The molecular sociology of the cell

Carol V. Robinson[1], Andrej Sali[2] & Wolfgang Baumeister[3]

Proteomic studies have yielded detailed lists of the proteins present in a cell. Comparatively little is known, however, about how these proteins interact and are spatially arranged within the 'functional modules' of the cell: that is, the 'molecular sociology' of the cell. This gap is now being bridged by using emerging experimental techniques, such as mass spectrometry of complexes and single-particle cryo-electron microscopy, to complement traditional biochemical and biophysical methods. With the development of integrative computational methods to exploit the data obtained, such hybrid approaches will uncover the molecular architectures, and perhaps even atomic models, of many protein complexes. With these structures in hand, researchers will be poised to use cryo-electron tomography to view protein complexes in action within cells, providing unprecedented insights into protein-interaction networks.

A cell consists of hundreds of different functional modules, such as the RNA exosome, the proteasome and the nuclear pore complex (NPC). These modules, in turn, are composed of macromolecules, such as proteins and nucleic acids, as well as various small molecules. 'Molecular sociology' refers to the interactions of molecules within these functional modules.

At one end of the scale, there are highly stable interactions that are robust enough to withstand the rigours of purification. A large proportion of these stable structures are likely to be solved. The preferred method for determining the structures of assemblies at atomic resolution is X-ray crystallography[1]. Crystallography, however, is suitable only for functional modules that can be reconstituted *in vitro* and purified in sufficient quantity for crystallization. A landmark in structural biology occurred in 2000, when atomic structures of a large functional module — the ribosome from extremophile bacteria — were solved[2–4]. Progress has since been made towards determining the structures of similarly large complexes; however, in the past decade, there has not been a marked increase in the molecular mass of asymmetrical complexes that can be studied by crystallography.

At the other end of the scale, there are interactions that occur more fleetingly, in response to intracellular signalling, for example. The potential for determining the structures of such transient complexes by using any type of crystallography is relatively poor. For these complexes, as well as for stable complexes that are refractory to structure determination by traditional methods, integrative approaches are required[5–8]. These approaches combine information from varied sources. For example, individual subunits can be assembled into the whole complex by molecular docking that is restrained by knowledge of structurally defined homologous interactions, direct contact information provided by mass spectrometry[9] and other data[10,11]. Such approaches have been aided greatly by the availability of high-resolution structures of individual subunits from high-throughput structural-genomics consortia[12], and they are enabling the generation of atomic models and architectural models (in which the location and orientation of subunits within an assembly are defined) of previously intractable assemblies[6,9,13,14]. These models provide a basis for the development of testable hypotheses that could not be envisaged without a structural model. A spectacular

example of the use of a hybrid approach[15] is the molecular model of auxilin bound to clathrin (the main component of the coat of coated vesicles), which was obtained by fitting comparative protein-structure models of the components into a cryo-electron-microscopy map at 12 Å resolution[16] (Fig. 1). Difference mapping showed changes in the
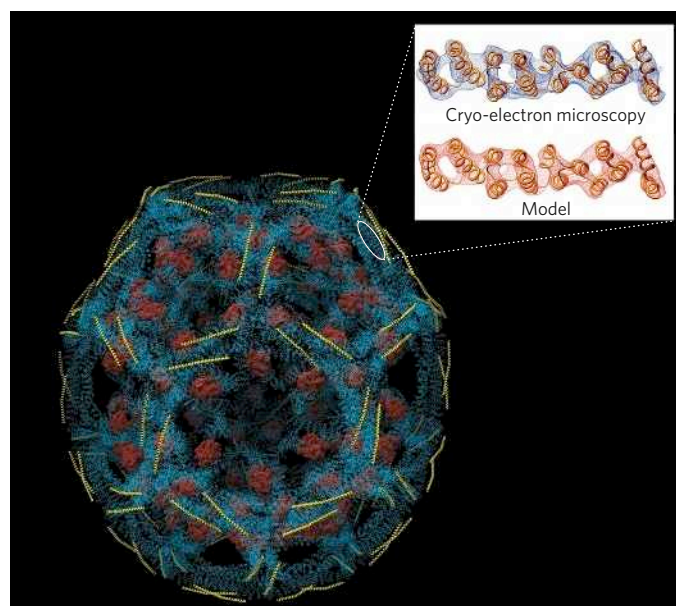

Figure 1 | A polypeptide-chain model for a clathrin D6 barrel. An α-carbon trace of the clathrin heavy (blue) and light (yellow) chains, derived by fitting atomic homology-based models into the density map from an 8 Å-resolution cryo-electron-microscopy reconstruction[16]. The position of a bound auxilin fragment (residues 547–910; red) was determined from a 12 Å-resolution cryo-electron-microscopy difference map. The inset zooms in to illustrate how closely the α-carbon coordinates of part of the heavy chain, as shown in the main figure (inset, lower), fit within the cryo-electron-microscopy density map (inset, upper). (Image reproduced, with permission, from ref. 16.)

[1]Department of Chemistry, Lensfield Road, University of Cambridge, Cambridge CB2 1EW, UK. [2]Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences, Byers Hall, Suite 503B, University of California at San Francisco, 1700 4th Street, San Francisco, California 94158, USA. [3]Max Planck Institute of Biochemistry, Department of Molecular Structural Biology, Am Klopferspitz 18, D-82152 Martinsried, Germany.

973

clathrin lattice when auxilin is bound, prompting the hypothesis that local destabilization of the lattice promotes uncoating of the membranes of coated vesicles.

To illustrate the emergence of integrative approaches to structure determination, we have chosen a series of molecular 'machines' with differences in molecular mass, robustness and abundance: from the comparatively moderate dimensions of the yeast RNA exosome (400 kDa)[17] to the 26S proteasome (2.5 MDa)[18] and culminating in the NPC (50–100 MDa)[19]. For the yeast RNA exosome, which is relatively robust, atomic models were constructed by using spatial restraints from mass spectrometry[9] as a guide for the computational docking of subunit comparative models. By contrast, the heterogeneity and lability of the 26S proteasome have so far made it impossible to obtain a high-resolution model. The low resolution of the cryo-electron-microscopy map and the absence of high-resolution structures of many of the components — with the notable exception of the 20S core — have precluded the use of hybrid approaches to generate an atomic-resolution model. However, there are valuable data on binary interactions between the components, obtained from the yeast two-hybrid system and from mass spectrometry, and these need to be integrated with the cryo-electron-microscopy map. This example highlights the difficulties in applying integrative approaches to less-robust protein complexes. For the NPC, the highest-resolution *in situ* characterization was achieved recently by

using cryo-electron tomography[20]. Moreover, it has also been possible to determine the configuration of the constituent proteins from a variety of proteomic and biophysical data[21]. Before presenting these examples, we consider the biophysical methods that can provide structural information about macromolecular assemblies.

Experimental methods for structure determination

Structures can be described at different levels of resolution. At the lowest level, the configuration of the components specifies the relative positions and interactions of the macromolecules. A higher-resolution description defines the molecular architecture, including the relative orientations of the components. For pseudo-atomic models, the positions of the atoms are specified but with errors larger than the size of an atom. The highest level of resolution is an atomic structure, which shows atomic positions with a precision smaller than the size of an atom.

Different experimental methods reveal different information about protein complexes. The stoichiometry and composition of an assembly, for example, can be determined by methods such as quantitative immunoblotting and mass spectrometry. The shape of the assembly can be revealed by cryo-electron microscopy and small-angle X-ray scattering (SAXS). In addition, cryo-electron microscopy can be used to determine the positions of the components, as can labelling techniques.
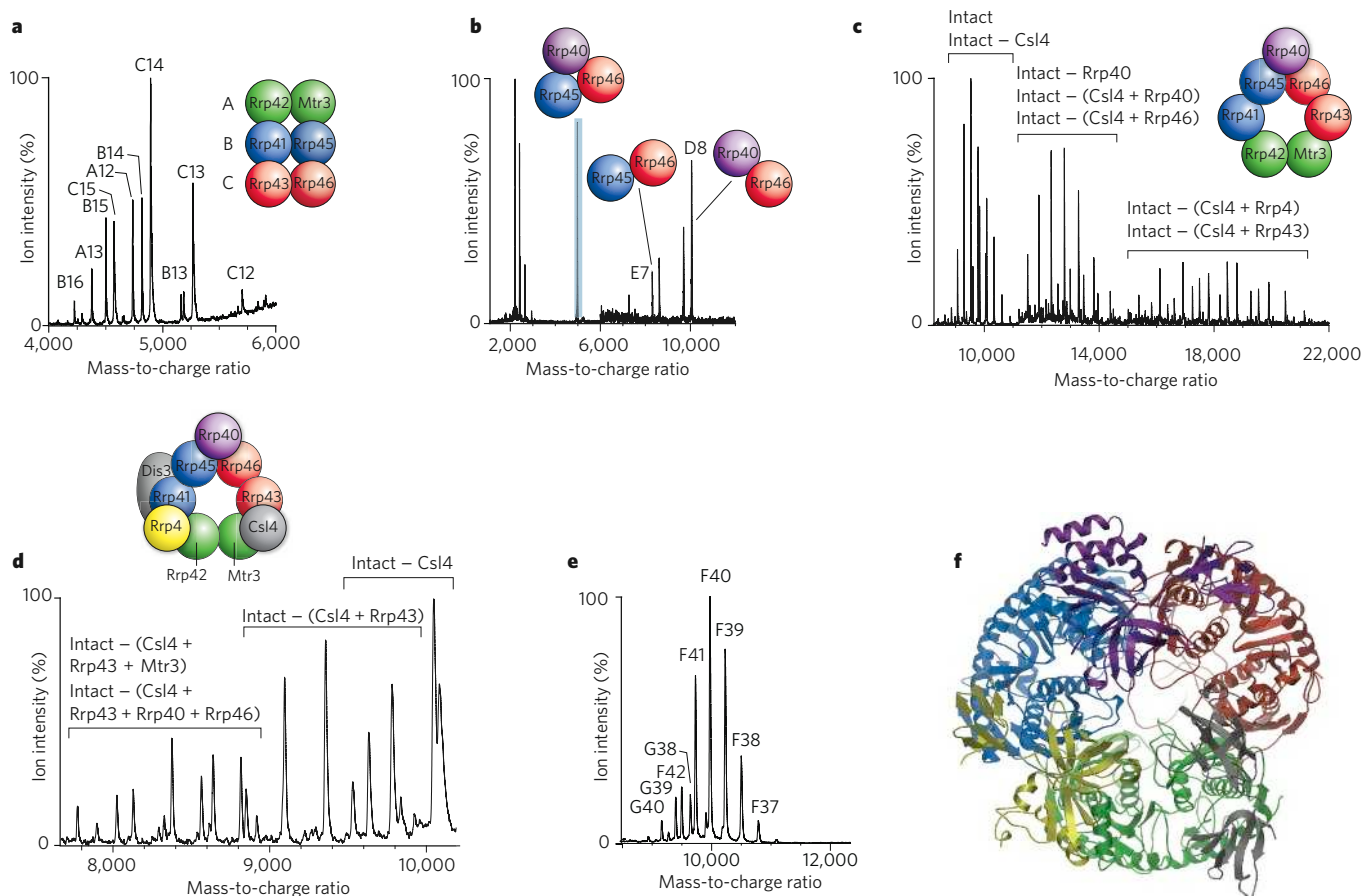

Figure 2 | Determining an atomic model of the yeast RNA exosome, by using mass spectrometry and comparative protein-structure modelling.
The figure shows a series of five mass spectra, recorded under different conditions, revealing the building blocks from which the overall structure was constructed. **a,** Intact RNA exosomes were isolated from yeast and partially denatured. Mass spectrometry showed the presence of three heterodimers (A, B and C), as determined from the mass-to-charge ratio of each peak. (The number of positive charges corresponding to each dimer is indicated; for example, the largest peak represents the heterodimer C with 14 positive charges.) **b,** After tandem mass spectrometry (see page 991) of a low-abundance complex, highlighted in blue, that was present in the solution of the intact complex (not visible in **a**), a heterotrimer was

identified that contained two of the subunits observed in (**a**) plus an additional subunit, Rrp40, enabling dimers B and C to be oriented within the ring. **c**, **d**, Using acceleration in the gas phase (**c**) and generation of complexes in solution (**d**), a series of related subcomplexes was produced, enabling the remaining subunits to be arranged in the ring, bridging subunits to be placed between the heterodimers, and the largest subunit, Dis3, to be located on the base of the complex. **e**, The intact complex confirms the single copy number of all ten subunits (F), with a small population of the complex having lost Csl4 during isolation (G). **f**, Comparative modelling was then used to produce an atomic model; the ribbons are depicted in colours corresponding to those in **a**–**d**. (Figure adapted, with permission, from ref. 9.)

Information can also be gained about whether particular components interact with each other, by using mass spectrometry, yeast two-hybrid experiments or affinity purification. Further information about interacting residues, as well as about the relative orientations of the components, can be inferred from cryo-electron microscopy, hydrogen–deuterium exchange, hydroxyl-radical footprinting and chemical crosslinking. At the highest resolution, information about the atomic structures of components and their interactions can be determined by X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. We outline some of these experimental methods in this section, and we highlight mass spectrometry and cryo-electron microscopy in the Boxes.

X-ray crystallography

The 'gold standard' for the structural analysis of proteins and protein complexes in terms of accuracy and resolution is X-ray crystallography[1]. Using this method, the amplitudes, and sometimes the phases, of structure factors in a crystal sample are measured. Together with a molecular-mechanics force field, this information is used in an optimization process that can result in an atomic structure of the assembly. In addition to the ribosome[2–4], X-ray crystallography has recently been used to solve structures of many macromolecular assemblies that involve protein–protein, protein–RNA and protein–DNA interactions, such as RNA polymerase[22], the RNA exosome[23] and the signal-recognition-particle complex[24].

NMR spectroscopy

NMR spectroscopy is increasingly used to determine which surfaces of components in a protein complex are interacting[25] (from chemical-shift perturbations[26] and residual dipolar coupling[27]), in addition to the structures of the individual protein components[28,29]. Such information can be combined with computational docking to obtain approximate structures of protein complexes[24]. A key attribute of NMR spectroscopy is that it allows the determination of atomic structures of complexes in solution in near-native conditions.

SAXS

Another method that enables structures to be determined in solution is SAXS. The data can be converted into a radial distribution function that provides low-resolution information about the shape of an assembly[30]. One of the advantages of SAXS is that it is suitable for assemblies of 50–250 kDa, which cannot easily be examined by cryo-electron microscopy or NMR spectroscopy. In addition, the ease of altering the solution conditions in which the sample is studied makes SAXS ideal for mapping differences between the conformational states of an assembly. The recent renaissance of SAXS largely results from efforts to integrate SAXS data with other structural information from complementary sources[31]. For example, the data obtained from SAXS studies of proteins or their complexes can be considered simultaneously with corresponding cryo-electron-microscopy maps[32]. SAXS spectra have also been incorporated into a protocol for structure determination by NMR spectroscopy[33]. Because SAXS data contain global information about the protein that is complementary to the short-range restraints from NMR spectroscopy, models of multidomain proteins are much more accurate than models based on NMR spectra alone. Examples of quaternary atomic structures obtained by using SAXS in conjunction with atomic structures of the protein components are calcium/calmodulin-dependent protein kinase II (ref. 34), the Ras activator son of sevenless (SOS)[35] and the various nucleotide-bound conformations of the ATPase GspE[36].

Labelling techniques

The approximate positions of protein components in an assembly can be determined by labelling techniques[37]. The protein component of interest is tagged with a probe, which can then be detected, for example by cryo-electron microscopy. The choice of labels depends on the known properties of the protein. For example, immuno-electron microscopy can be used to study proteins labelled with an antibody,

---

> ### Box 1 | Mass spectrometry and hybrid approaches
>
> Mass spectrometry has underpinned proteomics for many years, but recent developments have led to its integration into the structural-biologist's toolkit.
>
> **Determining the composition and stoichiometry of a complex**
> Analysis of intact complexes by mass spectrometry often requires modification of a conventional mass spectrometer, but it can reveal the stoichiometry and copy number of many protein complexes[58,82], from homomers (which consist of multiple copies of the same protein)[57] to complex heterogeneous structures such as ribosomes[83].
>
> When using mass spectrometry as part of a hybrid approach, the first step (after the complex has been isolated by affinity purification) is a traditional proteomics experiment[84]. The component proteins of the complex are separated using one-dimensional SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and are subjected to trypsin digestion. Mass spectra are then recorded for the resultant peptides. This first step, coupled with database searching to identify the peptides and therefore the subunits, provides an inventory of all components of the complex.
>
> With the identity of the components established, the masses of the intact components can then be determined, to define post-translational modifications. This is achieved by using a denaturing step, typically incubation in a low-pH solution, to disassemble the complex — sometimes after chromatographic separation[45,85].
>
> The next step is to generate subcomplexes by perturbing the complex in solution[9]. When the masses of the identified subunits and subcomplexes have been determined, a mass spectrum of the intact complex is used to define the overall copy number of all components in the complex. Computational analysis is then used to generate an interaction network[9]. The connections between each component are weighted according to the number of times each interaction is found in the population of networks that satisfy all of the nearest-neighbour restraints imposed by the subcomplexes.
>
> **Defining distance restraints by mass spectrometry**
> If two interacting proteins can be chemically crosslinked with molecular tethers, then this implies a distance restraint on the tethered residues[86]. Coupling this technique with mass spectrometry is appealing intellectually, but successes have so far been limited to a few small protein complexes[87]. Another way to define distance restraints is ion-mobility mass spectrometry[88]. This technique has only recently been applied to protein complexes, and, in these examples, measurement of collision cross-section has been used to examine complexes in the context of their X-ray structures[89]. One of the difficulties encountered when applying this approach is that protein complexes unfold when activated in the gas phase[90]. To restrain interaction models using collision cross-sections, it is necessary to establish conditions to maintain complexes as close to their native topology as possible. Although still in its infancy, ion-mobility mass spectrometry holds great promise for generating key structural information for the modelling of macromolecular assemblies.

---

which is typically conjugated to nanometre-sized gold beads to facilitate visualization[37]. Another option is to label protein components with histidine tags, which can be detected by using nickel-nitrilotriacetic acid (NiNTA)-conjugated gold beads[38]. Alternatively, proteins can be identified by exposing them to interacting proteins that have been covalently bound to gold beads[20].

Biochemical and biophysical methods

Information about the relative position, as well as the relative orientation, of the components in a complex can be gained from biochemical and biophysical methods. Site-directed mutagenesis, for example, can identify the amino-acid residues that mediate an interaction[14]. Approaching the same problem from a different angle, chemical footprinting[39] and hydrogen–deuterium exchange[40] can identify the surfaces that are buried when a complex forms[41]. Structural information

can also be obtained by measuring the proximities of labelled groups on interacting proteins, using fluorescence resonance energy transfer (FRET) spectroscopy[42]. For example, the protein organization of the spindle pole body in yeast cells was established largely from distances obtained in FRET experiments[43].

Proteomics experiments

Proteomics experiments are generating large amounts of data that provide information about the molecular architectures of functional modules[6,7,43–45]. Information about binary interactions between proteins can be gained by using various techniques: yeast two-hybrid experiments[46,47], protein-fragment complementation assays[48], a combination of phage display and other techniques[49], protein arrays[50], and solid-phase detection by using surface plasmon resonance[51]. Physical interactions between proteins have also been inferred from genetic interactions, through the reduced activity or viability of mutant yeast strains in which genes encoding both proteins have been knocked out[52]. Furthermore, affinity purification[53,54] can be used to characterize not only binary interactions but also higher-order interactions, by purifying protein complexes and then identifying their components by mass spectrometry[55]; proximity between the identified components is established because they are directly or indirectly associated with the same tagged 'bait' protein.

Integration of structural information from different sources

After structural data have been obtained by one or more of these experimental methods, they need to be converted into a structural model through computation. As mentioned earlier, when approaches dominated by a single source of information fail, a hybrid approach, in which all of the available information about the composition and the structure of a given assembly is simultaneously considered (irrespective of the source), can sometimes be sufficient to calculate a useful structural model[5,6,8]. Even when this model is of relatively low resolution and accuracy, it can still be helpful for studying the function and evolution of the assembly; it also provides the necessary starting point for a study

at higher resolution. An example of a simple hybrid approach is building a pseudo-atomic model of a large assembly by fitting the atomic structures of the subunits into the cryo-electron-microscopy map of the assembly[15,56,57]. In this section, we present three hybrid approaches, which were successfully applied to solve the structures of the RNA exosome, the 26S proteasome and the NPC.

One of the main difficulties encountered when structurally characterizing assemblies is the absence of information about direct contacts between subunits. Direct contacts can be identified by partial disruption of an assembly to yield a series of subcomplexes, followed by tandem mass spectrometry (which allows further disruption of a selected region of the mass spectrum) to determine the stoichiometry and the contacts between the components[58]. When enough subcomplexes have been characterized, an unequivocal protein–protein interaction network can be generated for the whole complex[7,9,45]. Such an approach has been applied to the yeast RNA exosome, which has ten subunits.

An atomic model of an RNA exosome

Despite its small size, attempts to analyse the eukaryotic RNA exosome by using X-ray crystallography have been repeatedly unsuccessful. Interesting structural insights have been gained, however, by overexpressing subunits of RNA exosomes from Archaea[59,60]. Moreover, a hybrid approach to studying the yeast RNA exosome has to some extent circumvented the challenges presented by crystallography. The yeast RNA exosome is present in both the nucleus and the cytoplasm, and is involved in RNA processing and turnover[17]. To obtain an architectural model of the yeast complex, the cytoplasmic form of the intact complex was isolated by tandem affinity purification[9]. Using partial denaturing agents, subcomplexes were generated and, after confirmation by tandem mass spectrometry, a protein–protein interaction network for the complex was determined (Fig. 2; Box 1). A key step in assembling the architectural model was the identification of three pairs of heterodimers that constitute a six-membered ring, a structure that had been observed in low-resolution electron-microscopy maps[61]. Experimental data also showed that several proteins — Csl4, Rrp4 and Rrp40 — bind to and
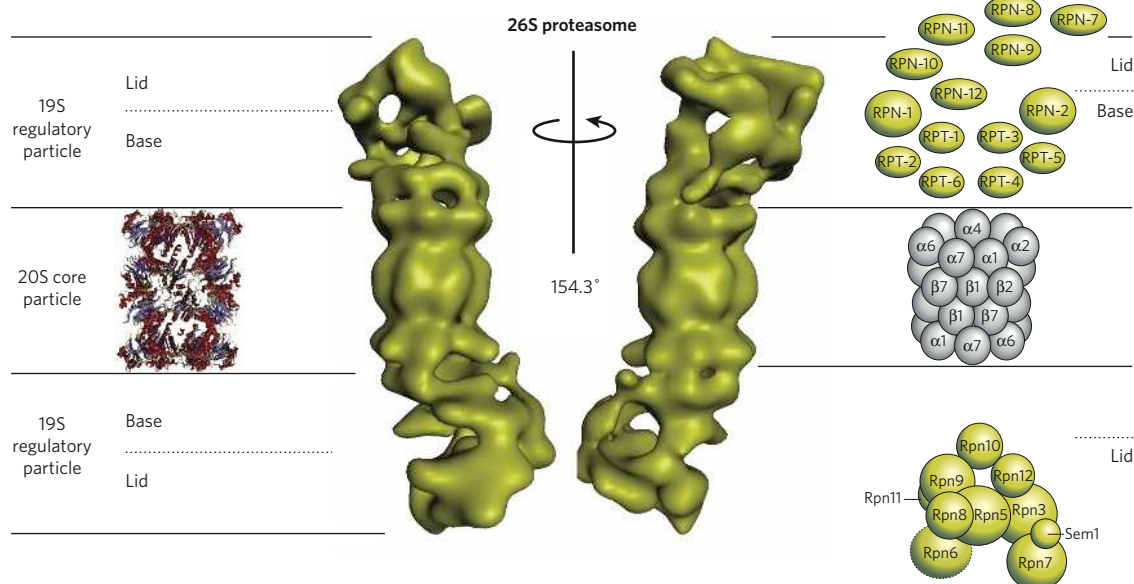

Figure 3 | The molecular architecture of the 26S proteasome. The 26S proteasome consists of 19S regulatory particles associated with the ends of a barrel-shaped 20S core particle. The part of each 19S regulatory subunit that is closest to the core is known as the base, and the part that is farthest away is known as the lid. Crystal structures have been obtained for archaeal, bacterial and eukaryotic 20S core particles[63,77–79] (left, α-helices in red, and β-sheets in blue). For the eukaryotic 26S holocomplex, only a low-resolution structure, obtained by cryo-electron microscopy[67], is available (centre; two orientations, rotated by 154.3°). Topological models of the regulatory particle have been deduced from yeast two-hybrid screens of *Caenorhabditis elegans* proteins[68] (upper right) and from mass spectrometry of yeast proteins[45] (lower right). These models agree reasonably well, albeit not completely. A topological model of the 20S core (centre right) that corresponds to the crystal structure (left) is also shown. No attempt has yet been made to obtain the molecular architecture of the entire 26S proteasome by integrating these topological models with the cryo-electron-microscopy map. RPN, non-ATPase subunit; RPT, ATPase subunit. (Central image reproduced, with permission, from ref. 65.)

strengthen the interfaces between the heterodimers, so these 'bridging' subunits were placed in the ring accordingly[9]. Given the similarity between the subunits in RNA exosomes from different species, models of the yeast proteins were then superimposed on the related archaeal ring structure[59]. The resultant model clearly shows the complementarity of the interactions within the various heterodimers and positions each of the bridging subunits between the heterodimers (Fig. 2). Restraints determined by mass spectrometry do not indicate whether the ring runs clockwise or anticlockwise, so the alternative enantiomer was also modelled. In this case, however, the interfaces within the heterodimers were less complementary than those in the first model, and the bridging subunits appear between the subunits within each heterodimer instead of between the heterodimers themselves. This arrangement is therefore not supported by experimental data on the bridging subunits[9]. Moreover, in this alternative model, the active sites of the catalytically active (RNase pleckstrin-homology (PH) domain) subunits — Rrp41 (also known as Ski6), Rrp46 and Mtr3 — are pointing towards the bridging subunits, which is in contrast to the known orientation of the Rrp41 equivalent in the archaeal RNA exosome[59]. An atomic model was then constructed (Fig. 2): this model is the best fit to the experimental data and is in close agreement with the structure of the related human RNA exosome, which was determined recently by using X-ray crystallography after reconstitution of nine subunits *in vitro*[23]. This example highlights the power of mass spectrometry and comparative protein-structure modelling to generate an atomic model of a complex protein assembly that has eluded determination by X-ray crystallography.

The architecture of the 26S proteasome

Determining the structure of the 26S proteasome presents an even greater challenge. Whereas the yeast RNA exosome can be isolated as a relatively homogeneous assembly, the 26S proteasome is labile and is therefore often heterogeneous. Moreover, unlike the yeast RNA exosome, there are few structures available for the components of the 26S proteasome, precluding atomic-resolution characterization.

The eukaryotic 26S proteasome is a large (2.5 MDa) molecular machine similar in size to the ribosome; it consists of one or two 19S regulatory complexes attached to the ends of a barrel-shaped 20S core complex. It has a central role in intracellular protein degradation, proteolytically cleaving proteins that have been marked for destruction by the attachment of multiple ubiquitin molecules[62]. The structure of the 20S core complex, which is highly conserved from Archaea to mammals, was solved by X-ray crystallography[63], revealing salient features of this protease[18]. A recent study also uncovered aspects of the structural changes that are involved in the functioning of the core complex, by using NMR spectroscopy[64]. By contrast, it has not been possible to crystallize the 26S holocomplex. The 19S regulatory subunits — which comprise at
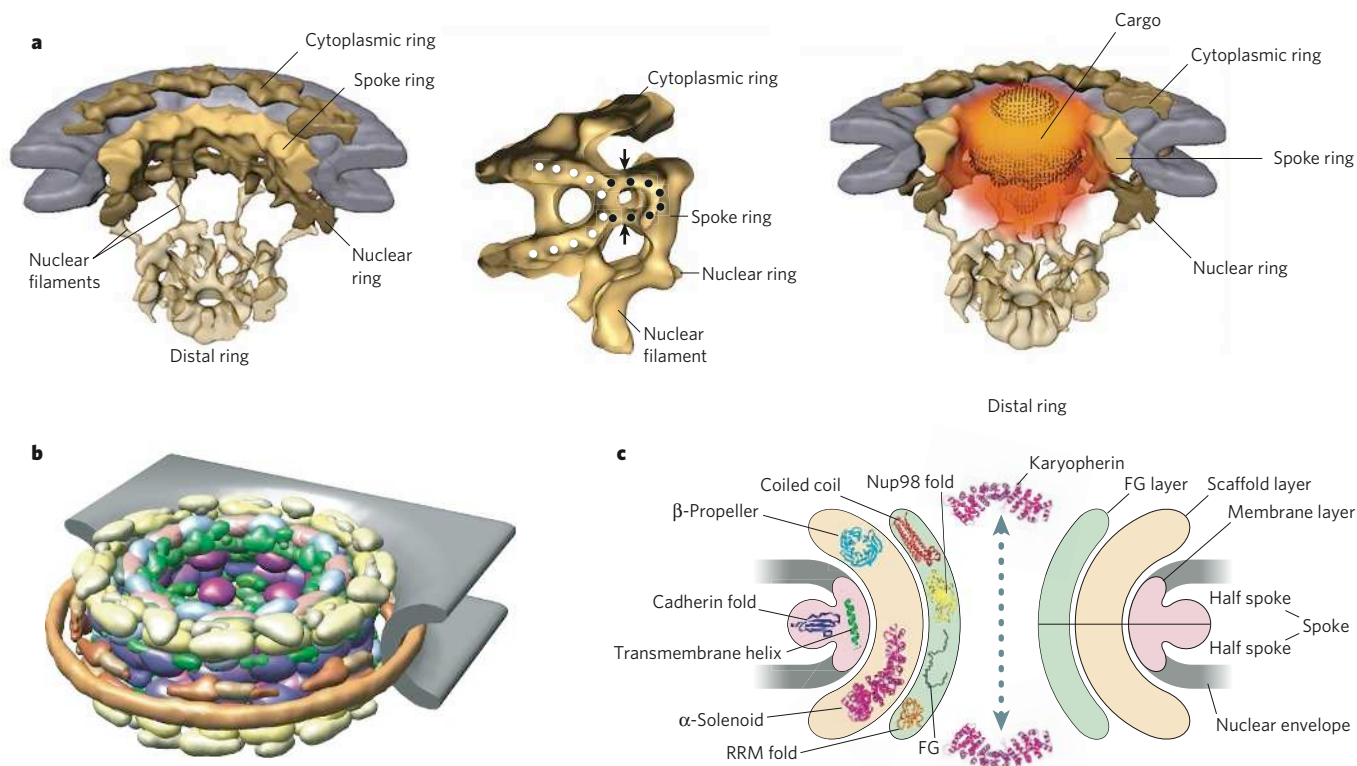
Figure 4 | The molecular architecture of the NPC. By using a variety of techniques, different aspects of the NPC structure have been revealed. **a**, Using cryo-electron tomography, a density map of the *Dictyostelium discoideum* NPC at 5.8 nm resolution was generated, allowing single molecules to be observed during nuclear import[20]. A cutaway view of the structure of rejoined asymmetrical units is shown (left), with subjective segmentation for the cytoplasmic ring, spoke ring and nuclear ring (brown and yellow), and the inner nuclear membrane and outer nuclear membrane (that is, the nuclear envelope; grey). For clarity, the central plug (that is, the transporter) has been omitted, and the basket with nuclear filaments and distal ring was rendered transparent. A cutaway view of a protomer is shown (centre). The fused inner nuclear membrane and outer nuclear membrane (white circles), as well as the clamp-shaped spoke structure (black circles), are indicated; arrows mark the entry and exit of what seems to be a channel. A cutaway view of the NPC structure with a three-dimensional probability distribution of import cargo is shown (right). The classical import cargo NLS–2GFP (Asn-Leu-Ser with two green fluorescent protein molecules attached) was labelled with gold, and the probability distribution for the cargo (orange; brightness indicates higher probability) is superimposed onto the central plug (brown dots). **b**, Various experimental data were integrated[7], revealing the configuration of the 456 core proteins (excluding FG (Phe-Gly) repeats in FG nucleoporins and the basket) that form the yeast NPC[21]. The inner and outer nuclear membranes (grey) are shown. The NPC proteins are coloured according to their assignment to various NPC modules: membrane rings (brown), outer rings (yellow), inner rings (purple, light and dark shades), linker nucleoporins (blue and pink, light shades) and FG nucleoporins (green). (Panel adapted, with permission, from ref. 7.) **c**, Structural folds were assigned to the domains of the NPC proteins, by comparing their sequences to those of known protein structures, revealing a simple fold composition and modular architecture for the NPC[72]. The architecture of the NPC ring, viewed as a transverse section, is segregated into three layers: membrane (pale pink), scaffold (pale yellow) and FG (pale green). The arrow denotes the direction of cargo transport. RRM, RNA-recognition motif.
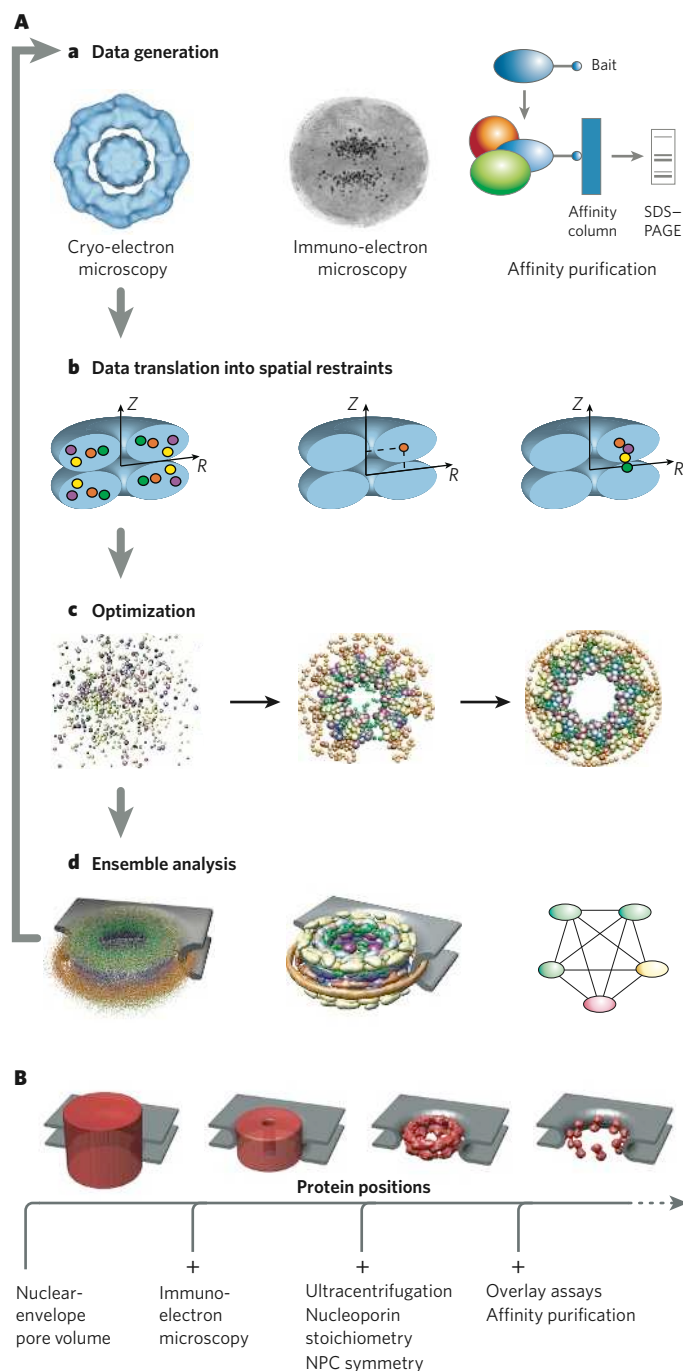
977

Figure 5 | Integrative structure determination. A, Using the NPC as an example[7], the four steps to determine a structure by integrating varied data are illustrated. These steps are data generation (**a**), data translation into spatial restraints (**b**), optimization (**c**) and ensemble analysis (**d**). **a**, First, structural data are generated by experiments, such as cryo-electron microscopy (left), immuno-electron microscopy (centre) and affinity purification of subcomplexes (right). Many other types of information can also be included. **b**, Second, the data and theoretical considerations are expressed as spatial restraints that ensure the observed symmetry and shape of the assembly (from cryo-electron microscopy, left), the positions of constituent gold-labelled proteins (from immuno-electron microscopy, centre) and the proximities of the constituent proteins (from affinity purification, right). The assembly is indicated in blue, and constituent proteins are indicated as coloured circles. **c**, Third, an ensemble of structural solutions that satisfy the data is obtained by minimizing the violations of the spatial restraints (from left to right). **d**, Fourth, the ensemble is clustered into sets of distinct solutions (left), and analysed in different representations, such as protein positions (centre) and protein–protein contacts (right). The integrative approach to structure determination has several advantages. First, synergy among the input data minimizes the drawbacks of incomplete, inaccurate and/or imprecise data sets. Each individual restraint contains little structural information, but by concurrently satisfying all restraints derived from independent experiments, the degeneracy of structural solutions can be markedly reduced. Second, this approach has the potential to produce all structures that are consistent with the data, not just one structure. Third, the variation between the structures that are consistent with the data allows an assessment of whether there are sufficient data and how precise the representative structure is. Last, this approach can make the process of structure determination more efficient, by indicating which measurements would be the most informative. **B**, When applying the process described in **A**, the position of each protein is specified with increasing accuracy and precision as each type of synergistic experimental information is added[7]. Each panel illustrates the localization volume (red) of 16 copies of nucleoporin 192 (Nup192) in the ensemble of NPC structures that satisfy the spatial restraints corresponding to the experimental data sets indicated. The smaller the volume, the better the proteins are localized. Further experiments could localize the proteins to a greater degree, as indicated by the dashed arrow. Therefore, the NPC structure is, in essence, 'moulded' into shape by the large quantity of diverse experimental data. (Panel reproduced with permission from ref. 7.)

least 18 subunits, including 6 ATPases — bind to ubiquitylated substrates and prepare them for degradation in the core complex. Structural studies of the 26S holocomplex, using cryo-electron microscopy, have been hampered by the low intrinsic stability of the complex, which tends to dissociate during purification and sample preparation. The dynamics of the complex present another problem: in addition to a set of 'canonical' subunits, there are several variable subunits; therefore, the composition of individual complexes varies, modulating proteasome function[65]. In principle, single-particle cryo-electron microscopy can handle heterogeneous samples that contain several distinct subsets of particles. Image classification allows particles to be sorted, thus achieving structural homogeneity *in silico*[66]. For a detailed classification, however, large sets of images are needed, and acquiring these is greatly facilitated by automated image recording[67]. At the present level of resolution (~2.5 nm), the spatial arrangement of the subunits of the 26S proteasome cannot be determined. Fortunately, there is a wealth of information on interactions between the proteasomal subunits, obtained from yeast two-hybrid

studies[68] and mass spectrometry[45], as well as other sources[69] (Fig. 3). The challenge therefore is to interpret the current cryo-electron-microscopy map in light of these data. This should not be done in an *ad hoc* manner but by a systematic search for all structures that satisfy the restraints implied by the data. The power of such an approach is illustrated by the recent description of the architecture of the NPC[7,21].

The architecture of the NPC

NPCs are large proteinaceous assemblies that span the nuclear envelope, where they function as the main mediators of bidirectional exchange between the nucleoplasmic and cytoplasmic compartments in all eukaryotes[19]. Cryo-electron-microscopy images of the NPC show that it forms a channel through the stacking of two similar rings, each consisting of eight copies of the basic symmetry unit of the NPC (that is, the 'half spoke')[70]. In yeast, each half spoke contains ~30 different proteins known as nucleoporins, resulting in 456 proteins in the whole NPC, which has a mass of ~50 MDa[71]. Owing to its size and flexibility,

---

Box 2 | Cryo-electron microscopy

Cryo-electron microscopy is a generic term that refers to various electron-microscopy imaging modalities when applied to samples embedded in amorphous ice[91]. Samples are vitrified by plunge freezing or high-pressure freezing. A short description of the three main branches of cryo-electron microscopy is provided below.

Electron crystallography
Electron crystallography relies on the availability of two-dimensional crystals, either natural or synthetic. It is particularly suited to studying membrane proteins, but its use is not restricted to this class of protein. Very high resolution can be attained by optimizing the imaging conditions and by applying image-processing strategies to compensate for imperfections in the crystal lattices. Data acquisition can be time-consuming because of difficulties in collecting data sets of consistent quality; image quality is often degraded, particularly at high tilt angles, for reasons that are not well understood at present.

Single-particle analysis
Single-particle analysis (arguably a misleading name) relies on the existence of multiple copies of the object. Molecules suspended in thin layers of ice occur in random orientations. After grouping them into classes that correspond to common orientations, class averages are generated. Three-dimensional reconstructions are obtained by assigning relative orientations to the class averages and placing them in a virtual tilt experiment. Single-particle analysis is particularly suited to studying macromolecular complexes — the larger, the better. Some degree of heterogeneity in the sample (for example, variations in subunit composition, stoichiometry or conformational states) is tolerable and can be taken into account by image classification. There is no fundamental reason why atomic resolution could not be attained, but until now this has remained an elusive goal. Medium-resolution maps (1–2 nm) can be obtained routinely. This resolution is usually sufficient for fitting high-resolution structures of components (that is, subunits or domains) obtained by other methods into the cryo-electron-microscopy maps of the

complex. At subnanometre resolution, elements of secondary structure can be discerned, enabling docking to be carried out with high accuracy. Efforts are under way to increase the speed of single-particle techniques, by automated data acquisition and image analysis[92].

Electron tomography
Electron tomography is unique in its capability to provide three-dimensional reconstructions of non-repetitive structures. Therefore, it enables insights into the molecular architecture of higher-order structures that have a degree of stochasticity. Objects are reconstructed from a series of transmission electron micrographs taken from different viewing angles. During data collection, the requirement for optimal sampling must be reconciled with the need to avoid radiation damage (through sustaining a low cumulative radiation dose). Tomograms taken in these conditions are rich in information, but the poor signal-to-noise ratio makes interpretation difficult. Tomograms of intact cells or organelles are images of their entire proteomes, and sophisticated pattern-recognition methods must be applied to make use of this information. At a resolution of 4–5 nm, typically obtained for intact cells, only large complexes can be visualized and mapped with an acceptable fidelity. With ongoing advances in instrumentation, however, resolutions of 2–3 nm are a realistic goal and will enable cells to be mapped more comprehensively. Better image-processing tools are needed to refine and validate such maps and to derive molecular-interaction patterns from them.

Generating pseudo-atomic models of assemblies
Fitting atomic structures and models of proteins and nucleic acids into cryo-electron-microscopy maps has resulted in pseudo-atomic models of many assemblies: complexes of viral subunits[93], ribosomes and ribosome-interacting proteins[94], the chaperone complex containing heat-shock protein 90 (ref. 95), cytoskeletal proteins and associated proteins[96,97], spliceosomal components[98], clathrin cages[16] and COPII cages[99]. Moreover, single-particle cryo-electron microscopy is becoming increasingly powerful at capturing assemblies in different conformational states[100].

---

detailed structural characterization of the complete NPC has proven to be extraordinarily difficult. Further compounding the problem, atomic structures have been solved only for domains that cover ~5% of the protein sequences[72]. As a result, the NPC is a challenging model system that is suitable for developing methods to map the molecular architectures of many other assemblies.

Cryo-electron tomography allows macromolecular assemblies to be studied *in situ*, eliminating the risk of preparation-induced artefacts and preserving the function of the structure[73] (discussed further in the next section). Thus, it is possible to take snapshots of molecular machines in action. This technique was applied to NPCs that were actively importing molecules into the intact nuclei of *Dictyostelium discoideum*. Many such snapshots were obtained and superimposed, yielding a map outlining the trajectories of the cargo[20] (Fig. 4a). Closer inspection of individually reconstructed NPCs shows substantial plasticity, probably reflecting both intrinsic dynamics and distortions that result from strain. To avoid the loss of resolution caused by averaging individually variable entities, a deformation analysis was carried out. This allows deviations from perfect eight-fold symmetry to be determined, and it provides the basis for the computational compensation of such distortions. Despite substantial improvements in resolution, the current resolution of 5.8 nm still falls short of that needed to determine the spatial arrangement of the component proteins.

The approximate spatial arrangement of the component proteins (Fig. 4b) can, however, be determined by integrating a variety of experimental data[7,21], using the approach outlined in Fig. 5. In a structure calculation, each of the 456 proteins in the yeast NPC was represented by a flexible chain consisting of a small number of connected beads (the numbers and radii of which were chosen to match the molecular masses and Stokes radii of the proteins). Next, to capture information about the structure of the NPC, a scoring function was constructed, which was a sum of

spatial restraints of various types. These restraints incorporated data about protein shapes (from the protein sequences and ultracentrifugation), component protein positions (from immuno-electron microscopy), protein contacts (from affinity purification), eight-fold and two-fold symmetries of the NPC (from cryo-electron microscopy) and nuclear-envelope shape (from cryo-electron microscopy). The relative positions and proximities of the constituent proteins of the NPC were then calculated by satisfying these spatial restraints.

The calculation started with a random protein configuration and then iteratively moved the proteins so as to minimize violations of the restraints, relying on conjugate gradients and molecular dynamics with simulated annealing. To sample comprehensively all possible structural solutions that are consistent with the data, an 'ensemble' of 1,000 independently calculated structures that satisfy the input restraints was obtained. After superimposing the structures, the ensemble was converted into the probability of any volume element being occupied by a given protein (that is, the localization probability). The resultant localization probabilities yielded single pronounced maxima for almost all nucleoporins, showing that the input restraints define one predominant architecture for the NPC. The average standard deviation for the separation between nucleoporins is 5 nm. Given that this is less than the diameter of many NPC constituents, the map is sufficient to determine the relative positions of the proteins in the NPC. Although each individual restraint contains little structural information, the degeneracy of the structural solutions is markedly reduced by concurrently satisfying all restraints.

The arrangement of the proteins in the NPC (Fig. 4b, c), determined by the above approach, revealed that half of the NPC consists of a core scaffold, which is structurally analogous to vesicle-coating complexes[21,72]. This scaffold forms an interlaced network that coats the entire curved surface of the nuclear envelope, within which the NPC is embedded. The selective barrier to transport between the nucleoplasmic and cytoplasmic
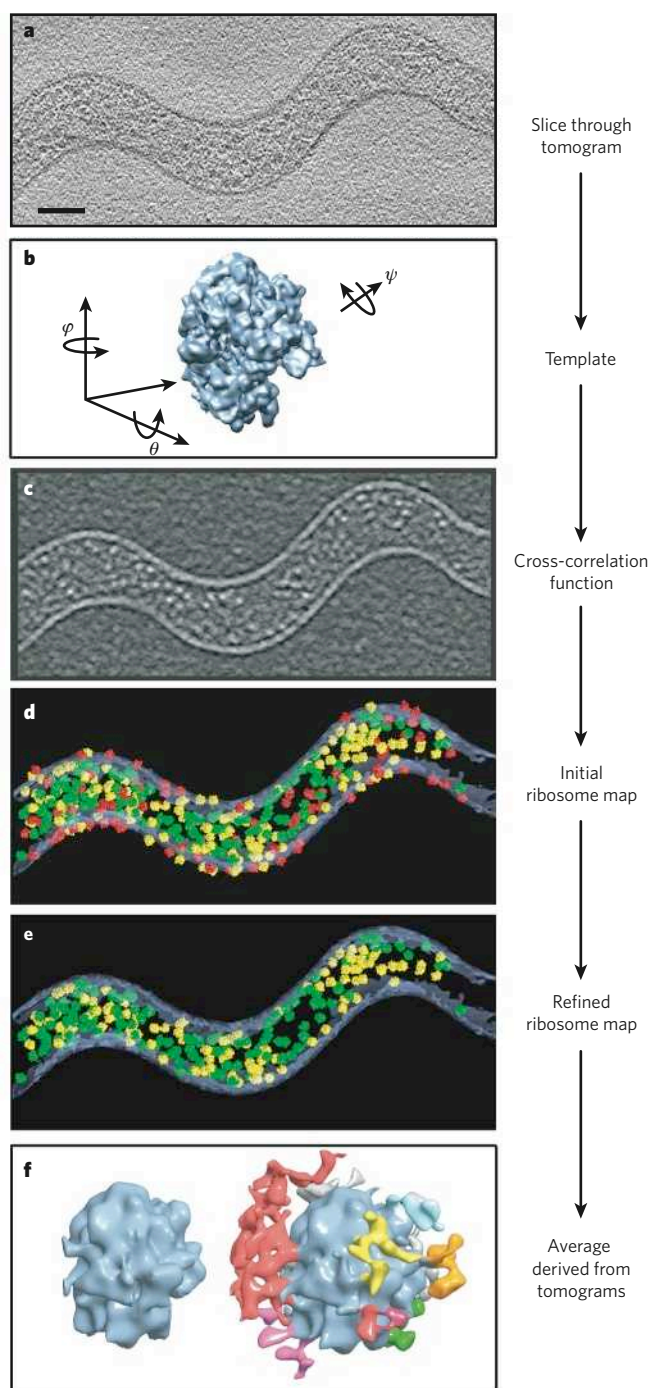
Figure 6 | Mapping of 70S ribosomes in a tomogram of the bacterium *Spiroplasma melliferum*[80]. a, An orthogonal slice through a tomogram of *S. melliferum* is shown. Scale bar, 100 nm. **b,** To determine the positions and orientations of the ribosomes in this cell, a template obtained by single-particle analysis[81] (resolution 11.5 Å) was correlated with the tomogram. **c,** In the cross-correlation function, white spots indicate sites where ribosomes were detected. **d,** From the cross-correlation function, a ribosome map was derived. Colours correspond to detection fidelity: high (green), intermediate (yellow) and low (red). **e,** After the initial ribosome map was generated, putative false positives were removed, leading to the refined map. The ribosomes that were identified and localized by template matching occupy ~5% of the cellular volume, which agrees well with estimates derived from other measurements. **f,** From the refined map, an average of the 70S ribosome was derived at a resolution of 45 Å (left). When the threshold for the isosurface representation of this map was lowered (right), distinct masses become visible near the ribosome. At present, these densities cannot be interpreted, but they most probably represent nascent chains, chaperones and other interacting factors. (Figure adapted, with permission, from ref. 80.)

compartments is formed by large numbers of FG nucleoporins, with disordered regions lining the inner face of the scaffold. The NPC consists of only a few structural modules. These modules resemble each other in terms of the configuration of their homologous constituents, thus providing clues to the ancient evolutionary origins of the NPC.

Studying functional modules *in situ*

Characterizing the NPC *in situ* required a non-invasive imaging technique. The technique used, cryo-electron tomography, generates images of large pleiomorphic objects — not only protein assemblies but also organelles. It does this by reconstructing three-dimensional objects from a series of two-dimensional transmission electron-microscopy images taken from different viewing angles.

Although the principles of electron tomography have been known for decades, its use has gathered momentum only recently. Technological advances have enabled the development of automated data-acquisition procedures, which in turn has reduced the total dose of electrons to a level at which radiation-sensitive biological materials, embedded in ice, can be studied[73] (Box 2). As a result, researchers are now poised to combine the potential of three-dimensional imaging with a 'close-to-life' preservation of biological specimens. At present, the resolution of cellular objects in cryo-electron-tomography studies is usually limited to 4–5 nm, but prospects for attaining molecular resolution (that is, 2–3 nm) are good[74].

Molecular-resolution tomograms of intact organelles or cells contain vast amounts of information. In essence, they are three-dimensional images of the entire proteome of a cell, and they should enable the spatial relationships of the macromolecules in a cell (the 'interactome') to be mapped (a process referred to as visual proteomics). Advanced pattern-recognition methods are needed to interpret the 'noisy' tomograms in an objective and systematic manner. This approach has two requirements: the proteomic 'inventory' must have been determined by mass-spectrometry analysis, and a library of template structures must be available so that tomograms can be interpreted by matching the cellular tomograms with the template structures[75]. Template structures can be generated by direct experimental methods, as well as by hybrid approaches. In the long term, with increasing numbers of structures of complexes deposited into the databases, template structures could be drawn from these databases.

We envisage a situation in which high-quality tomograms of a large range of cell types, generated with advanced instrumentation, will be made available to the scientific community, together with the software needed for their interpretation. This resource would enable researchers who have determined structures of complexes to use them as templates for exploring their functional environment. At the currently achievable resolution, only large complexes (such as ribosomes and proteasomes) can be mapped with an acceptable fidelity (Fig. 6; Box 2). But, with advances in instrumentation and methodology, today's imaging capabilities will improve, allowing proteomes to be mapped in a comprehensive manner. The remaining challenges are to untangle huge data sets, to derive interaction patterns from maps of intimidating complexity, and to understand the underlying molecular sociology.

Outlook

Constructing atomic models of functional modules in action will improve the current understanding of how cells function at many levels. To achieve this aim, new integrative methods are required, especially for dealing with the heterogeneity and dynamics of transient functional modules. One such hybrid approach that shows great promise is a combination of mass spectrometry and electron microscopy[76] in which isolation of functional modules is achieved in the gas phase. This allows selection of complexes on the basis of mass-to-charge ratio from a heterogeneous ensemble of closely related complexes. Subsequent 'soft landing' on suitable electron-microscopy grids then allows simultaneous characterization and visualization of transient complexes. These new hybrid methods, together with further computational integration, make revealing the molecular architecture of even fleeting social interactions within functional modules an enticing possibility. ∎

1. Blundell, T. L. & Johnson, L. *Protein Crystallography* (Academic, New York, 1976).
2. Wimberley, B. T. *et al.* Structure of the 30S ribosomal subunit. *Nature* **407,** 327–339 (2000).
3. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å. *Science* **289,** 905–920 (2000).
4. Schluenzen, F. *et al.* Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* **102,** 615–623 (2000).
5. Malhotra, A. & Harvey, S. C. A quantitative model of the *Escherichia coli* 16S RNA in the 30S ribosomal subunit. *J. Mol. Biol.* **240,** 308–340 (1994).
6. Alber, F., Kim, M. F. & Sali, A. Structural characterization of assemblies from overall shape and subcomplex compositions. *Structure* **13,** 435–445 (2005).
7. Alber, F. *et al.* Determining the architectures of macromolecular assemblies. *Nature* **450,** 683–694 (2007).
8. Sali, A., Glaeser, R., Earnest, T. & Baumeister, W. From words to literature in structural proteomics. *Nature* **422,** 216–225 (2003).
9. Hernandez, H., Dziembowski, A., Taverner, T., Seraphin, B. & Robinson, C. V. Subunit architecture of multimeric complexes isolated directly from cells. *EMBO Rep.* **7,** 605–610 (2006).
10. Davis, F. P. *et al.* Protein complex compositions predicted by structural similarity. *Nucleic Acids Res.* **34,** 2943–2952 (2006).
11. van Dijk, A. D. *et al.* Modeling protein–protein complexes involved in the cytochrome *c* oxidase copper-delivery pathway. *J. Proteome Res.* **6,** 1530–1539 (2007).
12. Todd, A. E., Marsden, R. L., Thornton, J. M. & Orengo, C. A. Progress of structural genomics initiatives: an analysis of solved target structures. *J. Mol. Biol.* **348,** 1235–1260 (2005).
13. Alber, F., Eswar, N. & Sali, A. in *Practical Bioinformatics* 1950–1954 (Springer, Heidelberg, 2004).
14. Sivasubramanian, A., Chao, G., Pressler, H. M., Wittrup, K. D. & Gray, J. J. Structural model of the mAb 806–EGFR complex using computational docking followed by computational and experimental mutagenesis. *Structure* **14,** 401–414 (2006).
15. Rossmann, M. G., Morais, M. C., Leiman, P. G. & Zhang, W. Combining X-ray crystallography and electron microscopy. *Structure* **13,** 355–362 (2005).
16. Fotin, A. *et al.* Structure of an auxilin-bound clathrin coat and its implications for the mechanism of uncoating. *Nature* **432,** 649–653 (2004).
17. Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. & Tollervey, D. The exosome: a conserved eukaryotic RNA processing complex containing multiple 3′→5′ exoribonucleases. *Cell* **91,** 457–466 (1997).
18. Baumeister, W., Walz, J., Zuhl, F. & Seemuller, E. The proteasome: paradigm of a self-compartmentalizing protease. *Cell* **92,** 367–380 (1998).
19. Lim, R. Y. & Fahrenkrog, B. The nuclear pore complex up close. *Curr. Opin. Cell Biol.* **18,** 342–347 (2006).
20. Beck, M., Lucic, V., Forster, F., Baumeister, W. & Medalia, O. Snapshots of nuclear pore complexes in action captured by cryo-electron tomography. *Nature* **449,** 611–615 (2007).
21. Alber, F. *et al.* The molecular architecture of the nuclear pore complex. *Nature* **450,** 695–701 (2007).
22. Meinhart, A. & Cramer, P. Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. *Nature* **430,** 223–226 (2004).
23. Liu, Q., Greimann, J. C. & Lima, C. D. Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* **127,** 1223–1237 (2006).
24. Egea, P. F. *et al.* Substrate twinning activates the signal recognition particle and its receptor. *Nature* **427,** 215–221 (2004).
25. Bonvin, A. M., Boelens, R. & Kaptein, R. NMR analysis of protein interactions. *Curr. Opin. Chem. Biol.* **9,** 501–508 (2005).
26. Zuiderweg, E. R. Mapping protein–protein interactions in solution by NMR spectroscopy. *Biochemistry* **41,** 1–7 (2002).
27. McCoy, M. A. & Wyss, D. F. Structures of protein–protein complexes are docked using only NMR restraints from residual dipolar coupling and chemical shift perturbations. *J. Am. Chem. Soc.* **124,** 2104–2105 (2002).
28. Wuthrich, K. The way to NMR structures of proteins. *Nature Struct. Biol.* **8,** 923–925 (2001).
29. Rieping, W., Habeck, M. & Nilges, M. Inferential structure determination. *Science* **309,** 303–306 (2005).
30. Vachette, P., Koch, M. H. & Svergun, D. I. Looking behind the beamstop: X-ray solution scattering studies of structure and conformational changes of biological macromolecules. *Methods Enzymol.* **374,** 584–615 (2003).
31. Nagar, B. & Kuriyan, J. SAXS and the working protein. *Structure* **13,** 169–170 (2005).
32. Tidow, H. *et al.* Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex. *Proc. Natl Acad. Sci. USA* **104,** 12324–12329 (2007).
33. Grishaev, A., Wu, J., Trewhella, J. & Bax, A. Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. *J. Am. Chem. Soc.* **127,** 16621–16628 (2005).
34. Rosenberg, O. S., Deindl, S., Sung, R. J., Nairn, A. C. & Kuriyan, J. Structure of the autoinhibited kinase domain of CaMKII and SAXS analysis of the holoenzyme. *Cell* **123,** 849–860 (2005).
35. Sondermann, H., Nagar, B., Bar-Sagi, D. & Kuriyan, J. Computational docking and solution X-ray scattering predict a membrane-interacting role for the histone domain of the Ras activator son of sevenless. *Proc. Natl Acad. Sci. USA* **102,** 16632–16637 (2005).
36. Yamagata, A. & Tainer, J. A. Hexameric structures of the archaeal secretion ATPase GspE and implications for a universal secretion mechanism. *EMBO J.* **26,** 878–890 (2007).
37. Hainfeld, J. F. & Powell, R. D. New frontiers in gold labeling. *J. Histochem. Cytochem.* **48,** 471–480 (2000).
38. Pye, V. E. *et al.* Structural insights into the p97-Ufd1-Npl4 complex. *Proc. Natl Acad. Sci. USA* **104,** 467–472 (2007).
39. Guan, J. Q., Almo, S. C., Reisler, E. & Chance, M. R. Structural reorganization of proteins revealed by radiolysis and mass spectrometry: G-actin solution structure is divalent cation dependent. *Biochemistry* **42,** 11992–12000 (2003).
40. Anand, G. S. *et al.* Identification of the protein kinase A regulatory R'α-catalytic subunit interface by amide H/²H exchange and protein docking. *Proc. Natl Acad. Sci. USA* **100,** 13264–13269 (2003).
41. Lee, T. *et al.* Docking motif interactions in MAP kinases revealed by hydrogen exchange mass spectrometry. *Mol. Cell* **14,** 43–55 (2004).
42. Yan, Y. & Marriott, G. Analysis of protein interactions using fluorescence technologies. *Curr. Opin. Chem. Biol.* **7,** 635–640 (2003).
43. Muller, E. G. *et al.* The organization of the core proteins of the yeast spindle pole body. *Mol. Biol. Cell* **16,** 3341–3352 (2005).
44. Gavin, A. C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440,** 631–636 (2006).
45. Sharon, M., Taverner, T., Ambroggio, X. I., Deshaies, R. J. & Robinson, C. V. Structural organization of the 19S proteasome lid: insights from MS of intact complexes. *PLoS Biol.* **4,** e267 (2006).
46. Parrish, J. R., Gulyas, K. D. & Finley, R. L. Yeast two-hybrid contributions to interactome mapping. *Curr. Opin. Biotechnol.* **17,** 387–393 (2006).
47. Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae. Nature* **403,** 623–627 (2000).
48. Michnick, S. W., Ear, P. H., Manderson, E. N., Remy, I. & Stefan, E. Universal strategies in research and drug discovery based on protein-fragment complementation assays. *Nature Rev. Drug Discov.* **6,** 569–582 (2007).
49. Landgraf, C. *et al.* Protein interaction networks by proteome peptide scanning. *PLoS Biol.* **2,** e14 (2004).
50. MacBeath, G. & Schreiber, S. L. Printing proteins as microarrays for high-throughput function determination. *Science* **289,** 1760–1763 (2000).
51. Piehler, J. New methodologies for measuring protein interactions *in vivo* and *in vitro. Curr. Opin. Struct. Biol.* **15,** 4–14 (2005).
52. Collins, S. R. *et al.* Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446,** 806–810 (2007).
53. Krogan, N. J., Cagney, G., Haiyuan, Y., Zhong, G. & Guo, X. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae. Nature* **440,** 637–643 (2006).
54. Collins, S. R. *et al.* Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae. Mol. Cell. Proteomics* **6,** 439–450 (2007).
55. Bauer, A. & Kuster, B. Affinity purification — mass spectrometry. Powerful tools for the characterization of protein complexes. *Eur. J. Biochem.* **270,** 570–578 (2003).
56. Rappas, M. *et al.* Structural insights into the activity of enhancer-binding proteins. *Science* **307,** 1972–1975 (2005).
57. Poliakov, A. *et al.* Macromolecular mass spectrometry and electron microscopy as complementary tools for investigation of the heterogeneity of bacteriophage portal assemblies. *J. Struct. Biol.* **157,** 371–383 (2007).
58. Hernandez, H. & Robinson, C. V. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nature Protoc.* **2,** 715–726 (2007).
59. Lorentzen, E. *et al.* The archaeal exosome core is a hexameric ring structure with three catalytic subunits. *Nature Struct. Mol. Biol.* **12,** 575–581 (2005).
60. Buttner, K., Wenig, K. & Hopfner, K. P. Structural framework for the mechanism of archaeal exosomes in RNA processing. *Mol. Cell* **20,** 461–471 (2005).
61. Aloy, P. *et al.* Structure-based assembly of protein complexes in yeast. *Science* **303,** 2026–2029 (2004).
62. Voges, D., Zwickl, P. & Baumeister, W. The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu. Rev. Biochem.* **68,** 1015–1068 (1999).
63. Groll, M. *et al.* Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* **386,** 463–471 (1997).
64. Sprangers, R. & Kay, L. E. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **445,** 618–622 (2007).
65. Hanna, J. & Finley, D. A proteasome for all occasions. *FEBS Lett.* **581,** 2854–2861 (2007).
66. Scheres, S. H. W. *et al.* Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nature Methods* **4,** 27–29 (2007).
67. Nickell, S. *et al.* Automated cryoelectron microscopy of 'single particles' applied to the 26S proteasome. *FEBS Lett.* **581,** 2751–2756 (2007).
68. Davy, A. *et al.* A protein–protein interaction map of the *Caenorhabditis elegans* 26S proteasome. *EMBO Rep.* **2,** 821–828 (2001).
69. Ferrell, K., Wilkinson, C. R., Dubiel, W. & Gordon, C. Regulatory subunit interactions of the 26S proteasome, a complex problem. *Trends Biochem. Sci.* **25,** 83–88 (2000).
70. Hinshaw, J. E., Carragher, B. O. & Milligan, R. A. Architecture and design of the nuclear pore complex. *Cell* **69,** 1133–1141 (1992).
71. Rout, M. P. *et al.* The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148,** 635–651 (2000).
72. Devos, D. *et al.* Simple fold composition and modular architecture of the nuclear pore complex. *Proc. Natl Acad. Sci. USA* **103,** 2172–2177 (2006).
73. Koster, A. J. *et al.* Perspectives of molecular and cellular electron tomography. *J. Struct. Biol.* **120,** 276–308 (1997).
74. Nickell, S., Kofler, C., Leis, A. P. & Baumeister, W. A visual approach to proteomics. *Nature Rev. Mol. Cell. Biol.* **7,** 225–230 (2006).
75. Baumeister, W. From proteomic inventory to architecture. *FEBS Lett.* **579,** 933–937 (2005).
76. Benesch, J. L., Ruotolo, B. T., Simmons, D. A. & Robinson, C. V. Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chem. Rev.* **107,** 3544–3567 (2007).
77. Lowe, J. *et al.* Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* **268,** 533–539 (1995).
78. Unno, M. *et al.* The structure of the mammalian 20S proteasome at 2.75 Å resolution. *Structure* **10,** 609–618 (2002).
79. Kwon, Y. D., Nagy, I., Adams, P. D., Baumeister, W. & Jap, B. K. Crystal structures of the *Rhodococcus* proteasome with and without its pro-peptides: implications for the role of the pro-peptide in proteasome assembly. *J. Mol. Biol.* **335,** 233–245 (2004).
80. Ortiz, J. O., Forster, F., Kurner, J., Linaroudis, A. A. & Baumeister, W. Mapping 70S ribosomes in intact cells by cryoelectron tomography and pattern recognition. *J. Struct. Biol.* **156,** 334–341 (2006).
81. Gabashvili, I. S. *et al.* Solution structure of the *E. coli* 70S ribosome at 11.5 Å resolution. *Cell* **100,** 537–549 (2000).
82. Sharon, M. & Robinson, C. V. The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu. Rev. Biochem.* **76,** 167–193 (2007).

83. Ilag, L. L. *et al.* Heptameric (L12)$_6$/L10 rather than canonical pentameric complexes are found by tandem MS of intact ribosomes from thermophilic bacteria. *Proc. Natl Acad Sci. USA* **102,** 8192–8197 (2005).

84. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422,** 198–207 (2003).

85. Synowsky, S. A., van den Heuvel, R. H., Mohammed, S., Pijnappel, P. W. & Heck, A. J. Probing genuine strong interactions and post-translational modifications in the heterogeneous yeast exosome protein complex. *Mol. Cell. Proteomics* **5,** 1581–1592 (2006).

86. Back, J. W., de Jong, L., Muijsers, A. O. & de Koster, C. G. Chemical cross-linking and mass spectrometry for protein structural modeling. *J. Mol. Biol.* **331,** 303–313 (2003).

87. Vasilescu, J. & Figeys, D. Mapping protein–protein interactions by mass spectrometry. *Curr. Opin. Biotechnol.* **17,** 394–399 (2006).

88. von Helden, G., Wyttenbach, T. & Bowers, M. T. Conformation of macromolecules in the gas phase: use of matrix-assisted laser desorption methods in ion chromatography. *Science* **267,** 1483–1485 (1995).

89. Ruotolo, B. T. *et al.* Evidence for macromolecular protein rings in the absence of bulk water. *Science* **310,** 1658–1661 (2005).

90. Ruotolo, B. T. *et al.* Ion mobility–mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes. *Angew. Chem. Int. Ed. Engl.* **46,** 8001–8004 (2007).

91. Henderson, R. Realizing the potential of electron cryo-microscopy. *Q. Rev. Biophys.* **37,** 3–13 (2004).

92. Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151,** 41–60 (2005).

93. Johnson, J. E. & Chiu, W. DNA packaging and delivery machines in tailed bacteriophages. *Curr. Opin. Struct. Biol.* **17,** 237–243 (2007).

94. Taylor, D. J. *et al.* Structures of modified eEF2 80S ribosome complexes reveal the role of GTP hydrolysis in translocation. *EMBO J.* **26,** 2421–2431 (2007).

95. Vaughan, C. K. *et al.* Structure of an Hsp90-Cdc37-Cdk4 complex. *Mol. Cell* **23,** 697–707 (2006).

96. Woodhead, J. L. *et al.* Atomic model of a myosin filament in the relaxed state. *Nature* **436,** 1195–1199 (2005).

97. Wang, H. W. & Nogales, E. Nucleotide-dependent bending flexibility of tubulin regulates microtubule assembly. *Nature* **435,** 911–915 (2005).

98. Stark, H. & Luhrmann, R. Cryo-electron microscopy of spliceosomal components. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 435–457 (2006).

99. Fath, S., Mancias, J. D., Bi, X. & Goldberg, J. Structure and organization of coat proteins in the COPII cage. *Cell* **129,** 1325–1336 (2007).

100. Mitra, K. & Frank, J. Ribosome dynamics: insights from atomic structure modeling into cryo-electron microscopy maps. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 299–317 (2006).