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 $(2F_o - F_c)$ 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 $(5F_o - 4F_c)$ 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

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

where $|F|_i^o$ and $|F|_i^c$ 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*):

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

where $d_{\max}$ 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 typically between about one-third and one-half of the corresponding largest likely R factor, depending on the resolution. It is therefore reasonable to expect the R factor for a well determined fibre structure to be between one-third and one-half of the largest likely R factor calculated for the structure. R factors should, therefore, generally be less than 0.15 to 0.25, depending on the particular structure and the resolution as illustrated by the examples presented in Millane & Stubbs (1992).

The free R factor (Brünger, 1997) has become popular in single-crystal crystallography as a tool for validation of refinements. The free R factor is more difficult to implement (but is probably even more important) in fibre diffraction studies because of the smaller data sets, but has been used to advantage in recent studies (Hudson *et al.*, 1997; Welsh *et al.*, 1998, 2000).

4.5.3. Electron crystallography of polymers

(D. L. DORSET)

4.5.3.1. *Is polymer electron crystallography possible?*

As a crystallographic tool, the electron microscope has also made an important impact in polymer science. Historically, single-crystal electron diffraction information has been very useful for the interpretation of cylindrically averaged fibre X-ray patterns (Atkins, 1989), particularly when there is an extensive overlap of diffracted intensities. An electron diffraction pattern aids indexing of the fibre pattern and facilitates measurement of unit-cell constants, and the observation of undistorted plane-group symmetry similarly places important constraints on the identification of the space group (Geil, 1963; Wunderlich, 1973).

The concept of using electron diffraction intensities by themselves for the quantitative determination of crystal structures of polymers or other organics often has been met with scepticism (Lipson & Cochran, 1966). Difficulties experienced in the quantitative interpretation of images and diffraction intensities from 'hard' materials composed of heavy atoms (Hirsch *et al.*, 1965; Cowley, 1981), for example, has adversely affected the outlook for polymer structure analysis, irrespective of whether these reservations are important or not for 'soft' materials comprising light atoms. Despite the still commonly held opinion that no new crystal structures will be determined that are solely based on data collected in the electron microscope, it can be shown that this extremely pessimistic outlook is unwarranted. With proper control of crystallization (*i.e.* crystal thickness) and data collection, the electron microscope can be used quite productively for the direct determination of macromolecular structures at atomic resolution, not only to verify some of the previous findings of fibre X-ray diffraction analysis, but, more importantly, to determine new structures, even of crystalline forms that cannot be studied conveniently by X-rays as drawn fibres (Dorset, 1995*b*). The potential advantages of electron crystallography are therefore clear. The great advantage in scattering cross section of matter for electrons over X-rays permits much smaller samples to be examined by electron diffraction as single-crystalline preparations (Vainshtein, 1964). (Typical dimensions are given below.)

Electron crystallography can be defined as the quantitative use of electron micrographs and electron diffraction intensities for the determination of crystal structures. In the electron microscope, an electron beam illuminates a semitransparent object and the microscope objective lens produces an enlarged representation of the object as an image. If the specimen is thin enough and/or the electron energy is high enough, the weak-phase-object or 'kinematical' approximation is valid (Cowley, 1981), see Chapter 2.5. That is to say, there is an approximate one-to-one mapping of density points between the object mass distribution and the image, within the resolution limits of the instrument (as set by the objective lens aberrations and electron wavelength). The spatial relationships between diffraction and image planes of an electron microscope objective lens are reciprocal and related by Fourier transform operations (Cowley, 1988). While it is easy to transform from the image to the diffraction pattern, the reverse Fourier transform of the diffraction pattern to a high-resolution image requires solution of the famous crystallographic phase problem (as discussed for electron diffraction in Section 2.5.7).

Certainly, in electron diffraction studies, one must still be cognizant of the limitations imposed by the underlying scattering theory. An approximate 'quasi-kinematical' data set is often sufficient for the analysis (Dorset, 1995*a*). However (Dorset, 1995*b*), there are other important perturbations to diffraction intensities which should be minimized. For example, the effects of radiation damage while recording a high-resolution image are minimized by so-called 'low-dose' procedures (Tsuji, 1989).

4.5.3.2. *Crystallization and data collection*

The success of electron crystallographic determinations relies on the possibility of collecting data from *thin* single microcrystals. These can be grown by several methods, including self-seeding, epitaxic orientation, *in situ* polymerization on a substrate, in a Langmuir–Blodgett layer, *in situ* polymerization within a thin layer

references