

## 2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

around the zero level increasing the chance of selecting an incorrect resolution point. In other criteria one tries to equate the resolution with the frequency at which noise begins to dominate the signal. A good choice of the cut-off level is  $SSNR = 1.0$ , a level at which the power of the signal in the reconstruction is equal to the power of the noise. According to (2.5.7.22), this corresponds to  $FSC = 0.333$ . Another often-used cut-off level is  $FSC = 0.5$ , at which the  $SSNR$  in the reconstruction is 2.0 (Böttcher *et al.*, 1997; Conway *et al.*, 1997; Penczek, 1998).

The main reason behind the determination of the resolution of the EM maps is the necessary step of low-pass filtration of the results before the interpretation of the map is attempted. In order to avoid mistakes, particularly the danger of overinterpretation, one has to remove from the map unreliable Fourier coefficients. Inclusion of Fourier coefficients with a low  $SNR$  will result in the creation of spurious details and artifacts in the map. Thus, the optimal filtration should be based on the  $SSNR$  distribution in the map and the solution is given by a Wiener filter:

$$W(u) = \frac{SSNR(u)}{SSNR(u) + 1}. \quad (2.5.7.24)$$

Based on the relation of  $FSC$  to  $SSNR$  (2.5.7.22), we can write (2.5.7.24) as

$$W(u) = \frac{2FSC(u)}{FSC(u) + 1}. \quad (2.5.7.25)$$

In practice, because of the irregular shape of typical  $FSC$  curves (particularly for small values of  $FSC$ ) it is preferable to approximate the shape of the Wiener filter (2.5.7.25) by one of the standard low-pass filters, such as Butterworth (Gonzalez & Woods, 2002) or hyperbolic tangent (Basokur, 1998).

The FRC/ $FSC$  methodology can be used to compare a noise-corrupted map with a noise-free ideal version of the same object. In single-particle reconstruction this situation emerges when an X-ray crystallographic structure of either the entire EM-determined structure or of some of its domains is available (Penczek *et al.*, 1999). In this case, we assume that in (2.5.7.19)  $f$  represents a sum of the signal and additive uncorrelated noise and  $g$  represents the noise-free signal, so is straightforward to calculate the expectation of (2.5.7.19) in order to obtain the relation between the cross-resolution ( $CRC$ ) and the  $SSNR$ :

$$E[CRC] \cong \left( \frac{SSNR}{SSNR + 1} \right)^{1/2}. \quad (2.5.7.26)$$

Thus

$$SSNR = \frac{CRC^2}{1 - CRC^2}. \quad (2.5.7.27)$$

Interestingly, for the same  $SSNR$  cut-off levels, corresponding values of  $CRC$  are higher than those for  $FSC$ . For example, for  $SSNR = 1$ ,  $CRC = 0.71$ , while  $FSC = 0.33$ . For  $SSNR = 2$ ,  $CRC = 0.82$ , while  $FSC = 0.5$ .

## 2.5.7.11. Analysis of 3D cryo-EM maps

The amount of structural information that can be derived from a structure of a macromolecular complex determined by cryo-EM depends on two factors: the resolution of the map and the availability of additional structural information about the system. Generally, we will refer to complexes at a resolution better than 7 Å as high-resolution structures, as at this resolution the

elements of secondary structure become directly visible. Maps at resolution lower than that we will call intermediate resolution, as at this scale of detail one can only determine a general arrangement of subunits. However, it is good to realize that there is a huge difference between the amount and reliability of information derived from a map of the same complex determined at 10 Å as compared to a map determined at 30 Å resolution. Similarly, very large complexes determined at 50 Å resolution will yield more information than very small complexes determined at 15 Å resolution. On the other hand, even intermediate-resolution EM maps provide extremely valuable information if they can be placed in the context of other structural work. The single-particle structure can be also investigated within a context of a more complex system using other, lower-resolution techniques, for example electron tomography. In this case, by using docking approaches one can determine the distribution, orientation and general arrangement of smaller cryo-EM determined complexes within larger subcellular systems. On a different scale of resolution, it is quite common to have structures of some domains or even of the entire complexes determined to atomic resolution by X-ray crystallography. Again, by using docking techniques it is possible to determine whether the conformation of the EM structure differs from that determined by X-ray crystallography or to map subunits and domains of the larger complex by fitting available atomic resolution structures.

The basic mode of visualization of cryo-EM maps is surface representation. The first step involves the choice of an appropriate threshold level for the displayed surface, particularly when the scaling of the cryo-EM data is arbitrary. A good guide is provided by the total molecular mass of the complex: given a pixel size of  $p$  Å, an average protein density  $d = 1.36 \times 10^{-24} \text{ g } \text{Å}^{-3}$  and the total molecular mass of the complex  $M$  Da, the number of voxels  $N_v$  occupied by the complex is

$$N_v = M / (p^3 d N_A), \quad (2.5.7.28)$$

where  $N_A$  is the Avogadro's number ( $6.02 \times 10^{23} \text{ atoms mole}^{-1}$ ). Based on that, one can find the threshold that for a given structure encompasses the determined number of voxels  $N_v$  [appropriate functions are implemented in *SPIDER* (Agrawal *et al.*, 1996; Frank *et al.*, 1996) and *SPARX* (Hohn *et al.*, 2007)]. At a sufficiently high resolution, cryo-EM maps can be analysed in the same manner as X-ray crystallographic maps and using the same graphical/analytical packages ('backbone tracing') (Jones *et al.*, 1991) (Fig. 2.5.7.6).

The complexity of cryo-EM maps of large macromolecular assemblies combined with their limited resolution invites attempts to automate some of the steps of analysis in an attempt to make the results more robust and less dependent on the researcher's bias. A good example of semi-automated analysis is the nucleic acid-protein separation in a 11.5 Å cryo-EM map of the 70S *E. coli* ribosome (Spahn *et al.*, 2000). In the procedure, the (continuous-valued) densities were analysed making use of (i) the difference in scattering density between protein and nucleic acids; (ii) continuity constraints that the image of any nucleic acid molecule must obey and (iii) knowledge of the molecular volumes of all proteins. As a result, it was possible to reproduce boundary assignments between ribosomal RNA (rRNA) and proteins made from higher-resolution X-ray maps of the ribosomal subunits with a high degree of accuracy, and allowed plausible predictions to be made for the placements of proteins and RNA components as yet unassigned. One of the conclusions derived from this separation was that the 23S rRNA is solely responsible for the catalysis of peptide-bond formation; thus, the ribosome is a ribozyme. The same conclusion was reached independently in the studies of the X-ray crystallographic structure of the 70S ribosome (Nissen *et al.*, 2000). The method by Spahn *et al.* cannot be easily extended to other

## 2.5. ELECTRON DIFFRACTION AND ELECTRON MICROSCOPY IN STRUCTURE DETERMINATION

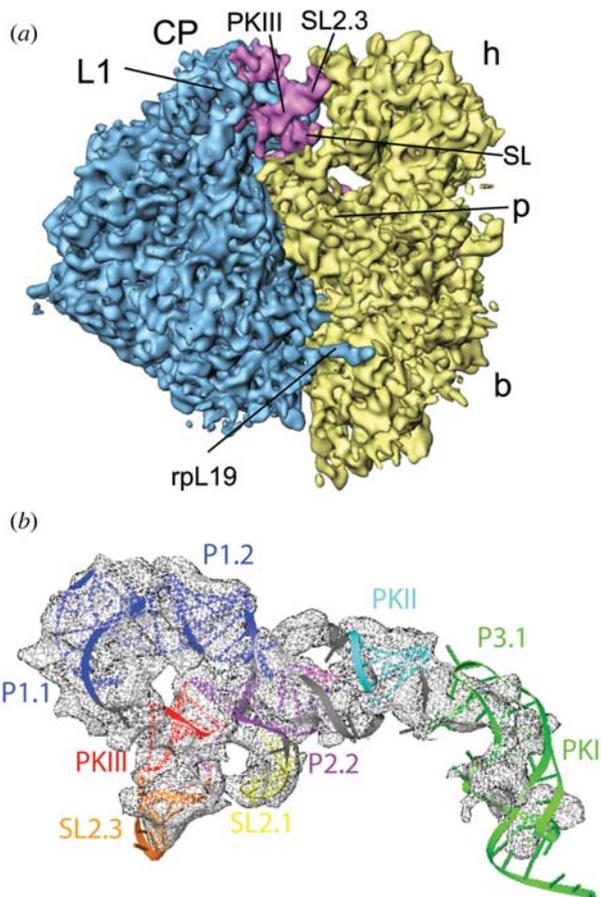


Fig. 2.5.7.6. A high-resolution cryo-EM map allows backbone tracing. (a) Cryo-EM map of the cricket paralysis virus (CrPV) IRES RNA in complex with the yeast 80S ribosome determined at 7.3 Å resolution. The map is shown from the L1 protuberance side with the ribosomal 40S subunit in yellow, the 60S subunit in blue and the CrPV IRES in magenta. Landmarks for the 40S subunit: b, body; h, head; p, platform. Landmarks for the 60S subunit: CP, central protuberance; L1, L1 protuberance. PKIII denotes helix PKIII of the CrPV IRES RNA and SL the two stem loops present in the secondary structure of the RNA. (b) Structure of the CrPV IRES RNA. Based on the cryo-EM map and additional biochemical knowledge, the complete chain of the RNA is found (189 nucleotides could be traced). The IRES molecular model is shown as a coloured ribbon docked into the cryo-EM density (grey mesh). PK and P denote the individual helical elements of the IRES and SL denotes the stem loops (Schüler *et al.*, 2006).

macromolecules that comprise only protein and generally it is very difficult to delineate at intermediate resolution subunits of large macromolecular assemblies, automatically or not, in the absence of independent knowledge about their shape. The reason is that both the density and shape of the subunits are affected by the limited resolution differently depending on their spatial context. In general, subunits that are isolated and located on the surface or protruding from the structure will have relatively lower density while at the same time their overall shape will be better preserved and easier to discern. Subunits located inside the structure and surrounded by other structural elements, while having higher density, are more difficult to recognize, as they fuse with the surrounding mass densities. Therefore, it is difficult to provide a general method that could cope with the problem of automated mass-density analysis.

As most cryo-EM structures are determined at intermediate resolution, the most common mode of analysis is either to compare the map with the available X-ray crystallographic structures of its domains or to consider the result in the context of larger, subcellular structures obtained by electron tomography. In both cases correlation techniques are used extensively to obtain objective results or to validate the results obtained by manual fitting.

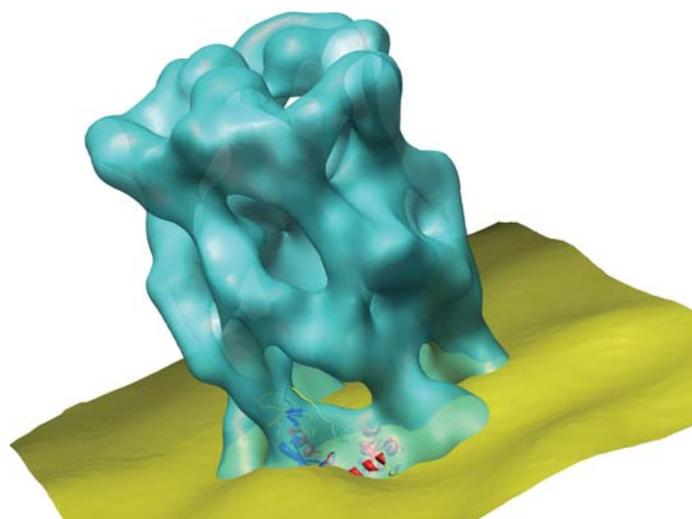


Fig. 2.5.7.7. Hrs (blue) embedded into the membrane (yellow) of an early endosome. Two functional domains have been docked into the cryo-EM density map of hexameric Hrs, determined to 16 Å resolution, and are shown as ribbons coloured by secondary structure (Pullan *et al.*, 2006). The structures of the VHS and FYVE domains have been determined crystallographically (Mao *et al.*, 2000) and are docked into the EM density map of the Hrs. The knowledge of the location of the FYVE and VHS domains, which are reported to bind to PI(3)P molecules found within the endosomal membrane (Kutateladze *et al.*, 1999), has guided the hypothetical placement of Hrs within the endosomal membrane. The immersion of the end caps of Hrs into the endosomal membrane demonstrates an 'end-on' binding model of the Hrs particles with the membrane. According to this model, either end cap can embed into the membrane, allowing the other end cap to carry out other essential protein trafficking functions, or to embed into another membrane, thus preventing fusion of membranes during early endosomal fusion.

In docking of X-ray crystallographic structures into EM maps, the first step is the conversion of atomic coordinates from X-ray molecular models, as given in Protein Data Bank (PDB) files, into an electron-density map in a way that would mimic the physical image formation process. Although sophisticated methods of computational emulation of the image-formation process in the electron microscope are available, very simple approaches to conversion yield quite satisfactory results at the resolution of the EM results. The most common one is to assume that the Coulomb potential of an atom is proportional to its atomic number and add these atomic numbers within a Euclidean grid with a cell size equal to the EM pixel size in Å. The atomic coordinates of atoms are interpolated within the grid using trilinear interpolation. After such conversion, the X-ray map can be handled using the general image-processing tools of a single-particle software package. Initial orientation (or orientations, if the general placement is not immediately visually apparent) of the X-ray map can be easily performed manually within any number of graphical packages, for example *Chimera* (Pettersen *et al.*, 2004). The initial six orientation parameters (three translations and three Eulerian angles) are next transferred to the EM package (for details see Baldwin & Penczek, 2007) and the manual docking is refined using correlation techniques (Fig. 2.5.7.7). Similarly, the handedness of the EM map can be established or confirmed by performing fitting of the X-ray determined structure to two EM maps that differ by their hand.

Docking of EM maps into the broader cellular context of structures determined by electron tomography can provide information about the distribution of complexes and their interactions within the cell. Conceptually, the approach is very similar to that of particle picking, *i.e.*, template matching, with the main difference being that calculations are performed in three instead of two dimensions. Given a 3D structure of a single-particle EM complex, a set of 3D templates is prepared by rotating the

## 2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

template around its centre of mass using the quasi-uniformly distributed three Eulerian angles [equation (2.5.7.17) with  $\Delta\psi = \delta\theta$ ]. However, in application to tomography the angular step  $\delta\theta$  can be relatively large, resulting in a much smaller number of templates than in two dimensions, the reason being the rather low resolution of typical electron tomograms (not exceeding 50 Å). Next, a brute-force 3D cross-correlation search with all templates is performed (Frangakis *et al.*, 2002). After windowing out 3D subvolumes containing putative complexes, subsequent averaging and classification can be performed.

Cryo-EM is a unique structural technique in its ability to detect conformational variability of large molecular assemblies within one sample that may contain a mixture of complexes in various conformational states. In addition to the expected conformational heterogeneity of the assemblies, due to fluctuations of the structure around the ground state one can expect to capture molecules in different functional states, especially if the binding of a ligand induces a conformational change in the macromolecular assembly. Therefore, a data set of images from an EM experiment must be interpreted as a mixture of projections from similar but not identical structures. The analysis of the extent of the resulting variability requires the calculation of the real-space distribution of 3D variance/covariance in macromolecules reconstructed from a set of their projections. The problem is difficult, as there is no clear relation between the variance in sets of projections that have the same angular direction and the variance of the 3D structure calculated from these projections. Penczek, Chao *et al.* (2006) proposed calculating the variance in the 3D mass distribution of the structure using a statistical bootstrap resampling technique, in which a new set of projections is selected with replacements from the available whole set of  $N$  projections. In the new set, some of the original projections will appear more than once, while others will be omitted. This selection process is repeated a number of times and for each new set of projections the corresponding 3D volume is calculated. Next, the voxel-by-voxel bootstrap variance  $\sigma_B^2$  of the resulting set of volumes is calculated. The target variance is obtained using a relationship between the variance of arithmetic means for sampling with replacements and the sample variance,

$$\sigma^2 = N\sigma_B^2. \quad (2.5.7.29)$$

The estimated structure-variance map can be used for (i) detection of different functional states (for example, those characterized by binding of a ligand) and subsequent classification of the data set into homogeneous groups (Penczek, Frank & Spahn, 2006a), (ii) analysis of the significance of small details in 3D reconstructions, (iii) analysis of the significance of details in difference maps, and (iv) docking of known structural domains into EM density maps.

The bootstrap technique also leads to the analysis of conformational modes of macromolecular complexes, and this is due to the fact that the covariance matrix of the structure can be directly calculated from the bootstrap volumes. The covariance matrix obtained this way would be very large. One possibility is to calculate only correlation coefficients between regions of interest that have large variance (Penczek, Chao *et al.*, 2006). Another possibility is to use the iterative Lanczos technique (Parlett, 1980) and calculate eigenvolumes directly from bootstrap volumes without forming the covariance matrix. These eigenvolumes are related to conformational modes of the molecule, as captured by the projection data of the sample (Penczek, Frank & Spahn, 2006b). Thus, this direct relation to the actual cryo-EM projection data positively distinguishes this approach from other techniques in which conformations are postulated based on flexible models of the EM map (Ming *et al.*, 2002; Mitra *et al.*, 2005).

### 2.5.8. Direct phase determination in electron crystallography

By D. L. DORSET

#### 2.5.8.1. Problems with 'traditional' phasing techniques

The concept of using experimental electron-diffraction intensities for quantitative crystal structure analyses has already been presented in Section 2.5.4. Another aspect of quantitative structure analysis, employing high-resolution images, has been presented in Sections 2.5.5 to 2.5.7. That is to say, electron micrographs can be regarded as an independent source of crystallographic phases.

Before direct methods (Chapter 2.2) were developed as the standard technique for structure determination in small-molecule X-ray crystallography, there were two principal approaches to solving the crystallographic phase problem. First, 'trial and error' was used, finding some means to construct a reasonable model for the crystal structure *a priori*, e.g. by matching symmetry properties shared by the point group of the molecule or atomic cluster and the unit-cell space group. Secondly, the autocorrelation function of the crystal, known as the Patterson function (Chapter 2.3), was calculated (by the direct Fourier transform of the available intensity data) to locate salient interatomic vectors within the unit cell.

The same techniques had been used for electron-diffraction structure analysis (nowadays known as *electron crystallography*). In fact, advocacy of the first method persists. Because of the perturbations of diffracted intensities by multiple-beam dynamical scattering (Chapter 5.2), it has often been suggested that trial and error be used to construct the scattering model for the unit crystal in order to test its convergence to observed data after simulation of the scattering events through the crystal. This indirect approach assumes that no information about the crystal structure can be obtained directly from observed intensity data. Under more favourable scattering conditions nearer to the kinematical approximation, *i.e.* for experimental data from thin crystals made up of light atoms, trial-and-error modelling, simultaneously minimizing an atom-atom nonbonded potential function with the crystallographic residual, has enjoyed widespread use in electron crystallography, especially for the determination of linear polymer structures (Brisse, 1989; Pérez & Chanzy, 1989).

Interpretation of Patterson maps has also been important for structure analysis in electron crystallography. Applications have been discussed by Vainshtein (1964), Zvyagin (1967) and Dorset (1994a). In face of the dynamical scattering effects for electron scattering from heavy-atom crystals realized later (*e.g.* Cowley & Moodie, 1959), attempts had also been made to modify this autocorrelation function by using a power series in  $|F_h|$  to sharpen the peaks (Cowley, 1956). (Here  $F_h \equiv \Phi_h$ , replacing the notation for the kinematical electron-diffraction structure factor employed in Section 2.5.4.) More recently, Vincent and co-workers have selected first-order-Laue-zone data from inorganics to minimize the effect of dynamical scattering on the interpretability of their Patterson maps (Vincent & Exelby, 1991, 1993; Vincent & Midgley, 1994). Vainshtein & Klechkovskaya (1993) have also reported use of the Patterson function to solve the crystal structure of a lead soap from texture electron-diffraction intensity data.

It is apparent that trial-and-error techniques are most appropriate for *ab initio* structure analysis when the underlying crystal structures are reasonably easy to model. The requisite positioning of molecular (or atomic) groups within the unit cell may be facilitated by finding atoms that fit a special symmetry position [see *IT A* (2005)]. Alternatively, it is helpful to know the molecular orientation within the unit cell (*e.g.* provided by the Patterson function) to allow the model to be positioned for a conformational or translational search. [Examples would include