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.

references