intensity $I(R,Z)$ into polar coordinates as $I(\rho,\sigma)$, or by simply sampling $I(R,Z)$ for fixed ρ and equally spaced samples of σ, $I_l(R)$ can be calculated from $I(\rho,\sigma)$ by deconvolution, usually by some appropriate solution of the resulting system of linear equations (Makowski, 1978). If the effects of coherence length are significant, as they often are, then equation (4.5.2.55) does not represent a convolution since the width of the Gaussian smearing function depends on σ through equation (4.5.2.20). However, the problem can still be posed as the solution of a system of linear equations and becomes one of profile fitting rather than deconvolution (Millane & Arnott, 1986). This allows the layer-line intensities to be extracted from the data beyond the resolution where they overlap, although there is a limiting resolution, owing to excessive overlap, beyond which reliable data cannot be obtained (Makowski, 1978; Millane & Arnott, 1986). This procedure requires that α_0 and l_c be known; these parameters can be estimated from the angular profiles at low resolution where there is no overlap, or they can be determined as part of the profile-fitting procedure.

For a diffraction pattern from a polycrystalline specimen containing Bragg reflections, the intensities $I_l(R_{hk})$ given by equation (4.5.2.24) need to be extracted from the intensity $I(R,Z)$ on the diffraction pattern mapped into reciprocal space. Each composite reflection $I_l(R_{hk})$ is smeared into a spot whose intensity profile is given by equation (4.5.2.27), and adjacent reflections may overlap. The intensity $I_l(R_{hk})$ is equal to the intensity $I(R,Z)$ integrated over the region of the spot, and the intensity at the centre of a spot is reduced, relative to $I_l(R_{hk})$, by a factor that increases with the degree of smearing.

The c repeat can be obtained immediately from the layer-line spacing. Initial estimates of the remaining cell constants can be made from inspection of the (R,Z) coordinates of low-order reflections. These values are refined by minimizing the difference between the calculated and measured (R,Z) coordinates of all the sharp reflections on the pattern.

One approach to measuring the intensities of Bragg reflections is to estimate the boundary of each spot (or a fixed proportion of the region occupied by each spot) and integrate the intensity over that region (Millane & Arnott, 1986; Hall *et al.*, 1987). For spots that overlap, an integration region that is the union of the region occupied by each contributing spot can be used, allowing the intensities for composite spots to be calculated (Millane & Arnott, 1986). This is more accurate than methods based on the measurement of the peak intensity followed by a correction for smearing. Integration methods suffer from problems associated with determining accurate spot boundaries and they are not capable of separating weakly overlapping spots. A more effective approach is one based on profile fitting. The intensity distribution on the diffraction pattern can be written as

$$I(R,Z) = \sum_l \sum_{h,k} I_l(R_{hk},R,Z), \qquad (4.5.2.56)$$

where $I_l(R_{hk},R,Z)$ denotes the intensity distribution of the spot $I_l(R_{hk})$, and the sums are over all spots on the diffraction pattern. Using equation (4.5.2.27) shows that equation (4.5.2.56) can be written as

$$I(R,Z) = \sum_l \sum_{h,k} I_l(R_{hk}) S(R_{hk};\, l/c;\, R;\, Z), \qquad (4.5.2.57)$$

where $S(R_{hk};\, l/c;\, R;\, Z)$ denotes the profile of the spot centred at $(R_{hk}, l/c)$ [which can be derived from equation (4.5.2.27)]. Given estimates of the parameters $l_{\rm lat}$, $l_{\rm axial}$ and α_0, equation (4.5.2.57) can be written as a system of linear equations that can be solved for the intensities $I_l(R_{hk})$ from the data $I(R,Z)$ on the diffraction pattern. The parameters $l_{\rm lat}$, $l_{\rm axial}$ and α_0, as well as the cell constants and possibly other parameters, can also be refined as part of the profile-fitting procedure using nonlinear optimization.

A suite of programs for processing fibre diffraction data is distributed (and often developed) by the Collaborative Computational Project for Fibre and Polymer Diffraction (CCP13) in the UK (www.dl.ac.uk/SRS/CCP13) (Shotton *et al.*, 1998).

4.5.2.6. *Structure determination*

4.5.2.6.1. *Overview*

Structure determination in fibre diffraction is concerned with determining atomic coordinates or some other structural parameters, from the measured cylindrically averaged diffraction data. Fibre diffraction analysis suffers from the phase problem and low resolution (diffraction data rarely extend beyond 3 Å resolution), but this is no worse than in protein crystallography where phases derived from, say, isomorphous replacement or molecular replacement, coupled with the considerable stereochemical information usually available on the molecule under study, together contribute enough information to lead to precise structures. What makes structure determination by fibre diffraction more difficult is the loss of information owing to the cylindrical averaging of the diffraction data. However, in spite of these difficulties, fibre diffraction has been used to determine, with high precision, the structures of a wide variety of biological and synthetic polymers, and other macromolecular assemblies. Because of the size of the repeating unit and the resolution of the diffraction data, methods for structure determination in fibre diffraction tend to mimic those of macromolecular (protein) crystallography, rather than small-molecule crystallography (direct methods).

For a noncrystalline fibre one can determine only the molecular structure from the continuous diffraction data, whereas for a polycrystalline fibre one can determine crystal structures from the Bragg diffraction data. However, there is little fundamental difference between methods used for structure determination with noncrystalline and polycrystalline fibres. For partially crystalline fibres, little has so far been attempted with regard to rigorous structure determination.

As is the case with protein crystallography, the precise methods used for structure determination by fibre diffraction depend on the particular problem at hand. A variety of tools are available and one selects from these those that are appropriate given the data available in a particular case. For example, the structure of a polycrystalline polynucleotide might be determined by using Patterson functions to determine possible packing arrangements, molecular model building to define, refine and arbitrate between structures, difference Fourier synthesis to locate ions or solvent molecules, and finally assessment of the reliability of the structure. As a second example, to determine the structure of a helical virus, one might use isomorphous replacement to obtain phase estimates, calculate an electron-density map, fit a preliminary model and refine it using simulated annealing alternating with difference Fourier analysis, and assess the results. The various tools available, together with indications of where and how they are used, are described in the following sections.

Although a variety of techniques are used to solve structures using fibre diffraction, most of the methods do fall broadly into one of three classes that depend primarily on the size of the helical repeat unit. The first class applies to molecules whose repeating units are small, *i.e.* are represented by a relatively small number of independent parameters or degrees of freedom (after all stereochemical constraints have been incorporated). The structure can then be determined by an exhaustive exploration of the parameter space using molecular model building. The first example above would belong to this class. The second class of methods is appropriate when the size of the helical repeating unit is such that

its structure is described by too many variable parameters for the parameter space to be explored *a priori*. It is then necessary to phase the fibre diffraction data and construct an electron-density map into which the molecular structure can be fitted and then refined. The second example above would belong to this class. The second class of methods therefore mimics conventional protein crystallography quite closely. The third class of problems applies when the structure is large, but there are too few diffraction data to attempt phasing and the usual determination of atomic coordinates. The solution to such problems varies from case to case and usually involves modelling and optimization of some kind.

An important parameter in structure determination by fibre diffraction is the degree of overlap (that results from the cylindrical averaging) in the data. This parameter is equal to the number of significant terms in equation (4.5.2.17) or the number of independent terms in equation (4.5.2.24), and depends on the position in reciprocal space and, for a polycrystalline fibre, the space-group symmetry. The number of degrees of freedom in a particular datum is equal to twice this number (since each structure factor generally has real and imaginary parts), and is denoted in this section by m. Determination of the $G_{nl}(R)$ from the cylindrically averaged data $I_l(R)$ therefore involves separating the $m/2$ amplitudes $|G_{nl}(R)|$ and assigning phases to each. The electron density can be calculated from the $G_{nl}(R)$ using equations (4.5.2.7) and (4.5.2.11).

4.5.2.6.2. *Helix symmetry, cell constants and space-group symmetry*

The first step in analysis of any fibre diffraction pattern is determination of the molecular helix symmetry u_v. Only the zero-order Bessel term contributes diffracted intensity on the meridian, and referring to equation (4.5.2.6) shows that the zero-order term occurs only on layer lines for which l is a multiple of u. Therefore, inspection of the distribution of diffraction along the meridian allows the value of u to be inferred. This procedure is usually effective, but can be difficult if u is large, because the first meridional maximum may be on a layer line that is difficult to measure. This difficulty was overcome in one case by Franklin & Holmes (1958) by noting that the second Bessel term on the equator is $n = u$, estimating $G_{00}(R)$ using data from a heavy-atom derivative (see Section 4.5.2.6.6), subtracting this from $I_0(R)$, and using the behaviour of the remaining intensity for small R to infer the order of the next Bessel term [using equation (4.5.2.14)] and thence u.

Referring to equations (4.5.2.6) and (4.5.2.14) shows that the distribution of $R_{\min}$ for $0 < l < u$ depends on the value of v. Therefore, inspection of the intensity distribution close to the meridian often allows v to be inferred. Note, however, that the distribution of $R_{\min}$ does not distinguish between the helix symmetries u_v and u_{u-v}. Any remaining ambiguities in the helix symmetry need to be resolved by steric considerations, or by detailed testing of models with the different symmetries against the available data.

For a polycrystalline system, the cell constants are determined from the (R, Z) coordinates of the spots on the diffraction pattern as described in Section 4.5.2.6.4. Space-group assignment is based on analysis of systematic absences, as in conventional crystallography. However, in some cases, because of possible overlap of systematic absences with other reflections, there may be some ambiguity in space-group assignment. However, the space group can always be limited to one of a few possibilities, and ambiguities can usually be resolved during structure determination (Section 4.5.2.6.4).

4.5.2.6.3. *Patterson functions*

In fibre diffraction, the conventional Patterson function cannot be calculated since the individual structure-factor intensities are not available. However, MacGillavry & Bruins (1948) showed that the *cylindrically averaged Patterson function* can be calculated from fibre diffraction data. Consider the function $\hat{Q}(r, z)$ defined by

$$\hat{Q}(r,z) = \sum_{l=0}^{\infty} \int_0^{\infty} \varepsilon_l I_l(R) J_0(2\pi R r) \cos(2\pi l z/c) 2\pi R \, \mathrm{d}R, \quad (4.5.2.58)$$

where $\varepsilon_l = 1$ for $l = 0$ and 2 for $l > 0$, which can be calculated from the intensity distribution on a continuous fibre diffraction pattern. Using equations (4.5.2.7), (4.5.2.10), (4.5.2.17) and (4.5.2.58) shows that $\hat{Q}(r, z)$ is the cylindrical average of the Patterson function, $\hat{P}(r, \varphi, z)$, of one molecule, *i.e.*

$$\hat{Q}(r,z) = (1/2\pi) \int_0^{2\pi} \hat{P}(r,\varphi,z) \, \mathrm{d}\varphi. \quad (4.5.2.59)$$

The $\hat{}$ symbols on $\hat{P}(r, \varphi, z)$ and $\hat{Q}(r, z)$ indicate that these are Patterson functions of a single molecule, as distinct from the usual Patterson function of a crystal, which contains intermolecular interatomic vectors and is periodic with the same periodicity as the crystal. $\hat{P}(r, \varphi, z)$ is periodic only along z and is therefore, strictly, a Patterson function along z and an autocorrelation function along x and y (Millane, 1990*b*). The cylindrically averaged Patterson contains information on interatomic separations along the axial direction and in the lateral plane, but no information on orientations of the vectors in the lateral plane.

For a polycrystalline system; consider the function $Q(r, z)$ given by

$$Q(r,z) = \sum_l \sum_{h,k} R_{hk} I_l(R_{hk}) J_0(2\pi R_{hk} r) \cos(2\pi l z/c), \quad (4.5.2.60)$$

where the sums are over all the overlapped reflections $I_l(R_{hk})$ on the diffraction pattern, given by equation (4.5.2.24). It is easily shown that $Q(r, z)$ is related to the Patterson function $P(r, \varphi, z)$ by

$$Q(r,z) = (1/2\pi) \int_0^{2\pi} P(r,\varphi,z) \, \mathrm{d}\varphi, \quad (4.5.2.61)$$

where, in this case, $P(r, \varphi, z)$ is the usual Patterson function (expressed in cylindrical polar coordinates), *i.e.* it contains all intermolecular (both intra- and inter-unit cell) interatomic vectors and has the same translational symmetry as the unit cell. The cylindrically averaged Patterson function for polycrystalline fibres therefore contains the same information as it does for noncrystalline fibres (*i.e.* no angular information in the lateral plane), except that it also contains information on intermolecular separations.

Low resolution and cylindrical averaging, in addition to the usual difficulties with interpretation of Patterson functions, has resulted in the cylindrically averaged Patterson function not playing a major role in structure determination by fibre diffraction. However, information provided by the cylindrically averaged Patterson function has, in a number of instances, been a useful component in fibre diffraction analyses. A good review of the application of Patterson functions in fibre diffraction is given by Stubbs (1987). Removing data from the low-resolution part (or all) of the equator when calculating the cylindrically averaged Patterson function removes the strong vectors related to axially invariant (or cylindrically symmetric) parts of the map, and can aid interpretation (Namba *et al.*, 1980; Stubbs, 1987). It is also important when calculating cylindrically averaged Patterson functions to use data only at a resolution that is appropriate to the size and spacings of features one is looking for (Stubbs, 1987).

Cylindrically averaged Patterson functions were used in early applications of fibre diffraction analysis (Franklin & Gosling, 1953; Franklin & Klug, 1955). The intermolecular peaks that usually dominate in a cylindrically averaged Patterson function can help to define the locations of multiple molecules in the unit cell.

Depending on the space-group symmetry, it is sometimes possible to calculate the complete three-dimensional Patterson function (or certain projections of it). This comes about because of the equivalence of the amplitudes of overlapping reflections in some high-symmetry space groups. The intensity of each reflection can then be determined and a full three-dimensional Patterson map calculated (Alexeev *et al.*, 1992). The only difficulty is that non-systematic overlaps are often present, although these are usually relatively few in number and the intensity can be apportioned equally amongst them, the resulting errors usually being small relative to the level of detail present in the Patterson map. For lower space-group symmetries, it may not be possible to calculate a three-dimensional Patterson map, but it may be possible to calculate certain projections of the map. For example, if the overlapped *hk*0 reflections have the same intensities, a projection of the Patterson map down the *c* axis can be calculated. Since such a projection is along the polymer axes, it gives the relative positions of the molecules in the *ab* plane. If the combined helix and space-group symmetry is high, an estimate of the electron density can be obtained by averaging appropriate copies of the three-dimensional Patterson function (Alexeev *et al.*, 1992).

4.5.2.6.4. *Molecular model building*

The majority of the structures determined by X-ray fibre diffraction analysis have been determined by molecular model building (Campbell Smith & Arnott, 1978; Arnott, 1980; Millane, 1988). Most applications of molecular model building have been to polycrystalline systems, although there have been a number of applications to noncrystalline systems (Park *et al.*, 1987; Millane *et al.*, 1988). The approach is to use spacings and symmetry information derived directly from the diffraction pattern, coupled with the primary structure and stereochemical information on the molecule under study, to construct models of all *kinds* of possible molecular or crystal structure. These models are each refined (optimized) against the diffraction data, as well as stereochemical restraints, to produce the best model of each kind. The optimized models can be compared using various figures of merit, and in favourable cases one model will be sufficiently superior to the remainder for it to represent unequivocally the correct structure. The principle of this approach is that by making use of stereochemical constraints, the molecular and crystal structure have few enough degrees of freedom that the parameter space has a sufficiently small number of local minima for these to be identified and individually examined to find the global minimum. The X-ray phases are therefore not determined explicitly.

There are three steps involved in structure determination by molecular model building: (1) construction of all possible molecular and crystal structure models, (2) refinement of each model against the X-ray data and stereochemical restraints, and (3) adjudication among the refined models. The overall procedure for determining polymer structures using molecular model building is summarized by the flow chart in Fig. 4.5.2.2, and is described below.

The helix symmetry of the molecule, or one of a few helix symmetries, can be determined as described in Section 4.5.2.6.2. Different kinds of molecular model may correspond to one of a few different helix symmetries, usually corresponding to different values of v. For example, helix symmetries u_v and u_{u-v}, which correspond to the left- and right-handed helices, cannot be distinguished on the basis of the overall intensity distribution alone. Other examples of different kinds of molecular model may include single, double or multiple helices, parallel or antiparallel double helices, different juxtapositions of chains within multiple helices and different conformational domains within the molecule. For polycrystalline systems, in addition to different kinds of molecular structures, there are often different kinds of possible packing arrangements within the unit cell. There may be a number of possible packings which correspond to different arrangements within the crystallographic asymmetric unit, and there may be more than one space group that needs to be considered.

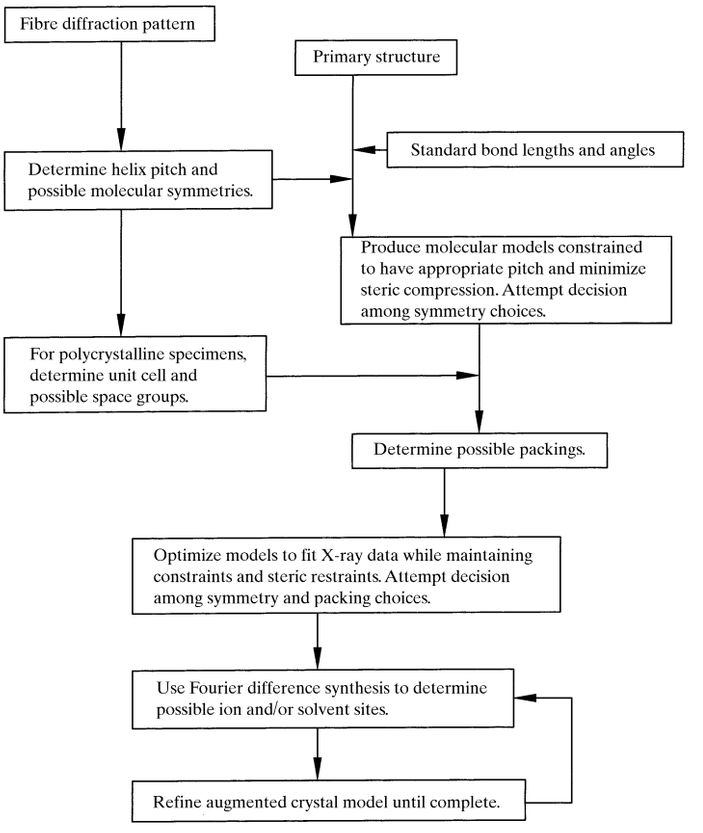

Fig. 4.5.2.2. Flow chart of the molecular-model-building approach to structure determination (Arnott, 1980).

Despite the apparent large number of potential starting models implied by the above discussion, in practice the number of feasible models is usually quite small, and many of these are often eliminated at an early stage. Definition and refinement of helical polymers [steps (1) and (2) above] are carried out using computer programs, the most popular and versatile being the *linked-atom least-squares* (*LALS*) system (Campbell Smith & Arnott, 1978; Millane *et al.*, 1985), originally developed by Arnott and co-workers in the early 1960s (Arnott & Wonacott, 1966). This system has been used to determine the structures of a wide variety of polynucleotides, polysaccharides, polyesters and polypeptides (Arnott, 1980; Arnott & Mitra, 1984; Chandrasekaran & Arnott, 1989; Millane, 1990*c*). Other refinement systems exist (Zugenmaier & Sarko, 1980; Iannelli, 1994), but the principles are essentially the same and the following discussion is in terms of the *LALS* system. The atomic coordinates are defined, using a linked-atom description, in terms of bond lengths, bond angles and conformation (torsion) angles (Campbell Smith & Arnott, 1978). Stereochemical constraints are imposed, and the number of parameters reduced, by fixing the bond lengths, often (but not always) the bond angles, and possibly some of the conformation angles. The molecular conformation is then defined by the remaining parameters. For polycrystalline systems, there are usually additional variable parameters that define the packing of the molecule(s) in the unit cell. A further source of stereochemical data is the requirement that a model exhibit no over-short nonbonded interatomic distances. These are incorporated by a quadratic nonbonded potential that is

matched to a Buckingham potential (Campbell Smith & Arnott, 1978). A variety of other restraints can also be incorporated.

In the *LALS* system, the quantity Ω given by

$$\Omega = \sum_m \omega_m \Delta F_m^2 + \sum_m k_m \Delta d_m^2 + \sum_m \lambda_m G_m = X + C + L \quad (4.5.2.62)$$

is minimized by varying a set of chosen parameters consisting of conformation angles, possibly bond angles, and packing parameters. The term X involves the differences ΔF_m between the model and experimental X-ray amplitudes – Bragg and/or continuous. The term C involves restraints to ensure that overshort nonbonded interatomic distances are driven beyond acceptable minimum values, that conformations are within desired domains, that hydrogen-bond and coordination geometries are close to the expected configurations, and a variety of other relationships are satisfied (Campbell Smith & Arnott, 1978). The ω_m and k_m are weights that are inversely proportional to the estimated variances of the data. The term L involves constraints which are relationships that are to be satisfied exactly $(G_m = 0)$ and the λ_m are Lagrange multipliers. Constraints are used, for example, to ensure connectivity from one helix pitch to the next and to ensure that chemical ring systems are closed. The cost function Ω is minimized using full-matrix nonlinear least squares and singular value decomposition (Campbell Smith & Arnott, 1978).

Structure determination usually involves first using equation (4.5.2.62) with the terms C and L only, to establish the stereochemical viability of each kind of possible molecular model and packing arrangement. It is worth emphasizing that it is usually advantageous if the specimen is polycrystalline, even though the continuous diffraction contains, in principle, more information than the Bragg reflections (since the latter are sampled). This is because the molecule in a noncrystalline specimen must be refined in steric isolation, whereas for a polycrystalline specimen it is refined while packed in the crystal lattice. The extra information provided by the intermolecular contacts can often help to eliminate incorrect models. This can be particularly significant if the molecule has flexible sidechains. The initial models that survive the steric optimization are then optimized also against the X-ray data, by further refinement with X included in equation (4.5.2.62). The ratios $(\Omega_P/\Omega_Q)^{1/2}$ and $(X_P/X_Q)^{1/2}$ can be used in Hamilton's test (Hamilton, 1965) to evaluate the differences between models P and Q. On the basis of these statistical tests, one can decide if one model is superior to the others at an acceptable confidence level. In the final stages of refinement, bond angles may be varied in a 'stiffly elastic' fashion from their mean values if there are sufficient data to justify the increase in the number of degrees of freedom.

If sufficient X-ray data are available, it is sometimes possible to locate additional ordered molecules such as counterions or solvent molecules by difference Fourier synthesis as described in Section 4.5.2.6.5. Their positions can then be co-refined with the polymer structure while hydrogen bonds and coordination geometries are optimized. The resulting structure can then be used to compute improved phases to search for additional molecules. Since the signal-to-noise ratio in fibre difference syntheses is usually low, difference maps must be interpreted with caution. The assignment of counterions or solvent molecules to peaks in the difference synthesis must be supported by plausible 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 $I_I(R_{hk}) = I_o$ 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.*

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

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*)

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

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

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

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

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

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

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

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

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

where $G_{00}^D(R)$ 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 $I_l^D(R) - I_l(R)$ 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 overdetermined 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

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

where E_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), $I_l^o(R_i)$ and $I_l^c(R_i)$ are the observed and calculated, respectively, cylindrically averaged diffraction intensities sampled at $R = R_i$, the w_{li} are weights for the observed intensities $I_l^o(R_i)$ 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 $(2F_o - F_c)$ 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