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

the monomer, G-actin, and 8 Å fibre diffraction data, by either treating the monomer as a rigid body or dividing it into four separate rigid domains, and using a search procedure followed by least-squares refinement. The results gave the orientation of the actin monomer in the actin helix. This structure has since been refined using a genetic algorithm (Lorenz *et al.*, 1993) and normal-mode analysis (Tirion *et al.*, 1995). The genetic algorithm involved a Monte Carlo method of selecting subdomains to be refined and nonlinear least squares to obtain the best fit for the selected domains. In the normal-mode analysis, the model was parameterized in terms of its low-frequency vibrational modes to allow low-energy conformational changes and reduce the number of parameters which were optimized against the fibre diffraction data using nonlinear least squares.

Squire *et al.* (1993) have refined a low-resolution model of the muscle thin-filament structure that consists of four spheres representing each of the F-actin monomer subdomains and five spheres (fixed relative to each other) representing tropomyosin. Steric restraints were placed on the actin subdomain and thin-filament structures. The positions of the actin subdomains and the orientation of the tropomyosin were refined using a search procedure against fibre diffraction data from both 'resting' and 'activated' muscle at 25 Å resolution. More recent work has used a low-resolution model of the myosin head (based on the single-crystal atomic structure), a search procedure and simulated-annealing refinements to study myosin head configuration (Hudson *et al.*, 1997) and myosin rod packing (Squire *et al.*, 1998).

4.5.2.6.8. Reliability

As with structure determination in any area of crystallography, assessment of the reliability or precision of a structure is critically important. The most commonly used measure of reliability in fibre diffraction is the R factor, calculated as

$$R = \frac{\sum_i ||F|_i^{\rm o} - |F|_i^{\rm c}|}{\sum_i |F|_i^{\rm o}}, \qquad (4.5.2.74)$$

where $|F|_i^{\rm o}$ and $|F|_i^{\rm c}$ denote the observed (measured) and calculated, respectively, amplitude of either the samples (along R) of the cylindrically averaged intensity $I_l^{1/2}(R)$ (for a noncrystalline specimen) or the cylindrically averaged structure factors $I_l^{1/2}(R_{hk})$ (for a polycrystalline specimen). One way of assessing the significance of the R factor obtained in a particular structure determination is by comparing it with the 'largest likely R factor' (Wilson, 1950), *i.e.* the expected value of the R factor for a random distribution of atoms. Wilson (1950) showed that the largest likely R factor is 0.83 for a centric crystal and 0.59 for an acentric crystal. Although it does not provide a quantitative measure of structural reliability, the largest likely R factor does provide a useful yardstick for evaluating the significance of R factors obtained in structure determinations.

The largest likely R factor for fibre diffraction can be calculated from the amplitude statistics, which depend on the number of degrees of freedom, m, in the measured intensity (Stubbs, 1989; Millane, 1990*a*). Making use of these statistics shows that the largest likely R factor, R_m, for m components is given by (Stubbs, 1989; Millane, 1989*a*)

$$R_m = 2 - 2^{2-m} m \binom{2m-1}{m} B_{1/2}\left(\frac{m+1}{2}, \frac{m}{2}\right), \qquad (4.5.2.75)$$

where $\binom{m}{n}$ is the binomial coefficient and $B_x(m, n)$ the incomplete beta function. The beta function in equation (4.5.2.75) can be replaced by a finite series that is easy to evaluate (Millane, 1989*a*). The expression in equation (4.5.2.75) for R_m can be written in various approximate forms (Millane, 1990*d*, 1992*a*), the simplest being

$$R_m \simeq (2/\pi m)^{1/2} \qquad (4.5.2.76)$$

(Millane, 1990*d*), which shows that the largest likely R factor falls off approximately as $m^{-1/2}$ with increasing m. This is because it is easier to match the sum of a number of structure amplitudes than to match each of them individually. The important conclusion is that the largest likely R factor is smaller in fibre diffraction than in conventional crystallography (where $m = 1$ or 2), and it is smaller when there are more overlapping reflections. This means that for equivalent precision, the R factor must be smaller for a structure determined by fibre diffraction than for one determined by conventional crystallography. How much smaller depends on the number of overlapping reflections on the diffraction pattern.

In a structure determination, the data have different values of m at different positions on the diffraction pattern. Using the definition of the R factor, equation (4.5.2.74), shows that the largest likely R factor for a structure determination is given by (Millane, 1989*b*)

$$R = \frac{\sum_m N_m R_m S_m}{\sum_m N_m S_m}, \qquad (4.5.2.77)$$

where the sums are over the values of m on the diffraction pattern, N_m is the number of data that have m components, R_m is given by equation (4.5.2.75) and S_m is given by

$$S_m = \frac{\Gamma\big((m/2) + (1/2)\big)}{\Gamma(m/2)}, \qquad (4.5.2.78)$$

where $\Gamma(\cdot)$ is the gamma function. The quantities on the right-hand side of equation (4.5.2.77) are easily determined for a particular data set. The largest likely R factor decreases (since m increases) with increasing resolution of the data, increasing diameter of the molecule and decreasing order u of the helix symmetry. For example, for TMV at 5 Å resolution the largest likely R factor is 0.37, and at 3 Å resolution it is 0.31, whereas for a tenfold nucleic acid structure at 3 Å resolution it is 0.40 (Millane, 1989*b*, 1992*b*). This underlines the importance of comparing R factors obtained in a fibre diffraction analysis with the largest likely R factor; an R factor of 0.25 that may indicate a good protein structure may, or may not, indicate a well determined fibre structure.

Using approximations for R_m, S_m and m allows the following approximation for the largest likely R factor for a noncrystalline fibre to be derived (Millane, 1992*b*):

$$R \simeq 0.261(u d_{\rm max}/r_{\rm max})^{1/2}, \qquad (4.5.2.79)$$

where $d_{\rm max}$ is the resolution of the data. The approximation (4.5.2.79) is generally not good enough for calculating accurate largest likely R factors, but it does show the general behaviour with helix symmetry, molecular diameter and diffraction-data resolution. Other approximations to largest likely R factors have been derived that are quite accurate and also include the effect of a minimum resolution for the data (Millane, 1992*b*).

Largest likely R factors in fibre diffraction studies are typically between about 0.3 and 0.5, depending on the particular structure (Millane, 1989*b*, 1992*b*; Millane & Stubbs, 1992). Although the largest likely R factor does not give a quantitative assessment of the significance of an R factor obtained in a particular structure determination, it can be used as a guide to the significance. R factors obtained for well determined protein structures are