interactions with the rest of the structure and, following refinement of the structure, by elimination of the peak in the difference map and by a significant improvement in the agreement between the calculated and measured X-ray amplitudes.

4.5.2.6.5. Difference Fourier synthesis

Difference Fourier syntheses are widely used in both protein and small-molecule crystallography to detect structural errors or to complete partial structures (Drenth, 1994). The difficulty in applying difference Fourier techniques in fibre diffraction is that the individual observed amplitudes $|F_o|$ are not available. However, difference syntheses have found wide use in fibre diffraction analysis, one of the earliest applications being to polycrystalline fibres of polynucleotides (*e.g.* Arnott *et al.*, 1967). Calculation of a three-dimensional difference map (for the unit cell) from Bragg fibre diffraction data requires that the observed intensity $I_l(R_{hk}) = I_o$ be apportioned among the contributing intensities $|F_{hkl}|^2 = |F_o|^2$. There are two ways of doing this. The intensities may be divided equally among the contributing $(m/2)$ reflections [*i.e.* $|F_o| = (2I_o/m)^{1/2}$], or they may be divided in the same proportions as those in the model, *i.e.*

$$|F_o| = \left(\frac{I_o}{\sum|F_c|^2}\right)^{1/2}|F_c|. \qquad (4.5.2.63)$$

The advantage of the former is that it is unbiased, and the advantage of the latter is that it may be more accurate but is biased towards the model. Equal division of the intensities is often (but not always) used to minimize model bias. Once the observed amplitudes have been apportioned, an $|F_o| - |F_c|$ map can be calculated as in conventional crystallography, although noise levels will be higher owing to errors in apportioning the amplitudes. As a result of overlapping of the reflections, a synthesis based on coefficients $m|F_o| - (m-1)|F_c|$ gives a more accurate estimate of the true density than does one based on $2|F_o| - |F_c|$, as is described below. Difference syntheses for polycrystalline specimens calculated in this way have been used, for example, to locate cations and water molecules in polynucleotide and polysaccharide structures (*e.g.* Cael *et al.*, 1978), to help position molecules in the unit cell (*e.g.* Chandrasekaran *et al.*, 1994) and to help position side chains, and have also been applied in neutron fibre diffraction studies of polynucleotides (Forsyth *et al.*, 1989).

Sim (1960) has shown that the mean-squared error in difference syntheses can be minimized by weighting the coefficients based on the agreement between the calculated and observed structure amplitudes. Such an analysis has recently been conducted for fibre diffraction, and shows that the optimum difference synthesis is obtained by using coefficients (Millane & Baskaran, 1997; Baskaran & Millane, 1999*a*)

$$\left[w_m\frac{|F_c|(I_o)^{1/2}}{(\sum|F_c|^2)^{1/2}} - |F_c|\right]\exp(i\alpha_c), \qquad (4.5.2.64)$$

where m is the number of degrees of freedom as defined in Section 4.5.2.6.1. If the reflections contributing to I_o are either all centric or all acentric, then the weights are given by

$$w_m = \frac{I_{m/2}(X)}{I_{m/2-1}(X)}, \qquad (4.5.2.65)$$

where $I_m(\cdot)$ denotes the modified Bessel function of the first kind of order m, and X is given by

$$X = \frac{\kappa(I_o)^{1/2}(\sum|F_c|^2)^{1/2}}{\sum_j f_j^2}, \qquad (4.5.2.66)$$

where $\kappa = 1$ for centric reflections and 2 for acentric reflections. The form of the weighting function is more complicated if both centric and acentric reflections contribute, but it can be approximated as w' given by

$$w' = (w_{2N_a} + w_{N_c})/2, \qquad (4.5.2.67)$$

where N_a and N_c are the number of acentric and centric reflections, respectively, contributing. Use of the weighted maps reduces bias towards the model (Baskaran & Millane, 1999*b*).

For continuous diffraction data from noncrystalline specimens, the situation is essentially identical except that one works in cylindrical coordinates. Referring to equations (4.5.2.7) and (4.5.2.10), the desired difference synthesis, $\Delta g(r, \varphi, z)$, is the Fourier–Bessel transform of $G_o - G_c$ where G_o and G_c denote the observed and calculated, respectively, Fourier–Bessel structure factors $G_{nl}(R)$. Since G_o is not known, the synthesis is based on the Fourier–Bessel transform of $(|G_o| - |G_c|)\exp(i\alpha_c)$, where α_c is the phase of G_c. As in the polycrystalline case, the individual $|G_o|$ need to be estimated from the data $I_o^{1/2}$ given by equation (4.5.2.17), and can be based on either equal division of the data, or division in the same proportion as the amplitudes from the model.

Namba & Stubbs (1987*a*) have shown that the peak heights in a difference synthesis are $1/m$ times their true value, as opposed to half their true value in a conventional difference synthesis. The best estimate of the true map is therefore provided by a synthesis based on the coefficients $[m|F_o| - (m-1)|F_c|]\exp(i\alpha_c)$, rather than on $(2|F_o| - |F_c|)\exp(i\alpha_c)$. Test examples showed that the noise in the synthesis can be reduced by using a value for m that is fixed over the diffraction pattern and approximately equal to the average value of m over the pattern (Namba & Stubbs, 1987*a*). Difference Fourier maps for noncrystalline systems have been used in studies of helical viruses to locate heavy atoms, to correct errors in atomic models and to locate water molecules (Mandelkow *et al.*, 1981; Lobert *et al.*, 1987; Namba, Pattanayek & Stubbs, 1989; Wang & Stubbs, 1994).

4.5.2.6.6. Multidimensional isomorphous replacement

At low enough resolution, only one Fourier–Bessel structure factor contributes on each layer line of a fibre diffraction pattern, so that only the phase needs to be determined and the situation is no different to that in protein crystallography. If heavy-atom-derivative specimens can be prepared, the usual method of multiple isomorphous replacement (MIR) (Drenth, 1994) can be applied, which in principle requires only two heavy-atom derivatives. At higher resolution, however, more than one Fourier–Bessel structure factor contributes on each layer line. A generalized form of isomorphous replacement which involves using diffraction data from several heavy-atom derivatives to determine the real and imaginary components of each contributing $G_{nl}(R)$ is referred to as *multidimensional isomorphous replacement* (MDIR) (Namba & Stubbs, 1985). MDIR was first described and used to determine the structure of TMV at 6.7 Å resolution (Stubbs & Diamond, 1975; Holmes *et al.*, 1975), and has since been used to extend the resolution to 2.9 Å (Namba, Pattanayek & Stubbs, 1989). A consequence of cylindrical averaging is that large numbers of heavy-atom derivatives are required: at least two for each Bessel term to be separated. The theory of MDIR is outlined here.

The first step in MDIR is location of the heavy atoms in the derivative structures. The radial coordinate of a heavy atom can be determined by analysis of the intensity distribution in the low-

resolution region of the equator where only the $G_{00}(R)$ Bessel term contributes. Since $G_{00}(R)$ is real, and $I_l(R)$ can be measured continuously in R, inspection of the positions of the minima and maxima in the low-resolution region of the equator generally allows the sign of $G_{00}(R)$ to be assigned to $I_0^{1/2}(R)$, *i.e.* $G_{00}(R)$ can be determined from $I_0(R)$. If the sign is determined for both the native and a heavy-atom derivative, referring to equation (4.5.2.13) shows that

$$G_{00}^{\rm D}(R) - G_{00}(R) = o_h f_h J_0(2\pi R r_h), \tag{4.5.2.68}$$

where $G_{00}^{\rm D}(R)$ is the value derived from the derivative data, o denotes the occupancy and the subscript h denotes values for the heavy atom. The parameters o_h and r_h on the right-hand side of equation (4.5.2.68) can be searched in a trial-and-error fashion to obtain the best agreement with the left-hand side (calculated from the data) to determine the radial coordinate r_h of the heavy atom (Mandelkow & Holmes, 1974). Lobert *et al.* (1987) applied the same method to cucumber green mottle mosaic virus (CGMMV), except that the sign of $G_{00}(R)$ was taken from that of TMV.

Two approaches have been used to determine the angular and axial coordinates of the heavy atom. Mandelkow & Holmes (1974) and Holmes *et al.* (1975) used a search procedure in which the quantity $\Phi = -n\varphi_h + 2\pi l z_h/c$ is varied and used to calculate the intensity of the Fourier–Bessel structure factor for the heavy atom alone. This is compared to $I_l^{\rm D}(R) - I_l(R)$ on each layer line, where only one Bessel order contributes, and Φ chosen to minimize the mean-square difference. The values of Φ found for each layer line can then be combined to determine φ_h and z_h. In the case of CGMMV, Lobert *et al.* (1987) used the phases and Bessel-order separations from TMV to calculate Fourier–Bessel difference maps between the native and derivative data to determine the heavy-atom coordinates (r_h, φ_h, z_h).

Consider a set of isomorphous heavy-atom derivatives indexed by j. Since the analysis is applied at any point (l, R) on the fibre diffraction pattern, the symbol G_n will be used for $G_{nl}(R)$ where no confusion arises. Denote by $G_{n,j}$ the value of G_n for the jth derivative, so that

$$G_{n,j} = G_n + g_{n,j}, \tag{4.5.2.69}$$

where $g_{n,j}$ denotes the Fourier–Bessel structure factor of a structure containing the heavy atom only. Denote by A_n and B_n the real and imaginary parts, respectively, of G_n (for the native structure), and by $a_{n,j}$ and $b_{n,j}$ the real and imaginary parts of $g_{n,j}$, *i.e.* for the jth heavy-atom structure alone. Equation (4.5.2.17) can then be written as

$$I = \sum_n (A_n^2 + B_n^2) \tag{4.5.2.70}$$

for the native and

$$I_j = \sum_n [(A_n + a_{n,j})^2 + (B_n + b_{n,j})^2] \tag{4.5.2.71}$$

for the jth derivative. If intensity data are available from J heavy-atom derivatives, $a_{n,j}$ and $b_{n,j}$ can be calculated from the heavy-atom positions, and equations (4.5.2.70) and (4.5.2.71) represent a system of $J+1$ second-order equations for the m unknowns A_n and B_n. If $J+1 > m$, then the system of equations is over-determined and can be solved for the A_n and B_n. The solution of this nonlinear system can be eased by deriving a system of linear equations by substituting from (4.5.2.70) into (4.5.2.71), giving

$$\sum_n (A_n a_{n,j} + B_n b_{n,j}) = (1/2)\left[I_j - I - \sum_n (a_{n,j}^2 + b_{n,j}^2)\right]. \tag{4.5.2.72}$$

Equation (4.5.2.72) is a system of linear equations for the unknowns A_n and B_n, the solution being subject to the constraint equation (4.5.2.70). However, since the original problem is second-order, there may be up to m local minima. Stubbs & Diamond (1975) describe a numerical procedure for locating *all* the local minima and selecting the best of these based on 'continuity' of the $G_{nl}(R)$. This method was used to determine the structure of TMV at 6.7 Å resolution (Holmes *et al.*, 1975) and 4 Å resolution (Stubbs *et al.*, 1977). In current applications of MDIR a more direct solution technique is used in which the phase-determining equations (4.5.2.70) and (4.5.2.71) are solved by first solving the linear equations (4.5.2.72) by linear least squares to obtain an approximate solution, which is then refined by solving the quadratic equations (4.5.2.70) and (4.5.2.71) directly using nonlinear least squares (Namba & Stubbs, 1985).

The number of heavy-atom derivatives required can be quite demanding experimentally, although phasing with fewer heavy-atom derivatives is possible, particularly if additional information is available, such as from a related structure. The different Bessel terms may be assumed to contribute the same amplitude each, or, if the structure of a related molecule is known, the ratios of the amplitudes can be taken as being the same as those for the related molecule. Using the amplitude estimates derived using either of these two approaches, applied to both native and derivative data, the phases of the Bessel terms can be estimated using conventional MIR and data from at least two heavy-atom derivatives, allowing an initial electron-density map to be calculated. If only one heavy-atom derivative is available then two phase solutions are obtained, but the method of conventional single isomorphous replacement (SIR) (Drenth, 1994) can be used to obtain an estimate of the electron density. The electron density obtained by MIR, and particularly by SIR, in this way tends to be noisy and low contrast as a result of inaccurate division of the intensities, as well as the usual sources of errors in MIR. The electron density can, however, be improved using solvent levelling. If *no* heavy-atom derivatives are available, both the relative amplitudes *and* the phases can be based on those of a related structure. Model bias can, however, be more serious than in conventional crystallography since both the phases and the relative amplitudes are based on the model.

The feasibility of structure determination with a limited number of heavy-atom derivatives was first demonstrated by Namba & Stubbs (1987*b*) using data from TMV at 4 Å resolution. The structure of CGMMV has been determined at 5 Å resolution using data from two heavy-atom derivatives and the techniques described above (Lobert *et al.*, 1987; Lobert & Stubbs, 1990). Structure determination at this resolution using MDIR would theoretically require six heavy-atom derivatives. Initial separation of the Bessel-term amplitudes was based on the equal-amplitude assumption and also on the relative amplitudes for (homologous) TMV.

In general, the equal-amplitude assumption appears to produce reliable electron-density maps where only two or three Bessel terms contribute. The corresponding resolution depends on the helix symmetry and the molecular diameter, but can be relatively high for molecules with high helix symmetry. At higher resolution where more Bessel terms contribute, use of related or partial structures can be used to calculate initial Bessel-term amplitudes and can lead to successful phasing.

If the molecule has only approximate helix symmetry, then layer-line splitting (Section 4.5.2.3.3) can provide additional information which reduces the number of heavy-atom derivatives required. The degree of splitting is usually significantly less than

the breadth of the layer lines so that the different Bessel terms within a (split) layer line overlap. The effect of splitting can be observed, however, since the centre of a layer line, at a particular value of R, is shifted towards the position of the stronger Bessel term contributing at that radius. The shift depends on the relative magnitudes of the contributing Bessel terms, and can be measured and used in phase determination as detailed by Stubbs & Makowski (1982). If P of the heavy-atom derivatives (in addition to the native) give accurate splitting information, then an additional P linear equations [analogous to equation (4.5.2.72)] and one quadratic equation [analogous to equation (4.5.2.70)] are available for solution of the phase problem, and the number of heavy-atom derivatives required is reduced by a factor of up to two. The value of layer-line splitting was first demonstrated by recalculating an electron-density map of TMV at 6.7 Å resolution using only two derivatives, rather than using six derivatives without the use of splitting data (Stubbs & Makowski, 1982). Layer-line splitting was subsequently used in a structure determination of TMV at 3.6 Å resolution (Namba & Stubbs, 1985).

Macromolecular fibre structures that have been built into an electron-density map have been refined using both restrained least-squares (RLS) and molecular-dynamics (MD) refinements. Restrained least squares has been used to refine the structure of TMV at 2.9 Å resolution (Namba, Pattanayek & Stubbs, 1989); however, Wang & Stubbs (1993) have shown that a larger radius of convergence is obtained using MD refinement (as in protein crystallography).

Molecular-dynamics refinement in fibre diffraction has been implemented by adding a fibre diffraction option (Wang & Stubbs, 1993) to the *X-PLOR* program (Brünger, 1992). This involves including the cylindrically averaged fibre diffraction intensities in the energy term and taking account of the inter-helical subunit contacts and covalent connections in the same way as described above for RLS refinement. The effective potential-energy function E used is

$$E = E_{\rm e} + S\sum_l \sum_i w_{li}\{[I_l^{\rm o}(R_i)]^{1/2} - k[I_l^{\rm c}(R_i)]^{1/2}\}^2, \qquad (4.5.2.73)$$

where $E_{\rm e}$ is the empirical energy function (which typically includes bond-length, bond-angle and torsion-angle distortions, van der Waals and electrostatic interactions, and other terms such as ring planarity), $I_l^{\rm o}(R_i)$ and $I_l^{\rm c}(R_i)$ are the observed and calculated, respectively, cylindrically averaged diffraction intensities sampled at $R = R_i$, the w_{li} are weights for the observed intensities $I_l^{\rm o}(R_i)$ and k is a scale factor between the calculated and observed data. The quantity S is a weight to make the gradients of the two terms in equation (4.5.2.73) comparable (Wang & Stubbs, 1993), and can be estimated using the method of Brünger (1992). Molecular-dynamics refinement has been successfully used to refine the structure of CGMMV at 3.4 Å resolution (Wang & Stubbs, 1994). In the case of ribgrass mosaic virus (RMV), the close isomorphism with TMV (identical helix symmetry, similar repeat distance, significant sequence homology and similar diffraction pattern) allowed an initial model to be built based on the TMV structure, and a solution obtained at 2.9 Å by alternating molecular-dynamics refinement with difference-map and omit-map calculations (Wang *et al.*, 1997).

4.5.2.6.7. Other techniques

Aside from the techniques for structure determination described in the previous sections, a variety of other techniques have been applied to specific problems where the methods described above are not suitable. This situation usually arises where the diffraction data available are far too few, by themselves, to determine the individual atomic coordinates of a structure, even with the usual stereochemical constraints. Often only relatively low-resolution data are available, but they can be supplemented by either a low-resolution or high-resolution model of either a whole molecule or relatively large subunits. Structure determination often amounts to positioning the molecules or subunits within a larger assembly. The results can be quite precise, depending on the information available. The problem is almost always one of refinement or optimization, since it invariably involves optimizing some kind of model directly against the fibre diffraction data. The problem is usually twofold: (1) parameterizing the model with few enough parameters to obtain a usable data-to-parameter ratio, but retaining enough degrees of freedom to represent the important structural features; and (2) devising an optimization procedure that will locate the global minimum of the resulting complicated cost function. There have been numerous such applications in fibre diffraction, and rather than attempt to be exhaustive or detailed, I will briefly mention a few of the more prominent applications and techniques.

The structure of the bacteriophage Pf1 was determined at 7 Å resolution using a model in which the α-helical segments of the structure were represented by rods of electron density of appropriate dimensions and spacings (Makowski *et al.*, 1980). The positions and orientations of the rods were refined in an iterative procedure that alternated between real space and reciprocal space and also incorporated solvent levelling. Neutron fibre diffraction data have been collected from specifically deuterated phages and, starting with a model of the kind described above, iterative application of difference maps (between the deuterated and native data) was used to locate 15 (of the 46) residues, allowing construction of a model of the coat protein (Stark *et al.*, 1988; Nambudripad *et al.*, 1991).

Pf1 undergoes a temperature-induced structural transition that involves a small change in the helix symmetry. The low-temperature form has 71_{13} helix symmetry with a c repeat of 216.5 Å, and the high-temperature form (that discussed in the previous paragraph) has 27_5 helix symmetry and a c repeat of 78.3 Å. These two symmetries are very similar since $71/3 \simeq 27/5$ and $216.5/71 \simeq 78.3/27$, *i.e.* the rotations and translations from one subunit to the next are very similar in both structures.

The structure of the low-temperature form of Pf1 has been determined at 3.3 Å resolution by starting with an α-helical polyalanine model (Marvin *et al.*, 1987) and alternating rounds of molecular-dynamics refinement and model rebuilding based on $(2F_{\rm o} - F_{\rm c})$ maps and omit maps (Gonzalez *et al.*, 1995). The structure of the high-temperature form of Pf1 was determined using data to 3 Å resolution, starting with a model based on the low-temperature form, making small adjustments to satisfy the slightly different helix symmetry, and refining the model using molecular dynamics (Welsh *et al.*, 2000).

The bacteriophage Pf3 is related to Pf1 but does not undergo a structural transition, and fibre diffraction patterns are similar to those from the high-temperature form of Pf1. An α-helical polyalanine model of Pf3 based on the Pf1 structure was used to separate and phase the Bessel terms, which were then used to calculate $(5F_{\rm o} - 4F_{\rm c})$ maps. These maps were used to align and position the polypeptide chain, and the resulting model was refined by molecular dynamics (Welsh *et al.*, 1998).

The R-type bacterial flagellar filament structure (that has a very high molecular weight subunit) has been determined at 9 Å resolution by X-ray fibre diffraction (Yamashita *et al.*, 1998). Accurate intensities were taken from high-quality X-ray diffraction patterns and combined with phases obtained from electron cryomicroscopy, and solvent levelling was used to refine the phases.

Some studies of muscle provide a good example of the use of low-resolution fibre diffraction data, coupled with high-resolution crystal structures of some of the component molecules, to determine the structure of a complex. Holmes *et al.* (1990) constructed a model of F-actin based on the crystal structure of