low-resolution information about the relative position and orientation of domains in a larger complex. These methods include: site-directed mutagenesis, which can identify the residues that mediate the interaction [87]; various forms of footprinting, such as hydrogen-deuterium exchange [88,89•] and hydroxyl radical footprinting [90], which can identify surfaces buried upon complex formation; chemical cross-linking [91–93], which can identify interacting residues; fluorescence resonance energy transfer (FRET) [94,95], which can determine the distance between labeled groups on the interacting proteins; and Fourier transform IR spectroscopy (FTIR), which describes structural changes upon complex formation [96]. Small angle X-ray scattering (SAXS) is another biophysical method that can provide low-resolution information about the shape of a complex. For instance, SAXS has recently been used to study the dynamics of conformational change in Bruton tyrosine kinase [97,98].

Computational protein–protein docking

When atomic structures of the individual proteins involved in an interaction are known, either by experiment or by modeling, several computational methods are available that suggest the structure of the interaction [99].

Most of these docking methods aim to predict the atomic model of a complex by maximizing the shape and chemical complementarity between a given pair of interacting proteins [99–102]. Docking strategies usually rely on a two-stage approach: they first generate a set of possible orientations of the two docked proteins and then score them in the hope that the native complex will be ranked highly. Katchalski-Katzir, Vakser and co-workers [103] have pioneered a fast Fourier transform (FFT)-based method for rapidly searching through the space of possible docked configurations. Due to its computational efficiency, it has also been incorporated into programs such as FTDock and 3D-Dock [104–106], GRAMM [107], DOT [108], ZDOCK [109] and HEX [110]. Other docking programs include ICM [111] and ROSETTA [112]. The searches may be restrained by other considerations, such as the known binding site location. These methods differ in protein representation, in the scoring of different configurations and in the search for the best solutions. Some methods boldly model the actual diffusion/collision trajectories involved in the docking process [113,114].

Although docking methods are not sufficiently accurate to predict whether or not two proteins actually interact with each other, they can sometimes correctly identify the interacting surfaces between two structurally defined subunits [115]. Docking methods are systematically assessed through blind trials in the Critical Assessment of PRedicted Interactions (CAPRI) [101,116•,117]. Predictions are made just before the structures are solved experimentally, followed by the assessment of the models at the CAPRI meetings. The best of the methods assessed in the last CAPRI experiment correctly predicted three of the seven target complexes [116•].

Methods that are able to work with comparative protein structure models [118] instead of experimentally determined subunit structures would extend the applicability of docking to many more biological problems, but would probably have poorer performance. Currently, docking is often applied in concert with experimental techniques, including site-directed mutagenesis [119], amide hydrogen-deuterium exchange [89•] and NMR spectroscopy [120•,121], as well as solid-state binding and surface plasmon resonance [122].

Inferring interactions by homology

Protein interactions can also be modeled by similarity [123,124••,125]. If a complex of known structure comprising homologs of a pair of interacting proteins is available, it is usually possible to build a model by comparative modeling [126]. There are now approximately 2000 distinct interaction types of known structure (i.e. whereby interacting domains sharing 30% or greater sequence identity are considered to be a single type; P Aloy, RB Russell, unpublished).

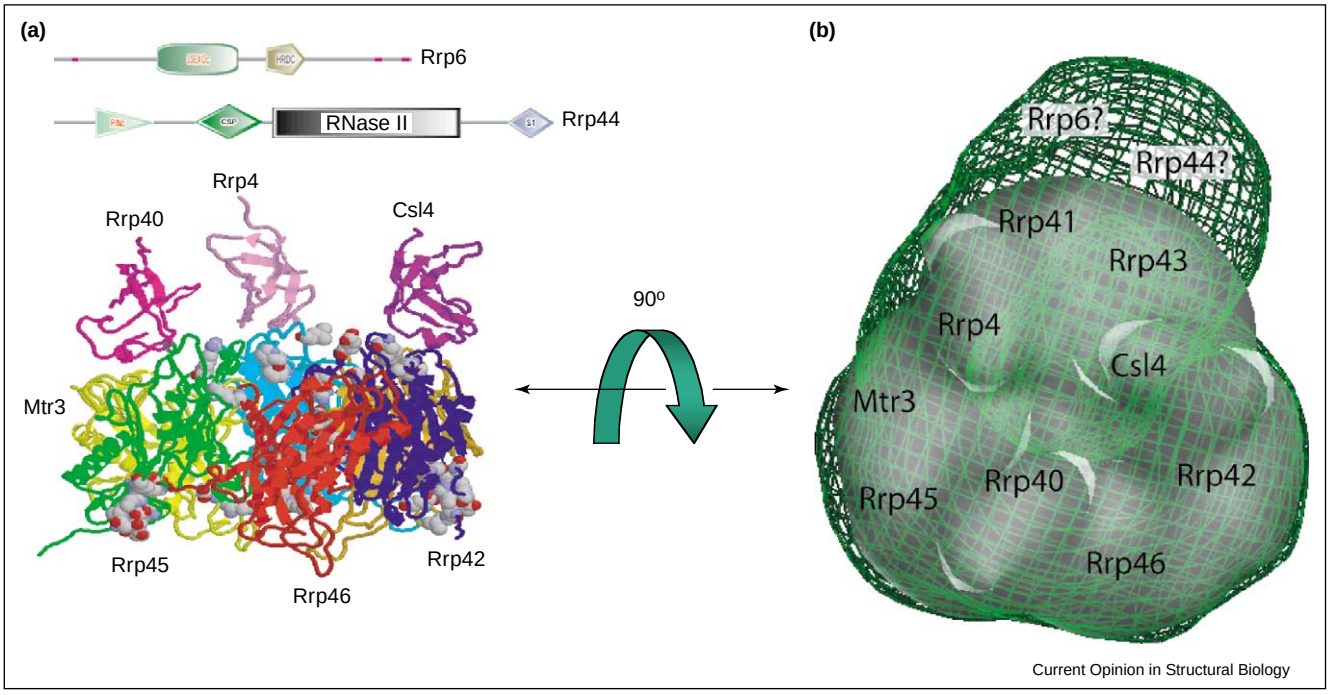
Building a model of the interaction between a pair of proteins based on the known structure of the complex between interacting homologs raises the question of whether or not homology of the subunits implies similarity of interaction. It was found that interactions between proteins of the same fold tend to be similar when the sequence identity is above approximately 30% [127•]. Below this cutoff, there is a twilight zone where interactions may or may not be similar geometrically.

Given a template, it is possible to model an interaction using standard comparative modeling techniques [126]. However, frequently there are multiple templates for the same interaction type. In addition, a single interaction template can be used to model many putative interactions in a single organism. Therefore, it is important to assess the likelihood of these potential interactions, particularly in the absence of experimental validation [128••]. For example, each of the dozens of fibroblast growth factors (FGFs) interacts with one or more of seven receptors with different affinities [129]. Two approaches have been developed recently that attempt to predict specificity by modeling interactions. The first approach, implemented by InterPReTS [123,130] and ModBase [125], uses empirical pair potentials derived from interfaces of known structure to score how well a pair of homologous proteins fits a known complex structure. The second approach, MULTIPROSPECTOR, is similar, although it attempts to study more distantly related protein sequences by threading sequences onto a library of interacting templates, followed by scoring how well the individual sequences fit their proposed folds and the interface between them [131]. Both approaches have since been applied to study large collections of sequences and interactions [124••,125,132••].

For some large complexes, the specificity of interactions within a family of homologous subunits is an important determinant of complex assembly. For instance, the chaperonin CCT consists of eight homologous subunits that are all similar to the single subunit comprising the thermosome [133]. Thus, building CCT using the thermosome requires the conversion of a seven-subunit ring into an eight-subunit ring, and then choosing the correct arrangement from the 5040 (8!/8) possibilities. It is possible to guide this process by experiment, such as the detection of subcomplexes that reveal preferred interacting pairs [134] or the application of the two-hybrid system [135]. InterPReTS was also applied to select one of the 120 possible arrangements of six exosome subunits (Figure 3) [136], with mixed results.

In eukaryotes, many of the protein–protein interactions in regulatory signaling networks are mediated by modular protein interaction domains. Such domains appear to have been used in a modular fashion throughout evolution to generate novel connections between proteins.

Figure 3



Putative structure through modeling and low-resolution EM. **(a)** Exosome subunits. The top of the panel shows the domain organization of two subunits present in the complex, but lacking any detectable similarity to known three-dimensional structures. The model for the nine other subunits (bottom) was constructed by predicting binary interactions using InterPReTS [130] and building models based on a homologous complex structure using comparative modeling. **(b)** EM density map (green mesh) with the best fit of the model shown as a gray surface and the predicted locations of the subunits labeled. The question marks indicate those subunits for which no structures could be modeled.

However, the repeated use of such domains presents the problem of specificity: how are biologically unique connections made? Experimental evidence suggests that some of these interactions are remarkably specific [137], whereas others show overlapping specificity [85••]. These findings highlight the need for the development and validation [128••] of accurate computational methods that capture the structural principles of protein interaction specificity.

Low-resolution computational methods

Even when docking or modeling is not feasible, it may still be possible to get some structural insight into a protein–protein interaction using other computational approaches. Various methods combine structures with sequence alignments and phylogenetic trees to identify sites on the surface that are likely to be involved in function or specificity [138–145]. Other computational methods perform alanine scanning to identify 'hot spots' at known protein interfaces. A comparative analysis of such hot spots may reveal determinants of specificity and cross-reactivity [146–148]. There are also many computational methods for the prediction of protein–protein interactions when no structural information is available (see the review by Bork *et al.* in this section).

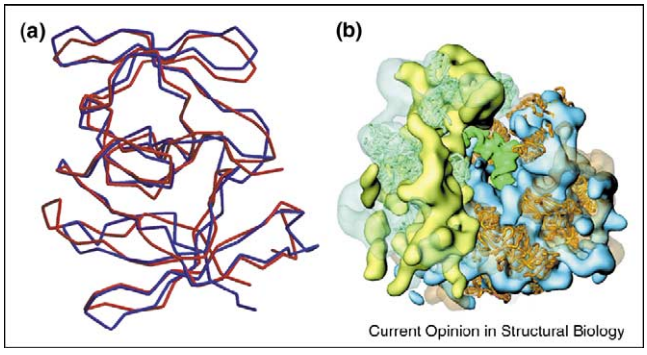
Hybrid methods

In the absence of atomic-resolution data, approximate atomic models of assemblies can be derived by combining low-resolution cryo-EM data on complete protein assemblies with computational docking of atomic-resolution structures of their subunits [149–156]. It has been estimated that using such fitting techniques improves the accuracy to up to one-tenth the resolution of the original EM reconstruction.

Hybrid approaches involving the fitting of subunits into EM maps are illustrated by pseudo-atomic models of the actin–myosin complex [157], the yeast ribosome [158,159] (Figure 4), the bacteriophage T4 baseplate [160••], pre-mRNA splicing complex SF3b [161], the RAD51 system involved in homologous recombination and DNA repair [162•], and complex virus structures [163,164].

Unfortunately, experimentally determined atomic-resolution structures of the isolated subunits are frequently not available. Even when available, the induced fit can severely limit their utility in the reconstruction of the whole assembly. In such cases, it might be possible to get useful models of the subunits by comparative protein

Figure 4



Hybrid assembly of the 80S ribosome from yeast [158]. **(a)** Superposition of a comparative protein structure model (red) of a domain from ribosomal protein L2 from *Bacillus stearothermophilus* with the actual structure (blue) (PDB code 1RL2). **(b)** A partial molecular model of the whole yeast ribosome calculated by fitting atomic rRNA (not shown) and comparative protein structure models (ribbon representation) into the electron density of the 80S ribosomal particle.

structure modeling [126,165–168]. The number of models that can be constructed with useful accuracy is already two orders of magnitude greater than the number of available experimentally determined structures. Models with at least the correct fold can be constructed for domains in approximately 58% of known protein sequences [125]. Comparative modeling will be increasingly more applicable and accurate because of structural genomics initiatives [169]. One of the main goals of structural genomics is to determine a sufficient number of appropriately selected structures from each domain family, such that all sequences are within modeling distance of at least one known protein structure [170,171].

Structural genomics may in fact contribute to a comprehensive and efficient structural description of complexes in an additional way. Although structural genomics currently focuses on single proteins or their domains, it could be expanded to the sampling of domain–domain interactions [127•,172,173]. Such an effort would provide a repertoire of templates for binary interactions, which would facilitate the building of higher order complexes.

Although X-ray crystallography and EM in combination with atomic structure docking have been successfully employed to solve the structures of protein assemblies, they are not capable of efficiently characterizing the myriad of complexes that exist in a cell. For example, most transient complexes cannot be addressed at all with these approaches. Therefore, there is a great need for the additional development of hybrid methods through which accuracy, high throughput, completeness and resolution are improved by integrating information from all available sources [19•,136,174].

The dynamics of complexes

By trapping complexes in different conformations and configurations, hybrid methods can be used to study the functional role of assembly dynamics. For instance, models of the two different functional states of the *Escherichia coli* 70S ribosome demonstrated that the complex changes from a compact to a looser conformation, and showed rearrangements of many of the ribosomal proteins [63•]. Similarly, T antigen double hexamers (a replicative helicase of simian virus 40) were assembled at the origin of replication using 27.5 Å cryo-EM maps at different degrees of bending along the DNA axis [175]. Fitting the crystal structure of the Tag helicase domain [176] into the three-dimensional cryo-EM density map ascertained that the C-terminal domains are rotated relative to each other in the complex. The results were combined with the available biochemical data to propose an integrated model for the initiation of viral DNA replication. Such a comparison also revealed details that are key to understanding filament function. Fitting atomic models of actin and the myosin cross-bridge into 14 Å cryo-EM maps showed that the closing of the actin-binding cleft upon actin binding is structurally coupled to the opening of the nucleotide-binding pocket [67•].

The dynamics of assembly models can also be studied by theoretical calculations [177–180]. A vibrational analysis of elastic models was employed to capture the essential motions of clamp closure in bacterial RNA polymerase, the ratcheting of the 30S and 50S subunits of the ribosome, and the dynamic flexibility of chaperonin CCT [181]. Also, a quantized elastic deformational model provided a basis for the simulation of conformational fluctuations related to the expansion and contraction of the truncated E2 core from the pyruvate dehydrogenase complex [182].

Conclusions

There is a wide spectrum of experimental and computational methods for the identification and structural characterization of macromolecular complexes. These methods need to be combined into hybrid approaches to achieve greater accuracy, coverage, resolution and efficiency than any of the individual methods. New methods must be capable of generating possible alternative models consistent with information such as stoichiometry, interaction data, homology to known structures, docking results and low-resolution images. There is a need to describe the structures and dynamics of both stable and transient complexes.

Structural biology is a great unifying discipline of biology. Thus, structural characterization of many protein complexes may be the way to bridge the gaps between genome sequencing, functional genomics, proteomics and systems biology. The goal seems daunting, but the prize will be commensurate with the effort invested,

given the importance of molecular machines and functional networks in biology and medicine.

Acknowledgements

We are grateful to Tanja Kortemme, Damien Devos, MS Madhusudan, Narayanan Eswar, Mike Kim, Matt Baker, Wah Chiu, Wolfgang Baumeister and David Agard for discussions about the modeling of assembly structures. We also acknowledge the support of the NIH, NSF, HFSP, SUN, IBM, Intel and The Sandler Family Supporting Foundation (AS).

References and recommended reading

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

- • of special interest
- •• of outstanding interest

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al.*: **Initial sequencing and analysis of the human genome**. *Nature* 2001, **409**:860-921.

2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA *et al.*: **The sequence of the human genome**. *Science* 2001, **291**:1304-1351.

3. Alberts B: **The cell as a collection of protein machines: preparing the next generation of molecular biologists**. *Cell* 1998, **92**:291-294.

4. Goto NK, Zor T, Martinez-Yamout M, Dyson HJ, Wright PE: **Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the KIX domain**. *J Biol Chem* 2002, **277**:43168-43174.

5. Grakoui A, Bromley SK, Sumen C, Davis MM, Shaw AS, Allen PM, Dustin ML: **The immunological synapse: a molecular machine controlling T cell activation**. *Science* 1999, **285**:221-227.

6. Courey AJ: **Cooperativity in transcriptional control**. *Curr Biol* 2001, **11**:R250-R252.

7. Noji H, Yoshida M: **The rotary machine in the cell, ATP synthase**. *J Biol Chem* 2001, **276**:1665-1668.

8. Nogales E, Grigorieff N: **Molecular machines: putting the pieces together**. *J Cell Biol* 2001, **152**:F1-F10.

9. Rout MP, Aitchison JD, Suprapto A, Hjertaas K, Zhao Y, Chait BT: **The yeast nuclear pore complex: composition, architecture, and transport mechanism**. *J Cell Biol* 2000, **148**:635-651.

10. Murakami KS, Darst SA: **Bacterial RNA polymerases: the wholo story**. *Curr Opin Struct Biol* 2003, **13**:31-39.

11. Nogales E: **Recent structural insights into transcription preinitiation complexes**. *J Cell Sci* 2000, **113**:4391-4397.

12. Vale RD: **The molecular motor toolbox for intracellular transport**. *Cell* 2003, **112**:467-480.

13. Goldstein LS, Yang Z: **Microtubule-based transport systems in neurons: the roles of kinesins and dyneins**. *Annu Rev Neurosci* 2000, **23**:39-71.

14. Vale RD, Milligan RA: **The way things move: looking under the hood of molecular motor proteins**. *Science* 2000, **288**:88-95.

15. Kennedy MB: **Signal-processing machines at the postsynaptic density**. *Science* 2000, **290**:750-754.

16. Park J, Lappe M, Teichmann SA: **Mapping protein family interactions: intramolecular and intermolecular protein family interaction repertoires in the PDB and yeast**. *J Mol Biol* 2001, **307**:929-938.

17. Aloy P, Russell RB: **Potential artefacts in protein-interaction networks**. *FEBS Lett* 2002, **530**:253-254.

18. Edwards A, Kus B, Jansen R, Greenbaum D, Greenblatt J, Gerstein M: **Bridging structural biology and genomics: assessing protein-interaction data with known complexes**. *Trends Genet* 2002, **18**:529-536.

19. Sali A, Glaeser R, Earnest T, Baumeister W: **From words to • literature in structural proteomics**. *Nature* 2003, **422**:216-225.
This review summarizes current efforts in structural proteomics. It provides a plethora of references for the various experimental and theoretical methods used to obtain structural information about macromolecular assemblies.

20. Henrick K, Thornton JM: **PQS: a protein quaternary structure file server**. *Trends Biochem Sci* 1998, **23**:358-361.

21. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P *et al.*: **A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae***. *Nature* 2000, **403**:623-627.

22. Ito T, Chiba T, Yoshida M: **Exploring the protein interactome using comprehensive two-hybrid projects**. *Trends Biotechnol* 2001, **19**:S23-S27.

23. Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, Lenzen G, Petel F, Wojcik J, Schachter V *et al.*: **The protein-protein interaction map of *Helicobacter pylori***. *Nature* 2001, **409**:211-215.

24. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM *et al.*: **Functional organization of the yeast proteome by systematic analysis of protein complexes**. *Nature* 2002, **415**:141-147.

25. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K *et al.*: **Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry**. *Nature* 2002, **415**:180-183.

26. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E *et al.*: **A protein interaction map of *Drosophila melanogaster***. *Science* 2003, **302**:1727-1736.

27. Phizicky E, Bastiaens PI, Zhu H, Snyder M, Fields S: **Protein analysis on a proteomic scale**. *Nature* 2003, **422**:208-215.

28. Aebersold R, Mann M: **Mass spectrometry-based proteomics**. • *Nature* 2003, **422**:198-207.
This review provides a good overview of the role of mass-spectrometry-based proteomics in the study of protein–protein interactions, the mapping of cell organelles and the generation of quantitative protein profiles for a variety of species.

29. Andersen JS, Wilkinson CJ, Mayor T, Mortensen P, Nigg EA, Mann M: **Proteomic characterization of the human centrosome by protein correlation profiling**. *Nature* 2003, **426**:570-574.

30. Kumar A, Snyder M: **Protein complexes take the bait**. *Nature* 2002, **415**:123-124.

31. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P: **Comparative assessment of large-scale data sets of protein-protein interactions**. *Nature* 2002, **417**:399-403.

32. Abbott A: **Proteomics: the society of proteins**. *Nature* 2002, **417**:894-896.

33. Baumeister W: **Electron tomography: towards visualizing the molecular organization of the cytoplasm**. *Curr Opin Struct Biol* 2002, **12**:679-684.

34. Sali A, Kuriyan J: **Challenges at the frontiers of structural biology**. *Trends Cell Biol* 1999, **9**:M20-M24.

35. Alber F, Eswar N, Sali A: **Structure determination of macromolecular complexes by experiment and computation**. In *Nucleic Acids and Molecular Biology, Volume 15, Practical Bioinformatics*. Edited by Bujnicki JM. : Springer-Verlag; 2004:73-96.

36. Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA: **Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution**. *Cell* 1999, **98**:811-824.

37. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA: **The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution**. *Science* 2000, **289**:905-920.

38. Carter AP, Clemons WM, Brodersen DE, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V: **Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics**. *Nature* 2000, **407**:340-348.

39. Harms J, Schluenzen F, Zarivach R, Bashan A, Gat S, Agmon I, Bartels H, Franceschi F, Yonath A: **High resolution structure of the large ribosomal subunit from a mesophilic eubacterium**. *Cell* 2001, **107**:679-688.

40. Wimberly BT, Brodersen DE, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V: **Structure of the 30S ribosomal subunit**. *Nature* 2000, **407**:327-339.

41. Schluenzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, Janell D, Bashan A, Bartels H, Agmon I, Franceschi F *et al.*: **Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution**. *Cell* 2000, **102**:615-623.

42. Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH, Noller HF: **Crystal structure of the ribosome at 5.5 Å resolution**. *Science* 2001, **292**:883-896.

43. Lowe J, Stock D, Jap B, Zwickl P, Baumeister W, Huber R: **Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution**. *Science* 1995, **268**:533-539.

44. Braig K, Otwinowski Z, Hegde R, Boisvert DC, Joachimiak A, Horwich AL, Sigler PB: **The crystal structure of the bacterial chaperonin GroEL at 2.8 Å**. *Nature* 1994, **371**:578-586.

45. Robinson RC, Turbedsky K, Kaiser DA, Marchand JB, Higgs HN, Choe S, Pollard TD: **Crystal structure of Arp2/3 complex**. *Science* 2001, **294**:1679-1684.

46. Ben-Shem A, Frolow F, Nelson N: **Crystal structure of plant photosystem I**. *Nature* 2003, **426**:630-635.

47. Liu Z, Yan H, Wang K, Kuang T, Zhang J, Gui L, An X, Chang W: **Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution**. *Nature* 2004, **428**:287-292.

48. • Egea PF, Shan SO, Napetschnig J, Savage DF, Walter P, Stroud RM: **Substrate twinning activates the signal recognition particle and its receptor**. *Nature* 2004, **427**:215-221.
The 1.9 Å X-ray structure of a complex formed by two GTPases, the signal recognition particle (SRP) and its receptor (SR), demonstrates a novel mode of substrate twinning. The structure suggests a unique activation mechanism for the SRP family of GTPases.

49. Grimes J, Basak AK, Roy P, Stuart D: **The crystal structure of bluetongue virus VP7**. *Nature* 1995, **373**:167-170.

50. Oda Y, Saeki K, Takahashi Y, Maeda T, Naitow H, Tsukihara T, Fukuyama K: **Crystal structure of tobacco necrosis virus at 2.25 Å resolution**. *J Mol Biol* 2000, **300**:153-169.

51. Nakagawa A, Miyazaki N, Taka J, Naitow H, Ogawa A, Fujimoto Z, Mizuno H, Higashi T, Watanabe Y, Omura T *et al.*: **The atomic structure of rice dwarf virus reveals the self-assembly mechanism of component proteins**. *Structure* 2003, **11**:1227-1238.

52. Fiaux J, Bertelsen EB, Horwich AL, Wuthrich K: **NMR analysis of a 900K GroEL GroES complex**. *Nature* 2002, **418**:207-211.

53. Fushman D, Xu R, Cowburn D: **Direct determination of changes of interdomain orientation on ligation: use of the orientational dependence of ^{15}N NMR relaxation in Abl SH(32)**. *Biochemistry* 1999, **38**:10225-10230.

54. Nakanishi T, Miyazawa M, Sakakura M, Terasawa H, Takahashi H, Shimada I: **Determination of the interface of a large protein complex by transferred cross-saturation measurements**. *J Mol Biol* 2002, **318**:245-249.

55. Zuiderweg ER: **Mapping protein-protein interactions in solution by NMR spectroscopy**. *Biochemistry* 2002, **41**:1-7.

56. Frickel EM, Riek R, Jelesarov I, Helenius A, Wuthrich K, Ellgaard L: **TROSY-NMR reveals interaction between ERp57 and the tip of the calreticulin P-domain**. *Proc Natl Acad Sci USA* 2002, **99**:1954-1959.

57. Pellecchia M, Sebbel P, Hermanns U, Wuthrich K, Glockshuber R: **Pilus chaperone FimC-adhesin FimH interactions mapped by TROSY-NMR**. *Nat Struct Biol* 1999, **6**:336-339.

58. Fernandez C, Wider G: **TROSY in NMR studies of the structure and function of large biological macromolecules**. *Curr Opin Struct Biol* 2003, **13**:570-580.

59. Velyvis A, Vaynberg J, Yang Y, Vinogradova O, Zhang Y, Wu C, Qin J: **Structural and functional insights into PINCH LIM4 domain-mediated integrin signaling**. *Nat Struct Biol* 2003, **10**:558-564.

60. Frank J: **Single-particle imaging of macromolecules by cryo-electron microscopy**. *Annu Rev Biophys Biomol Struct* 2002, **31**:303-319.

61. Baumeister W, Grimm R, Walz J: **Electron tomography of molecules and cells**. *Trends Cell Biol* 1999, **9**:81-85.

62. Nogales E, Wolf SG, Downing KH: **Structure of the alpha beta tubulin dimer by electron crystallography**. *Nature* 1998, **391**:199-203.

63. • Gao H, Sengupta J, Valle M, Korostelev A, Eswar N, Stagg SM, Roey PV, Agrawal RK, Harvey SC, Sali A *et al.*: **Study of the structural dynamics of the *E. coli* 70S ribosome using real-space refinement**. *Cell* 2003, **113**:789-801.
Forty-four atomic models built by comparative modeling were fit into the cryo-EM density map of the 70S unit of the *E. coli* ribosome (11.5 Å resolution) in two different functional states. The work demonstrates that the ribosome changes from a compact structure to a looser one, coupled with the rearrangement of many of the proteins.

64. Halic M, Becker T, Pool MR, Spahn CM, Grassucci RA, Frank J, Beckmann R: **Structure of the signal recognition particle interacting with the elongation-arrested ribosome**. *Nature* 2004, **427**:808-814.

65. Davis JA, Takagi Y, Kornberg RD, Asturias FA: **Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction**. *Mol Cell* 2002, **10**:409-415.

66. Yonekura K, Maki-Yonekura S, Namba K: **Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy**. *Nature* 2003, **424**:643-650.

67. • Holmes KC, Angert I, Kull FJ, Jahn W, Schroder RR: **Electron cryo-microscopy shows how strong binding of myosin to actin releases nucleotide**. *Nature* 2003, **425**:423-427.
The fitting of atomic models of actin and the myosin cross-bridge into 14 Å cryo-EM maps shows that the closing of the actin-binding cleft upon actin binding is structurally coupled to the opening of the nucleotide-binding pocket.

68. • Jiang W, Li Z, Zhang Z, Baker ML, Prevelige PE Jr, Chiu W: **Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions**. *Nat Struct Biol* 2003, **10**:131-135.
Structural analysis of bacteriophage P22 in two different functional states using cryo-EM density at subnanometer resolution shows that a large conformational change of the P22 capsid during maturation transition involves both domain movement of individual subunits and refolding of the capsid protein.

69. Zhang X, Walker SB, Chipman PR, Nibert ML, Baker TS: **Reovirus polymerase lambda 3 localized by cryo-electron microscopy of virions at a resolution of 7.6 Å**. *Nat Struct Biol* 2003, **10**:1011-1018.

70. • Zhang W, Chipman PR, Corver J, Johnson PR, Zhang Y, Mukhopadhyay S, Baker TS, Strauss JH, Rossmann MG, Kuhn RJ: **Visualization of membrane protein domains by cryo-electron microscopy of dengue virus**. *Nat Struct Biol* 2003, **10**:907-912.
Three-dimensional cryo-EM reconstruction (9.5 Å resolution) reveals secondary structural features of 180 envelope and 180 membrane proteins in the lipid envelope of mature dengue virus. This is one of only a few determinations of the disposition of transmembrane proteins *in situ* and it shows that the nucleocapsid core and envelope proteins do not directly interact in the mature virus.

71. Grunewald K, Desai P, Winkler DC, Heymann JB, Belnap DM, Baumeister W, Steven AC: **Three-dimensional structure of herpes simplex virus from cryo-electron tomography**. *Science* 2003, **302**:1396-1398.

72. Grunewald K, Medalia O, Gross A, Steven AC, Baumeister W: **Prospects of electron cryotomography to visualize macromolecular complexes inside cellular compartments: implications of crowding**. *Biophys Chem* 2003, **100**:577-591.

73. Medalia O, Weber I, Frangakis AS, Nicastro D, Gerisch G, Baumeister W: **Macromolecular architecture in eukaryotic cells**

visualized by cryoelectron tomography. *Science* 2002, **298**:1209-1213.

74. Bohm J, Frangakis AS, Hegerl R, Nickell S, Typke D, Baumeister W: **Toward detecting and identifying macromolecules in a cellular context: template matching applied to electron tomograms**. *Proc Natl Acad Sci USA* 2000, **97**:14245-14250.

75. Frangakis AS, Bohm J, Forster F, Nickell S, Nicastro D, Typke D, Hegerl R, Baumeister W: **Identification of macromolecular complexes in cryoelectron tomograms of phantom cells**. *Proc Natl Acad Sci USA* 2002, **99**:14153-14158.

76. Grimm R, Singh H, Rachel R, Typke D, Zillig W, Baumeister W: **Electron tomography of ice-embedded prokaryotic cells**. *Biophys J* 1998, **74**:1031-1042.

77. Plitzko JM, Frangakis AS, Nickell S, Forster F, Gross A, Baumeister W: ***In vivo* veritas: electron cryotomography of cells**. *Trends Biotechnol* 2002, **20**:S40-S44.

78. Stagljar I, Fields S: **Analysis of membrane protein interactions using yeast-based technologies**. *Trends Biochem Sci* 2002, **27**:559-563.

79. Burchett SA, Flanary P, Aston C, Jiang L, Young KH, Uetz P, Fields S, Dohlman HG: **Regulation of stress response signaling by the N-terminal dishevelled/EGL-10/pleckstrin domain of Sst2, a regulator of G protein signaling in *Saccharomyces cerevisiae***. *J Biol Chem* 2002, **277**:22156-22167.

80. Michnick SW: **Exploring protein interactions by interaction-induced folding of proteins from complementary peptide fragments**. *Curr Opin Struct Biol* 2001, **11**:472-477.

81. Hu CD, Kerppola TK: **Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis**. *Nat Biotechnol* 2003, **21**:539-545.

82. • Ranish JA, Yi EC, Leslie DM, Purvine SO, Goodlett DR, Eng J, Aebersold R: **The study of macromolecular complexes by quantitative proteomics**. *Nat Genet* 2003, **33**:349-355.
This paper describes a new generic strategy for determining the specific composition, changes in composition and changes in the abundance of protein complexes using mass spectrometry. Two examples are studied: the identification of genuine components of the RNA polymerase II pre-initiation complex within a high background of co-purifying proteins, and the detailed quantitative changes in the abundance and composition of immunopurified STE12 protein complexes from yeast cells exposed to different environmental conditions.

83. Himeda CL, Ranish JA, Angello JC, Maire P, Aebersold R, Hauschka SD: **Quantitative proteomic identification of six4 as the trex-binding factor in the muscle creatine kinase enhancer**. *Mol Cell Biol* 2004, **24**:2132-2143.

84. Tong AH, Drees B, Nardelli G, Bader GD, Brannetti B, Castagnoli L, Evangelista M, Ferracuti S, Nelson B, Paoluzi S *et al.*: **A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules**. *Science* 2002, **295**:321-324.

85. •• Landgraf C, Panni S, Montecchi-Palazzi L, Castagnoli L, Schneider-Mergener J, Volkmer-Engert R, Cesareni G: **Protein interaction networks by proteome peptide scanning**. *PLoS Biol* 2004, **2**:E14.
A variant of the work described in [84]. Here, candidate peptides in the yeast genome that match the consensus derived from phage display are synthesized and fixed to cellulose membranes, and then probed by SH3 domains. This reveals distinct classes of binding preferences, as well as some overlapping specificities. The strategy can be easily modified to probe the entire proteome.

86. Aloy P, Russell RB: **The third dimension for protein interactions and complexes**. *Trends Biochem Sci* 2002, **27**:633-638.

87. Cunningham BC, Jhurani P, Ng P, Wells JA: **Receptor and antibody epitopes in human growth hormone identified by homolog-scanning mutagenesis**. *Science* 1989, **243**:1330-1336.

88. Lanman J, Lam TT, Barnes S, Sakalian M, Emmett MR, Marshall AG, Prevelige PE Jr: **Identification of novel interactions in HIV-1 capsid protein assembly by high-resolution mass spectrometry**. *J Mol Biol* 2003, **325**:759-772.

89. • Anand GS, Law D, Mandell JG, Snead AN, Tsigelny I, Taylor SS, Eyck LFT, Komives EA: **Identification of the protein kinase A regulatory RIalpha-catalytic subunit interface by amide H/2H exchange and protein docking**. *Proc Natl Acad Sci USA* 2003, **100**:13264-13269.
The authors identify the interface between the two subunits (catalytic and regulatory) of protein kinase A (PKA) by computational docking and subsequent filtering of the solutions based on amide hydrogen-deuterium exchange interface protection data.

90. Guan JQ, Almo SC, Reisler E, Chance MR: **Structural reorganization of proteins revealed by radiolysis and mass spectrometry: G-actin solution structure is divalent cation dependent**. *Biochemistry* 2003, **42**:11992-12000.

91. Trester-Zedlitz M, Kamada K, Burley SK, Fenyo D, Chait BT, Muir TW: **A modular cross-linking approach for exploring protein interactions**. *J Am Chem Soc* 2003, **125**:2416-2425.

92. Back JW, de Jong L, Muijsers AO, de Koster CG: **Chemical cross-linking and mass spectrometry for protein structural modeling**. *J Mol Biol* 2003, **331**:303-313.

93. Serino G, Su H, Peng Z, Tsuge T, Wei N, Gu H, Deng XW: **Characterization of the last subunit of the *Arabidopsis* COP9 signalosome: implications for the overall structure and origin of the complex**. *Plant Cell* 2003, **15**:719-731.

94. Truong K, Ikura M: **The use of FRET imaging microscopy to detect protein-protein interactions and protein conformational changes *in vivo***. *Curr Opin Struct Biol* 2001, **11**:573-578.

95. Yan Y, Marriott G: **Analysis of protein interactions using fluorescence technologies**. *Curr Opin Chem Biol* 2003, **7**:635-640.

96. Kariakin A, Davydov D, Peterson JA, Jung C: **A new approach to the study of protein-protein interaction by FTIR: complex formation between cytochrome P450BM-3 heme domain and FMN reductase domain**. *Biochemistry* 2002, **41**:13514-13525.

97. Marquez JA, Smith CI, Petoukhov MV, Lo Surdo P, Mattsson PT, Knekt M, Westlund A, Scheffzek K, Saraste M, Svergun DI: **Conformation of full-length Bruton tyrosine kinase (Btk) from synchrotron X-ray solution scattering**. *EMBO J* 2003, **22**:4616-4624.

98. Svergun DI, Aldag I, Sieck T, Altendorf K, Koch MH, Kane DJ, Kozin MB, Gruber G: **A model of the quaternary structure of the *Escherichia coli* F1 ATPase from X-ray solution scattering and evidence for structural changes in the delta subunit during ATP hydrolysis**. *Biophys J* 1998, **75**:2212-2219.

99. Gray JJ, Moughon SE, Kortemme T, Schueler-Furman O, Misura KM, Morozov AV, Baker D: **Protein-protein docking predictions for the CAPRI experiment**. *Proteins* 2003, **52**:118-122.

100. Smith GR, Sternberg MJ: **Prediction of protein-protein interactions by docking methods**. *Curr Opin Struct Biol* 2002, **12**:28-35.

101. Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJ, Vajda S, Vakser I, Wodak SJ: **CAPRI: a Critical Assessment of PRedicted Interactions**. *Proteins* 2003, **52**:2-9.

102. Schneidman-Duhovny D, Inbar Y, Polak V, Shatsky M, Halperin I, Benyamini H, Barzilai A, Dror O, Haspel N, Nussinov R *et al.*: **Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking**. *Proteins* 2003, **52**:107-112.

103. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA: **Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques**. *Proc Natl Acad Sci USA* 1992, **89**:2195-2199.

104. Gabb HA, Jackson RM, Sternberg MJ: **Modelling protein docking using shape complementarity, electrostatics and biochemical information**. *J Mol Biol* 1997, **272**:106-120.

105. Moont G, Sternberg MJ: **Modeling protein-protein and protein-DNA docking**. Edited by Lengauer T. Weinheim: Wiley-VCH; 2001.

106. Jackson RM, Gabb HA, Sternberg MJ: **Rapid refinement of protein interfaces incorporating solvation: application to the docking problem**. *J Mol Biol* 1998, **276**:265-285.

107. Vakser IA: **Protein docking for low-resolution structures**. *Protein Eng* 1995, **8**:371-377.

108. Mandell JG, Roberts VA, Pique ME, Kotlovyi V, Mitchell JC, Nelson E, Tsigelny I, Ten Eyck LF: **Protein docking using continuum electrostatics and geometric fit**. *Protein Eng* 2001, **14**:105-113.

109. Chen R, Li L, Weng Z: **ZDOCK: an initial-stage protein-docking algorithm**. *Proteins* 2003, **52**:80-87.

110. Ritchie DW, Kemp GJ: **Protein docking using spherical polar Fourier correlations**. *Proteins* 2000, **39**:178-194.

111. Fernandez-Recio J, Totrov M, Abagyan R: **Soft protein-protein docking in internal coordinates**. *Protein Sci* 2002, **11**:280-291.

112. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D: **Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations**. *J Mol Biol* 2003, **331**:281-299.

113. Gabdoulline RR, Wade RC: **Protein-protein association: investigation of factors influencing association rates by Brownian dynamics simulations**. *J Mol Biol* 2001, **306**:1139-1155.

114. Fitzjohn PW, Bates PA: **Guided docking: first step to locate potential binding sites**. *Proteins* 2003, **52**:28-32.

115. Strynadka NC, Eisenstein M, Katchalski-Katzir E, Shoichet BK, Kuntz ID, Abagyan R, Totrov M, Janin J, Cherfils J, Zimmerman F *et al.*: **Molecular docking programs successfully predict the binding of a beta-lactamase inhibitory protein to TEM-1 beta-lactamase**. *Nat Struct Biol* 1996, **3**:233-239.

116. • Mendez R, Leplae R, De Maria L, Wodak SJ: **Assessment of blind predictions of protein-protein interactions: current status of docking methods**. *Proteins* 2003, **52**:51-67.
This paper reviews the results from the first two rounds of the CAPRI community-wide docking experiment, in which 19 groups attempted to predict the structures of seven protein–protein complexes. The targets, as well as the docking methods and protocols used in the predictions, are described. In total, five of the seven target complexes were predicted to an acceptable accuracy.

117. Vajda S, Camacho CJ: **Protein-protein docking: is the glass half-full or half-empty?** *Trends Biotechnol* 2004, **22**:110-116.

118. Tovchigrechko A, Wells CA, Vakser IA: **Docking of protein models**. *Protein Sci* 2002, **11**:1888-1896.

119. Morillas M, Gomez-Puertas P, Rubi B, Clotet J, Arino J, Valencia A, Hegardt FG, Serra D, Asins G: **Structural model of a malonyl-CoA-binding site of carnitine octanoyltransferase and carnitine palmitoyltransferase I: mutational analysis of a malonyl-CoA affinity domain**. *J Biol Chem* 2002, **277**:11473-11480.

120. • Dobrodumov A, Gronenborn AM: **Filtering and selection of structural models: combining docking and NMR**. *Proteins* 2003, **53**:18-32.
The authors demonstrate an approach to protein complex structure determination that first generates structural models by computational docking of the subunits and then filters these using experimental NMR constraints (residual dipolar couplings and chemical shift mapping).

121. Dominguez C, Boelens R, Bonvin AM: **HADDOCK: a protein-protein docking approach based on biochemical or biophysical information**. *J Am Chem Soc* 2003, **125**:1731-1737.

122. Romijn RA, Westein E, Bouma B, Schiphorst ME, Sixma JJ, Lenting PJ, Huizinga EG: **Mapping the collagen-binding site in the von Willebrand factor-A3 domain**. *J Biol Chem* 2003, **278**:15035-15039.

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

124. •• Aloy P, Bottcher B, Ceulemans H, Leutwein C, Mellwig C, Fischer S, Gavin AC, Bork P, Superti-Furga G, Serrano L *et al.*: **Structure-based assembly of protein complexes in yeast**. *Science* 2004, **303**:2026-2029.
Large-scale structural analysis of complexes in yeast using bioinformatics and EM. As complete models as possible were built for over 100 yeast complexes and all complex–complex interactions. For some, EM maps could be used to confirm models or combine separately modeled subcomplexes.

125. Pieper U, Eswar N, Braberg H, Madhusudhan MS, Davis FP, Stuart AC, Mirkovic N, Rossi A, Marti-Renom MA, Fiser A *et al.*: **MODBASE, a database of annotated comparative protein structure models, and associated resources**. *Nucleic Acids Res* 2004, **32**:D217-D222.

126. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A: **Comparative protein structure modeling of genes and genomes**. *Annu Rev Biophys Biomol Struct* 2000, **29**:291-325.

127. • Aloy P, Ceulemans H, Stark A, Russell RB: **The relationship between sequence and interaction divergence in proteins**. *J Mol Biol* 2003, **332**:989-998.
This study finds that the structural binding modes of proteins are conserved at around 30% sequence identity. Below this cutoff, the binding mode may or may not be similar.

128. •• Kortemme T, Joachimiak LA, Bullock AN, Schuler AD, Stoddard BL, Baker D: **Computational redesign of protein-protein interaction specificity**. *Nat Struct Mol Biol* 2004, **11**:371-379.
The authors develop a computational strategy to redesign specificity at protein–protein interfaces and apply it to create new specifically interacting DNase–inhibitor protein pairs. In addition to *in vitro* and *in vivo* biochemical confirmation of the specificities, a crystal structure of one of the pairs confirms the computationally designed interface structure to an impressive 0.62 Å all-atom rmsd.

129. Ornitz DM, Xu J, Colvin JS, McEwen DG, MacArthur CA, Coulier F, Gao G, Goldfarb M: **Receptor specificity of the fibroblast growth factor family**. *J Biol Chem* 1996, **271**:15292-15297.

130. Aloy P, Russell RB: **InterPreTS: protein interaction prediction through tertiary structure**. *Bioinformatics* 2003, **19**:161-162.

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

132. •• Lu L, Arakaki AK, Lu H, Skolnick J: **Multimeric threading-based prediction of protein-protein interactions on a genomic scale: application to the *Saccharomyces cerevisiae* proteome**. *Genome Res* 2003 June, **13**:1146-1154.
The authors predict the protein interaction network in *S. cerevisiae* by applying their MULTIPROSPECTOR threading algorithm to the entire proteome.

133. Valpuesta JM, Martin-Benito J, Gomez-Puertas P, Carrascosa JL, Willison KR: **Structure and function of a protein folding machine: the eukaryotic cytosolic chaperonin CCT**. *FEBS Lett* 2002, **529**:11-16.

134. Liou AK, Willison KR: **Elucidation of the subunit orientation in CCT (chaperonin containing TCP1) from the subunit composition of CCT micro-complexes**. *EMBO J* 1997, **16**:4311-4316.

135. Raijmakers R, Noordman YE, van Venrooij WJ, Pruijn GJ: **Protein-protein interactions of hCsl4p with other human exosome subunits**. *J Mol Biol* 2002, **315**:809-818.

136. Aloy P, Ciccarelli FD, Leutwein C, Gavin AC, Superti-Furga G, Bork P, Boettcher B, Russell RB: **A complex prediction: three-dimensional model of the yeast exosome**. *EMBO Rep* 2002, **3**:628-635.

137. Zarrinpar A, Park SH, Lim WA: **Optimization of specificity in a cellular protein interaction network by negative selection**. *Nature* 2003, **426**:676-680.

138. Ma B, Elkayam T, Wolfson H, Nussinov R: **Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces**. *Proc Natl Acad Sci USA* 2003, **100**:5772-5777.

139. Yao H, Kristensen DM, Mihalek I, Sowa ME, Shaw C, Kimmel M, Kavraki L, Lichtarge O: **An accurate, sensitive, and scalable method to identify functional sites in protein structures**. *J Mol Biol* 2003, **326**:255-261.

140. Fariselli P, Pazos F, Valencia A, Casadio R: **Prediction of protein–protein interaction sites in heterocomplexes with neural networks**. *Eur J Biochem* 2002, **269**:1356-1361.

141. del Sol Mesa A, Pazos F, Valencia A: **Automatic methods for predicting functionally important residues**. *J Mol Biol* 2003, **326**:1289-1302.

142. Lichtarge O, Bourne HR, Cohen FE: **An evolutionary trace method defines binding surfaces common to protein families**. *J Mol Biol* 1996, **257**:342-358.

143. Hannenhalli SS, Russell RB: **Analysis and prediction of functional sub-types from protein sequence alignments**. *J Mol Biol* 2000, **303**:61-76.

144. Aloy P, Querol E, Aviles FX, Sternberg MJE: **Automated structure-based prediction of functional sites in proteins: applications to assessing the validity of inheriting protein function from homology in genome annotation and to protein docking**. *J Mol Biol* 2001, **311**:395-408.

145. Landgraf R, Xenarios I, Eisenberg D: **Three-dimensional cluster analysis identifies interfaces and functional residue clusters in proteins**. *J Mol Biol* 2001, **307**:1487-1502.

146. Kortemme T, Baker D: **A simple physical model for binding energy hot spots in protein-protein complexes**. *Proc Natl Acad Sci USA* 2002, **99**:14116-14121.

147. Boulanger MJ, Bankovich AJ, Kortemme T, Baker D, Garcia KC: **Convergent mechanisms for recognition of divergent cytokines by the shared signaling receptor gp130**. *Mol Cell* 2003, **12**:577-589.

148. McFarland BJ, Kortemme T, Yu SF, Baker D, Strong RK: **Symmetry recognizing asymmetry: analysis of the interactions between the C-type lectin-like immunoreceptor NKG2D and MHC class I-like ligands**. *Structure* 2003, **11**:411-422.

149. Volkmann N, Hanein D, Ouyang G, Trybus KM, DeRosier DJ, Lowey S: **Evidence for cleft closure in actomyosin upon ADP release**. *Nat Struct Biol* 2000, **7**:1147-1155.

150. Roseman AM: **Docking structures of domains into maps from cryo-electron microscopy using local correlation**. *Acta Crystallogr D Biol Crystallogr* 2000, **56**:1332-1340.

151. Wriggers W, Birmanns S: **Using situs for flexible and rigid-body fitting of multiresolution single-molecule data**. *J Struct Biol* 2001, **133**:193-202.

152. Ceulemans H, Russell RB: **Fast fitting of atomic structures to lower resolution electron density maps by surface overlap maximization**. *J Mol Biol* 2004, **338**:783-793.

153. Volkmann N, Hanein D: **Quantitative fitting of atomic models into observed densities derived by electron microscopy**. *J Struct Biol* 1999, **125**:176-184.

154. Rossmann MG, Bernal R, Pletnev SV: **Combining electron microscopic with x-ray crystallographic structures**. *J Struct Biol* 2001, **136**:190-200.

155. Chiu W, Baker ML, Jiang W, Zhou ZH: **Deriving folds of macromolecular complexes through electron cryomicroscopy and bioinformatics approaches**. *Curr Opin Struct Biol* 2002, **12**:263-269.

156. Chacon P, Wriggers W: **Multi-resolution contour-based fitting of macromolecular structures**. *J Mol Biol* 2002, **317**:375-384.

157. Volkmann N, Ouyang G, Trybus KM, DeRosier DJ, Lowey S, Hanein D: **Myosin isoforms show unique conformations in the actin-bound state**. *Proc Natl Acad Sci USA* 2003, **100**:3227-3232.

158. Spahn CM, Beckmann R, Eswar N, Penczek PA, Sali A, Blobel G, Frank J: **Structure of the 80S ribosome from *Saccharomyces cerevisiae*–tRNA- ribosome and subunit-subunit interactions**. *Cell* 2001, **107**:373-386.

159. Beckmann R, Spahn CM, Eswar N, Helmers J, Penczek PA, Sali A, Frank J, Blobel G: **Architecture of the protein-conducting channel associated with the translating 80S ribosome**. *Cell* 2001, **107**:361-372.

160. •• Kostyuchenko VA, Leiman PG, Chipman PR, Kanamaru S, van Raaij MJ, Arisaka F, Mesyanzhinov VV, Rossmann MG: **Three-dimensional structure of bacteriophage T4 baseplate**. *Nat Struct Biol* 2003, **10**:688-693.

Three-dimensional cryo-EM reconstruction of the baseplate–tail tube complex of bacteriophage T4 at 12 Å resolution. The X-ray structures of six proteins were fitted into the electron density and the location of four others was determined with the help of biochemical data. The overall structure suggests a mechanism of baseplate triggering and structural transition during the initial stages of T4 infection.

161. Golas MM, Sander B, Will CL, Luhrmann R, Stark H: **Molecular architecture of the multiprotein splicing factor SF3b**. *Science* 2003, **300**:980-984.

162. • Shin DS, Pellegrini L, Daniels DS, Yelent B, Craig L, Bates D, Yu DS, Shivji MK, Hitomi C, Arvai AS *et al.*: **Full-length archaeal Rad51 structure and mutants: mechanisms for RAD51 assembly and control by BRCA2**. *EMBO J* 2003, **22**:4566-4576.

The crystal structure of full-length RAD51 (2.85 Å resolution), in combination with SAXS, cryo-EM and mutagenesis, was used to characterize a macromolecular machine with a central role in repairing DNA double-strand breaks by homologous recombination. The study suggests how BRC repeats may disrupt RAD51 ring assembly and direct RAD51 to form fibers in response to cellular DNA damage.

163. Zhou ZH, Baker ML, Jiang W, Dougherty M, Jakana J, Dong G, Lu G, Chiu W: **Electron cryomicroscopy and bioinformatics suggest protein fold models for rice dwarf virus**. *Nat Struct Biol* 2001, **8**:868-873.

164. Baker ML, Jiang W, Bowman BR, Zhou ZH, Quiocho FA, Rixon FJ, Chiu W: **Architecture of the herpes simplex virus major capsid protein derived from structural bioinformatics**. *J Mol Biol* 2003, **331**:447-456.

165. Blundell TL, Sibanda BL, Sternberg MJ, Thornton JM: **Knowledge-based prediction of protein structures and the design of novel molecules**. *Nature* 1987, **326**:347-352.

166. Greer J: **Comparative modeling methods: application to the family of the mammalian serine proteases**. *Proteins* 1990, **7**:317-334.

167. Sali A, Blundell TL: **Comparative protein modelling by satisfaction of spatial restraints**. *J Mol Biol* 1993, **234**:779-815.

168. Sauder JM, Dunbrack RL Jr: **Genomic fold assignment and rational modeling of proteins of biological interest**. *Proc Int Conf Intell Syst Mol Biol* 2000, **8**:296-306.

169. Burley SK, Almo SC, Bonanno JB, Capel M, Chance MR, Gaasterland T, Lin D, Sali A, Studier FW, Swaminathan S: **Structural genomics: beyond the human genome project**. *Nat Genet* 1999, **23**:151-157.

170. Vitkup D, Melamud E, Moult J, Sander C: **Completeness in structural genomics**. *Nat Struct Biol* 2001, **8**:559-566.

171. Baker D, Sali A: **Protein structure prediction and structural genomics**. *Science* 2001, **294**:93-96.

172. Apic G, Gough J, Teichmann SA: **Domain combinations in archaeal, eubacterial and eukaryotic proteomes**. *J Mol Biol* 2001, **310**:311-325.

173. Sali A: **NIH workshop on structural proteomics of biological complexes**. *Structure* 2003, **11**:1043-1047.

174. Malhotra A, Tan RK, Harvey SC: **Prediction of the three-dimensional structure of *Escherichia coli* 30S ribosomal subunit: a molecular mechanics approach**. *Proc Natl Acad Sci USA* 1990, **87**:1950-1954.

175. Gomez-Lorenzo MG, Valle M, Frank J, Gruss C, Sorzano CO, Chen XS, Donate LE, Carazo JM: **Large T antigen on the simian virus 40 origin of replication: a 3D snapshot prior to DNA replication**. *EMBO J* 2003, **22**:6205-6213.

176. Li D, Zhao R, Lilyestrom W, Gai D, Zhang R, DeCaprio JA, Fanning E, Jochimiak A, Szakonyi G, Chen XS: **Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen**. *Nature* 2003, **423**:512-518.

177. Keskin O, Bahar I, Flatow D, Covell DG, Jernigan RL: **Molecular mechanisms of chaperonin GroEL-GroES function**. *Biochemistry* 2002, **41**:491-501.

178. Keskin O, Durell SR, Bahar I, Jernigan RL, Covell DG: **Relating molecular flexibility to function: a case study of tubulin**. *Biophys J* 2002, **83**:663-680.

179. Ming D, Kong Y, Wakil SJ, Brink J, Ma J: **Domain movements in human fatty acid synthase by quantized elastic deformational model**. *Proc Natl Acad Sci USA* 2002, **99**:7895-7899.

180. Tama F, Wriggers W, Brooks CL III: **Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory**. *J Mol Biol* 2002, **321**:297-305.

181. Chacon P, Tama F, Wriggers W: **Mega-Dalton biomolecular motion captured from electron microscopy reconstructions**. *J Mol Biol* 2003, **326**:485-492.

182. Kong Y, Ming D, Wu Y, Stoops JK, Zhou ZH, Ma J: **Conformational flexibility of pyruvate dehydrogenase complexes: a computational analysis by quantized elastic deformational model**. *J Mol Biol* 2003, **330**:129-135.

183. Symmons MF, Jones GH, Luisi BF: **A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation**. *Structure Fold Des* 2000, **8**:1215-1226.

184. Selenko P, Sprangers R, Stier G, Buhler D, Fischer U, Sattler M: **SMN tudor domain structure and its interaction with the Sm proteins**. *Nat Struct Biol* 2001, **8**:27-31.

185. Kronenberg S, Kleinschmidt JA, Bottcher B: **Electron cryo-microscopy and image reconstruction of adeno-associated virus type 2 empty capsids**. *EMBO Rep* 2001, **2**:997-1002.

186. Uetz P: **Two-hybrid arrays**. *Curr Opin Chem Biol* 2002, **6**:57-62.

187. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y: **A comprehensive two-hybrid analysis to explore the yeast protein interactome**. *Proc Natl Acad Sci USA* 2001, **98**:4569-4574.