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 $\sigma_{\rm B}^2$ 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,

$$\sigma^2 = N\sigma_{\rm B}^2. \qquad (2.5.7.29)$$

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, 2006*a*), (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, 2006*b*). 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 (1994*a*). 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 $|F_{\mathbf{h}}|$ to sharpen the peaks (Cowley, 1956). (Here $F_{\mathbf{h}} \equiv \Phi_{\mathbf{h}}$, replacing the notation for the kinematical electron-diffraction structure factor employed in Section 2.5.4.) More recently, Vincent and coworkers 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

the polymer-structure analyses cited above, as well as the layer-packing analysis of some phospholipids (Dorset, 1987).] While attempts at *ab initio* modelling of three-dimensional crystal structures, by searching an n-dimensional parameter space and seeking a global internal energy minimum, has remained an active research area, most success so far seems to have been realized with the prediction of two-dimensional layers (Scaringe, 1992). In general, for complicated unit cells, determination of a structure by trial and error is very difficult unless adequate constraints can be placed on the search.

Although Patterson techniques have been very useful in electron crystallography, there are also inherent difficulties in their use, particularly for locating heavy atoms. As will be appreciated from comparison of scattering-factor tables for X-rays [*IT* C (2004), Chapter 6.1] with those for electrons [*IT* C (2004), Chapter 4.3], the relative values of the electron form factors are more compressed with respect to atomic number than are those for X-ray scattering. As discussed in Chapter 2.3, it is desirable that the ratio of summed scattering-factor terms, $r = \sum_{\text{heavy}} Z^2 / \sum_{\text{light}} Z^2$, where Z is the scattering-factor value at $\sin\theta/\lambda = 0$, be near unity. A practical comparison would be the value of r for copper (DL-alaninate) solved from electron-diffraction data by Vainshtein *et al.* (1971). For electron diffraction, $r = 0.47$ compared to the value 2.36 for X-ray diffraction. Orientation of salient structural features, such as chains and rings, would be equally useful for light-atom moieties in electron or X-ray crystallography with Patterson techniques. As structures become more complicated, interpretation of Patterson maps becomes more and more difficult unless an automated search can be carried out against a known structural fragment (Chapter 2.3).

2.5.8.2. Direct phase determination from electron micrographs

The 'direct method' most familiar to the electron microscopist is the high-resolution electron micrograph of a crystalline lattice. Retrieval of an average structure from such a micrograph assumes that the experimental image conforms adequately to the 'weak phase object' approximation, as discussed in Section 2.5.5. If this is so, the use of image-averaging techniques, *e.g.* Fourier filtration or correlational alignment, will allow the unit-cell contents to be visualized after the electron-microscope phase contrast transfer function is deconvoluted from the average image, also discussed in Section 2.5.5. Image analyses can also be extended to three dimensions, as discussed in Section 2.5.6, basically by employing tomographic reconstruction techniques to combine information from the several tilt projections taken from the crystalline object. The potential distribution of the unit cell to the resolution of the imaging experiment can then be used, *via* the Fourier transform, to obtain crystallographic phases for the electron-diffraction amplitudes recorded at the same resolution. This method for phase determination has been the mainstay of protein electron crystallography.

Once a set of phases is obtained from the Fourier transform of the deconvoluted image, they must, however, be referred to an allowed crystallographic origin. For many crystallographic space groups, this choice of origin may coincide with the location of a major symmetry element in the unit cell [see *IT* A (2005)]. Hence, since the Fourier transform of translation is a phase term, if an image shift $[\delta(\mathbf{r} + \mathbf{r}_0)]$ is required to translate the origin of the repeating mass unit $\varphi(\mathbf{r})$ from the arbitrary position in the image to a specific site allowed by the space group,

$$\mathrm{g}(\mathbf{r}) = \varphi(\mathbf{r}) \otimes \delta(\mathbf{r} + \mathbf{r}_0) = \varphi(\mathbf{r} + \mathbf{r}_0),$$

where the operation '$\otimes$' denotes convolution. The Fourier transform of this shifted density function will be

$$G(\mathbf{s}) = F(\mathbf{s}) \exp(2\pi i \mathbf{s} \cdot \mathbf{r}_0) = |F(\mathbf{s})| \exp[i(\phi_s + 2\pi i \mathbf{s} \cdot \mathbf{r}_0)].$$

In addition to the crystallographic phases ϕ_s, it will, therefore, be necessary to find the additional phase-shift term $2\pi i \mathbf{s} \cdot \mathbf{r}_0$ that will access an allowed unit-cell origin. Such origin searches are carried out automatically by some commercial image-averaging computer-software packages.

In addition to applications to thin protein crystals (*e.g.* Henderson *et al.*, 1990; Jap *et al.*, 1991; Kühlbrandt *et al.*, 1994), there are numerous examples of molecular crystals that have been imaged to a resolution of 3–4 Å, many of which have been discussed by Fryer (1993). For π-delocalized compounds, which are the most stable in the electron beam against radiation damage, the best results (2 Å resolution) have been obtained at 500 kV from copper perchlorophthalocyanine epitaxically crystallized onto KCl. As shown by Uyeda *et al.* (1978–1979), the averaged images clearly depict the positions of the heavy Cu and Cl atoms, while the positions of the light atoms in the organic residue are not resolved. (The utility of image-derived phases as a basis set for phase extension will be discussed below.) A number of aromatic polymer crystals have also been imaged to about 3 Å resolution, as reviewed recently (Tsuji, 1989; Dorset, 1994*b*).

Aliphatic molecular crystals are much more difficult to study because of their increased radiation sensitivity. Nevertheless, monolamellar crystals of the paraffin n-$C_{44}H_{90}$ have been imaged to 2.5 Å resolution with a liquid-helium cryomicroscope (Zemlin *et al.*, 1985). Similar images have been obtained at room temperature from polyethylene (Revol & Manley, 1986) and also a number of other aliphatic polymer crystals (Revol, 1991).

As noted by J. M. Cowley and J. C. H. Spence in Section 2.5.1, dynamical scattering can pose a significant barrier to the direct interpretation of high-resolution images from many inorganic materials. Nevertheless, with adequate control of experimental conditions (limiting crystal thickness, use of high-voltage electrons) some progress has been made. Pan & Crozier (1993) have described 2.0 Å images from zeolites in terms of the phase-grating approximation. A three-dimensional structural study has been carried out on an aluminosilicate by Wenk *et al.* (1992) with thin samples that conform to the weak-phase-object approximation at the 800 kV used for the imaging experiment. Heavy and light (*e.g.* oxygen) atoms were located in the micrographs in good agreement with an X-ray crystal structure. Heavy-atom positions from electron microscopic and X-ray structure analyses have also been favourably compared for two heavy-metal oxides (Hovmöller *et al.*, 1984; Li & Hovmöller, 1988).

2.5.8.3. Probabilistic estimate of phase invariant sums

Conventional direct phasing techniques, as commonly employed in X-ray crystallography (*e.g.* see Chapter 2.2), have also been used for *ab initio* electron-crystallographic analyses. As in X-ray crystallography, probabilistic estimates of a linear combination of phases (Hauptman & Karle, 1953; Hauptman, 1972) are made after normalized structure factors are calculated *via* electron form factors, *i.e.*

$$|E_{\mathbf{h}}^2| = I_{\text{obs}} / \varepsilon \sum_i f_i^2, \text{ where } \langle |E|^2 \rangle = 1.000.$$

(Here, an overall temperature factor can be found from a Wilson plot. Because of multiple scattering, the value of B may be found occasionally to lie close to 0.0 Å^2.) The phase invariant sums

$$\psi = \phi_{\mathbf{h}_1} + \phi_{\mathbf{h}_2} + \phi_{\mathbf{h}_3} + \ldots$$