

## 4. DIFFUSE SCATTERING AND RELATED TOPICS

## 4.5.2. X-ray fibre diffraction analysis

BY R. P. MILLANE

## 4.5.2.1. Introduction

*X-ray fibre diffraction analysis* is a collection of crystallographic techniques that are used to determine molecular and crystal structures of molecules, or molecular assemblies, that form specimens (often fibres) in which the molecules, assemblies or crystallites are approximately parallel but not otherwise ordered (Arnott, 1980; French & Gardner, 1980; Hall, 1984; Vibert, 1987; Millane, 1988; Atkins, 1989; Stubbs, 1999). These are usually long, slender molecules and they are often inherently flexible, which usually precludes the formation of regular three-dimensional crystals suitable for conventional crystallographic analysis. X-ray fibre diffraction therefore provides a route for structure determination for certain kinds of specimens that cannot be crystallized. Although it may be possible to crystallize small *fragments* or *subunits* of these molecules, and determine the crystal structures of these, X-ray fibre diffraction provides a means for studying the intact, and often the biologically or functionally active, system. Fibre diffraction has played an important role in the determination of biopolymers such as polynucleotides, polysaccharides (both linear and branched), polypeptides and a wide variety of synthetic polymers (such as polyesters), as well as larger assemblies including rod-like helical viruses, bacteriophages, microtubules and muscle fibres (Arnott, 1980; Arnott & Mitra, 1984; Millane, 1990c; Squire & Vibert, 1987).

Specimens appropriate for fibre diffraction analysis exhibit rotational disorder (of the molecules, aggregates or crystallites) about a preferred axis, resulting in cylindrical averaging of the diffracted intensity in reciprocal space. Therefore, fibre diffraction analysis can be thought of as 'structure determination from cylindrically averaged diffraction intensities' (Millane, 1993). In a powder specimen the crystallites are completely (spherically) disordered, so that structure determination by fibre diffraction can be considered to be intermediate between structure determination from single crystals and from powders.

This section is a review of the theory and techniques of structure determination by X-ray fibre diffraction analysis. It includes descriptions of fibre specimens, the theory of diffraction by these specimens, intensity data collection and processing, and the variety of structure determination methods used for the various kinds of specimens studied by fibre diffraction. It does not include descriptions of specimen preparation (those can be found in the references given for specific systems), or of applications of X-ray diffraction to determining polymer morphology (*e.g.* particle or void sizes and shapes, texture, domain structure *etc.*).

## 4.5.2.2. Fibre specimens

A wide variety of kinds of fibre specimen exist. All exhibit preferred orientation; the variety results from variability in the degree of order (crystallinity) in the lateral plane (the plane perpendicular to the axis of preferred orientation). This leads to categorization of three kinds of fibre specimen: *noncrystalline fibres*, in which there is no order in the lateral plane; *polycrystalline fibres*, in which there is near-perfect crystallinity in the lateral plane; and *disordered fibres*, in which there is disorder either within the molecules or in their crystalline packing (or both). The kind of fibre specimen affects the kind of diffraction pattern obtained, the relationships between the molecular and crystal structures and the diffraction data, methods of data collection, and methods of structure determination.

Noncrystalline fibres are made up of a collection of molecules that are *oriented*. This means that there is a common axis in each molecule (referred to here as the *molecular axis*), the axes being parallel in the specimen. The direction of preferred orientation is

called the *fibre axis*. The molecule itself is usually considered to be a rigid body. There is no other ordering within the specimen. The molecules are therefore randomly positioned in the lateral plane and are randomly rotated about their molecular axes. Furthermore, if the molecule does not have a twofold rotation axis normal to the molecular axis, then the molecular axis has a *direction* associated with it, and the molecular axes are oriented randomly parallel or antiparallel to each other. This is often called *directional disorder*, or the molecules are said to be oriented *randomly up and down*. The average length of the ordered molecular segments in a noncrystalline fibre is referred to as the *coherence length*.

Polycrystalline fibres are characterized by molecular segments packing together to form well ordered microcrystallites within the specimen. The crystallites effectively take the place of the molecules in a noncrystalline specimen as described above. The crystallites are oriented, and since the axis within each crystallite that is aligned parallel to those in other crystallites usually corresponds to the long axes of the constituent molecules, it is also referred to here as the molecular axis. The crystallites are randomly positioned in the lateral plane, randomly rotated about the molecular axis, and randomly oriented up or down. The size of the crystalline domains can be characterized by their average dimensions in the directions of the **a**, **b** and **c** unit-cell vectors. However, because of the rotational disorder of the crystallites, any differences between crystallite dimensions in different directions normal to the fibre axis tend to be smeared out in the diffraction pattern, and the crystallite size is usefully characterized by the average dimensions of the crystallites normal and parallel to the fibre axis.

The molecules or crystallites in a fibre specimen are not perfectly oriented, and the variation in inclinations of the molecular axes to the fibre axis is referred to as *disorientation*. Assuming that the orientation is axisymmetric, then it can be described by an *orientation density function*  $\Omega(\alpha)$  such that  $\Omega(\alpha) d\omega$  is the fraction of molecules in an element of solid angle  $d\omega$  inclined at an angle  $\alpha$  to the fibre axis. The exact form of  $\Omega(\alpha)$  is generally not known for any particular fibre and it is often sufficient to assume a Gaussian orientation density function, so that

$$\Omega(\alpha) = \frac{1}{2\pi\alpha_0^2} \exp\left(-\frac{\alpha^2}{2\alpha_0^2}\right), \quad (4.5.2.1)$$

where  $\alpha_0$  is a measure of the degree of disorientation.

Fibre specimens often exhibit various kinds of disorder. The disorder may be within the molecules or in their packing. Disorder affects the relationship between the molecular and crystal structure and the diffracted intensities. Disorder within the molecules may result from a degree of randomness in the chemical sequence of the molecule or from variability in the interactions between the units that make up the molecule. Such molecules may (at least in principle) form noncrystalline, polycrystalline or partially crystalline (described below) fibres. Disordered packing of molecules within crystallites can result from a variety of ways in which the molecules can interact with each other. Fibre specimens made up of disordered crystallites are referred to here as partially crystalline fibres.

## 4.5.2.3. Diffraction by helical structures

Molecules or assemblies studied by fibre diffraction are usually made up of a large number of identical, or nearly identical, residues, or subunits, that in an oriented specimen are distributed along an axis; this leads naturally to helical symmetry. Since a periodic structure with no helix symmetry can be treated as a onefold helix, the assumption of helix symmetry is not restrictive.

## 4.5. POLYMER CRYSTALLOGRAPHY

### 4.5.2.3.1. Helix symmetry

The presence of a unique axis about which there is rotational disorder means that it is convenient to use cylindrical polar coordinate systems in fibre diffraction. We denote by  $(r, \varphi, z)$  a cylindrical polar coordinate system in real space, in which the  $z$  axis is parallel to the molecular axes. The molecule is said to have  $u_v$  helix symmetry, where  $u$  and  $v$  are integers, if the electron density  $f(r, \varphi, z)$  satisfies

$$f(r, \varphi + (2\pi mv/u), z + (mc/u)) = f(r, \varphi, z), \quad (4.5.2.2)$$

where  $m$  is any integer. The constant  $c$  is the period along the  $z$  direction, which is referred to variously as the *molecular repeat distance*, the *crystallographic repeat*, or the *c repeat*. The helix pitch  $P$  is equal to  $c/v$ . Helix symmetry is easily interpreted as follows. There are  $u$  subunits, or *helix repeat units*, in one  $c$  repeat of the molecule. The helix repeat units are repeated by integral rotations of  $2\pi v/u$  about, and translations of  $c/u$  along, the molecular (or helix) axis. The helix repeat units may therefore be referenced to a helical lattice that consists of points at a fixed radius, with relative rotations and translations as described above. These points lie on a helix of pitch  $P$ , there are  $v$  turns (or pitch-lengths) of the helix in one  $c$  repeat, and there are  $u$  helical lattice points in one  $c$  repeat. A  $u_v$  helix is said to have ' $u$  residues in  $v$  turns'.

Since the electron density is periodic in  $\varphi$  and  $z$ , it can be decomposed into a Fourier series as

$$f(r, \varphi, z) = \sum_{l=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} g_{nl}(r) \exp(i[n\varphi - (2\pi lz/c)]), \quad (4.5.2.3)$$

where the coefficients  $g_{nl}(r)$  are given by

$$g_{nl}(r) = (c/2\pi) \int_0^{c/2\pi} \int_0^c f(r, \varphi, z) \exp(i[-n\varphi + (2\pi lz/c)]) d\varphi dz. \quad (4.5.2.4)$$

Assume now that the electron density has helical symmetry. Denote by  $g(r, \varphi, z)$  the electron density in the region  $0 < z < c/u$ ; the electron density being zero outside this region, *i.e.*  $g(r, \varphi, z)$  is the electron density of a single helix repeat unit. It follows that

$$f(r, \varphi, z) = \sum_{m=-\infty}^{\infty} g[r, \varphi + (2\pi mv/u), z + (mc/u)]. \quad (4.5.2.5)$$

Substituting equation (4.5.2.5) into equation (4.5.2.4) shows that  $g_{nl}(r)$  vanishes unless  $(l - nv)$  is a multiple of  $u$ , *i.e.* unless

$$l = um + vn \quad (4.5.2.6)$$

for any integer  $m$ . Equation (4.5.2.6) is called the *helix selection rule*. The electron density in the helix repeat unit is therefore given by

$$g(r, \varphi, z) = \sum_l \sum_n g_{nl}(r) \exp(i[n\varphi - (2\pi lz/c)]), \quad (4.5.2.7)$$

where

$$g_{nl}(r) = (c/2\pi) \int \int g(r, \varphi, z) \exp(i[-n\varphi + (2\pi lz/c)]) d\varphi dz, \quad (4.5.2.8)$$

and where in equation (4.5.2.7) (and in the remainder of this section) the sum over  $l$  is over all integers, the sum over  $n$  is over all integers satisfying the helix selection rule and the integral in equation (4.5.2.8) is over one helix repeat unit. The effect of helix symmetry, therefore, is to restrict the number of Fourier coefficients  $g_{nl}(r)$  required to represent the electron density to those whose index  $n$  satisfies the selection rule. Note that the selection rule is usually derived using a rather more complicated argument by considering the convolution of the Fourier transform of a continuous filamentary helix with a set of planes in reciprocal space (Cochran *et al.*, 1952). The approach described above, which follows that of Millane (1991), is much more straightforward.

### 4.5.2.3.2. Diffraction by helical structures

Denote by  $(R, \psi, Z)$  a cylindrical polar coordinate system in reciprocal space (with the  $Z$  and  $z$  axes parallel), and by  $F(R, \psi, Z)$  the Fourier transform of  $f(r, \varphi, z)$ . Since  $f(r, \varphi, z)$  is periodic in  $z$  with period  $c$ , its Fourier transform is nonzero only on the *layer planes*  $Z = l/c$  where  $l$  is an integer. Denote  $F(R, \psi, l/c)$  by  $F_l(R, \psi)$ ; using the cylindrical form of the Fourier transform shows that

$$F_l(R, \psi) = \int_0^c \int_0^{2\pi} \int_0^\infty f(r, \varphi, z) \exp(i2\pi[Rr \cos(\psi - \varphi) + (lz/c)]) r dr d\varphi dz. \quad (4.5.2.9)$$

It is convenient to rewrite equation (4.5.2.9) making use of the Fourier decomposition described in Section 4.5.2.3.1, since this allows utilization of the helix selection rule. The *Fourier-Bessel structure factors* (Klug *et al.*, 1958),  $G_{nl}(R)$ , are defined as the Hankel transform of the Fourier coefficients  $g_{nl}(r)$ , *i.e.*

$$G_{nl}(R) = \int_0^\infty g_{nl}(r) J_n(2\pi Rr) 2\pi r dr, \quad (4.5.2.10)$$

and the inverse transform is

$$g_{nl}(r) = \int_0^\infty G_{nl}(R) J_n(2\pi Rr) 2\pi R dR. \quad (4.5.2.11)$$

Using equations (4.5.2.7) and (4.5.2.11) shows that equation (4.5.2.9) can be written as

$$F_l(R, \psi) = \sum_n G_{nl}(R) \exp(in[\psi + (\pi/2)]), \quad (4.5.2.12)$$

where, as usual, the sum is over only those values of  $n$  that satisfy the helix selection rule. Using equations (4.5.2.8) and (4.5.2.10) shows that the Fourier-Bessel structure factors may be written in terms of the atomic coordinates as

$$G_{nl}(R) = \sum_j f_j(\rho) J_n(2\pi Rr_j) \exp(i[-n\varphi_j + (2\pi lz_j/c)]), \quad (4.5.2.13)$$

where  $f_j(\rho)$  is the (spherically symmetric) atomic scattering factor (usually including an isotropic temperature factor) of the  $j$ th

#### 4. DIFFUSE SCATTERING AND RELATED TOPICS

atom and  $\rho = (R^2 + l^2/Z^2)^{1/2}$  is the spherical radius in reciprocal space. Equations (4.5.2.12) and (4.5.2.13) allow the complex diffracted amplitudes for a helical molecule to be calculated from the atomic coordinates, and are analogous to expressions for the structure factors in conventional crystallography.

The significance of the selection rule is now more apparent. On a particular layer plane  $l$ , not all Fourier–Bessel structure factors  $G_n(R)$  contribute; only those whose Bessel order  $n$  satisfies the selection rule for that value of  $l$  contribute. Since any molecule has a maximum radius, denoted here by  $r_{\max}$ , and since  $J_n(x)$  is small for  $x < |n| - 2$  and diffraction data are measured out to only a finite value of  $R$ , reference to equation (4.5.2.10) [or equation (4.5.2.13)] shows that there is a maximum Bessel order that contributes significant value to equation (4.5.2.12) (Crowther *et al.*, 1970; Makowski, 1982), so that the infinite sum over  $n$  in equation (4.5.2.12) can be replaced by a finite sum. On each layer plane there is also a minimum value of  $|n|$ , denoted by  $n_{\min}$ , that satisfies the helix selection rule, so that the region  $R < R_{\min}$  is devoid of diffracted amplitude where

$$R_{\min} = \frac{n_{\min} - 2}{2\pi r_{\max}}. \quad (4.5.2.14)$$

The selection rule therefore results in a region around the  $Z$  axis of reciprocal space that is devoid of diffraction, the shape of the region depending on the helix symmetry.

##### 4.5.2.3.3. Approximate helix symmetry

In some cases the nature of the subunits and their interactions results in a structure that is not exactly periodic. Consider a helical structure with  $u + x$  subunits in  $v$  turns, where  $x$  is a small ( $x \ll 1$ ) real number; *i.e.* the structure has approximate, but not exact,  $u_v$  helix symmetry. Since the molecule has an *approximate* repeat distance  $c$ , only those layer planes close to those at  $Z = l/c$  show significant diffraction. Denoting by  $Z_{mn}$  the  $Z$  coordinate of the  $n$ th Bessel order and its associated value of  $m$ , and using the selection rule shows that

$$Z_{mn} = [(um + vn)/c] + (mx/c) = (l/c) + (mx/c), \quad (4.5.2.15)$$

so that the positions of the Bessel orders are shifted by  $mx/c$  from their positions if the helix symmetry is exactly  $u_v$ . At moderate resolution  $m$  is small so the shift is small. Hence Bessel orders that would have been coincident on a particular layer plane are now separated in reciprocal space. This is referred to as *layer-plane splitting* and was first observed in fibre diffraction patterns from tobacco mosaic virus (TMV) (Franklin & Klug, 1955). Splitting can be used to advantage in structure determination (Section 4.5.2.6.6).

As an example, TMV has approximately  $49_3$  helix symmetry with a  $c$  repeat of 69 Å. However, close inspection of diffraction patterns from TMV shows that there are actually about 49.02 subunits in three turns (Stubbs & Makowski, 1982). The virus is therefore more accurately described as a  $2451_{150}$  helix with a  $c$  repeat of 3450 Å. The layer lines corresponding to this larger repeat distance are not observed, but the effects of layer-plane splitting are detectable (Stubbs & Makowski, 1982).

##### 4.5.2.4. Diffraction by fibres

The kind of diffraction pattern obtained from a fibre specimen made up of helical molecules depends on the kind of specimen as described in Section 4.5.2.2. This section is divided into four parts. The first two describe diffraction patterns obtained from noncrystalline and polycrystalline fibres (which are the most common kinds used for structural analysis), and the last two describe diffraction by partially crystalline fibres.

##### 4.5.2.4.1. Noncrystalline fibres

A noncrystalline fibre is made up of a collection of helical molecules that are oriented parallel to each other, but are otherwise randomly positioned and rotated relative to each other. The recorded intensity,  $I_l(R)$ , is therefore that diffracted by a single molecule cylindrically averaged about the  $Z$  axis in reciprocal space *i.e.*

$$I_l(R) = (1/2\pi) \int_0^{2\pi} |F_l(R, \psi)|^2 d\psi; \quad (4.5.2.16)$$

using equation (4.5.2.12) shows that

$$I_l(R) = \sum_n |G_n(R)|^2, \quad (4.5.2.17)$$

where, as usual, the sum is over the values of  $n$  that satisfy the helix selection rule. On the diffraction pattern, reciprocal space  $(R, \psi, Z)$  collapses to the two dimensions  $(R, Z)$ . The  $R$  axis is called the *equator* and the  $Z$  axis the *meridian*. The layer planes collapse to *layer lines*, at  $Z = l/c$ , which are indexed by  $l$ . Equation (4.5.2.17) gives a rather simple relationship between the recorded intensity and the Fourier–Bessel structure factors.

Coherence length and disorientation, as described in Section 4.5.2.2, also affect the form of the diffraction pattern. These effects are described here, although they also apply to other than noncrystalline fibres. A finite coherence length leads to smearing of the layer lines along the  $Z$  direction. If the average coherence length of the molecules is  $l_c$ , the intensity distribution  $I_l(R, Z)$  about the  $l$ th layer line can be approximated by

$$I_l(R, Z) = I_l(R) \exp(-\pi l_c^2 [Z - (l/c)]^2). \quad (4.5.2.18)$$

It is convenient to express the effects of disorientation on the intensity distribution of a fibre diffraction pattern by writing the latter as a function of the polar coordinates  $(\rho, \sigma)$  (where  $\sigma$  is the angle with the  $Z$  axis) in  $(R, Z)$  space. Assuming a Gaussian orientation density function [equation (4.5.2.1)], if  $\alpha_0$  is small and the effects of disorientation dominate over those of coherence length (which is usually the case except close to the meridian), then the distribution of intensity about one layer line can be approximated by (Holmes & Barrington Leigh, 1974; Stubbs, 1974)

$$I(\rho, \sigma) \simeq \frac{I_l(R)}{2\pi\alpha_0 l_c \rho} \exp\left[-\frac{(\sigma - \sigma_l)^2}{2\beta^2}\right], \quad (4.5.2.19)$$

where (Millane & Arnott, 1986; Millane, 1989c)

$$\beta^2 = \alpha_0^2 + (1/2\pi l_c^2 \rho^2 \sin^2 \sigma_l) \quad (4.5.2.20)$$

and  $\sigma_l$  is the polar angle at the centre of the layer line, *i.e.*  $R = \rho \sin \sigma_l$ . The effect of disorientation, therefore, is to smear each layer line about the origin of reciprocal space.

##### 4.5.2.4.2. Polycrystalline fibres

A polycrystalline fibre is made up of crystallites that are oriented parallel to each other, but are randomly positioned and randomly rotated about their molecular axes. The recorded diffraction pattern is the intensity diffracted by a single crystallite, cylindrically averaged about the  $Z$  axis. On a fibre diffraction pattern, therefore, the Bragg reflections are cylindrically projected onto the  $(R, Z)$  plane and their positions are described

#### 4.5. POLYMER CRYSTALLOGRAPHY

by the cylindrically projected reciprocal lattice (Finkenstadt & Millane, 1998).

The molecules are periodic and are therefore usually aligned with one of the unit-cell vectors. Since the  $z$  axis is defined as the fibre axis, it is usual in fibre diffraction to take the  $\mathbf{c}$  lattice vector as the unique axis and as the lattice vector parallel to the molecular axes. It is almost always the case that the fibre is rotationally disordered about the molecular axes, *i.e.* about  $\mathbf{c}$ . Consider first the case of a monoclinic unit cell ( $\alpha = \beta = 90^\circ$ ) so that the reciprocal lattice is cylindrically projected about  $\mathbf{c}^*$ . The cylindrical coordinates of the projected reciprocal-lattice points are then given by

$$R_{hkl}^2 = h^2 a^{*2} + k^2 b^{*2} + 2hka^*b^* \cos \gamma^* \quad (4.5.2.21)$$

and

$$Z_{hkl} = lc^*, \quad (4.5.2.22)$$

so that  $R$  depends only on  $h$  and  $k$ , and  $Z$  depends only on  $l$ . Reflections with fixed  $h$  and  $k$  lie on straight *row lines*. Certain sets of distinct reciprocal-lattice points will have the same value of  $R_{hkl}$  and therefore superimpose in cylindrical projection. For example, for an orthorhombic system ( $\gamma = 90^\circ$ ) the reciprocal-lattice points  $(hkl)$ ,  $(\bar{h}kl)$ ,  $(h\bar{k}l)$  and  $(\bar{h}\bar{k}l)$  superimpose. Furthermore, the crystallites in a fibre specimen are usually oriented randomly up and down so that the reciprocal-lattice points  $(hkl)$  and  $(h\bar{k}l)$  superimpose, so that in the orthorhombic case eight reciprocal-lattice points superimpose. Also, as described below, reciprocal-lattice points that have similar values of  $R$  can effectively superimpose.

If the unit cell is either triclinic, or is monoclinic with  $\alpha \neq 90^\circ$  or  $\beta \neq 90^\circ$ , then  $\mathbf{c}^*$  is inclined to  $\mathbf{c}$  and the  $Z$  axis, and the reciprocal lattice is not cylindrically projected about  $\mathbf{c}^*$ . Equation (4.5.2.22) for  $Z_{hkl}$  still applies, but the cylindrical radius is given by

$$R_{hkl}^2 = h^2 a^{*2} + k^2 b^{*2} + l^2 [c^{*2} - (1/c^2)] + 2hka^*b^* \cos \gamma^* + 2hla^*c^* \cos \beta^* + 2klb^*c^* \cos \alpha^* \quad (4.5.2.23)$$

and the row lines are curved (Finkenstadt & Millane, 1998).

The most complicated situation arises if the crystallites are rotationally disordered about an axis that is inclined to  $\mathbf{c}$ . Reciprocal space is then rotated about an axis that is inclined to the normal to the  $a^*b^*$  plane,  $R_{hkl}$  and  $Z_{hkl}$  are both functions of  $h$ ,  $k$  and  $l$ , equation (4.5.2.23) does not apply, and reciprocal-lattice points for fixed  $l$  do not lie on layer lines of constant  $Z$ . Although this situation is rather unusual, it does occur (Daubeny *et al.*, 1954), and is described in detail by Finkenstadt & Millane (1998).

The observed fibre diffraction pattern consists of reflections at the projected reciprocal-lattice points whose intensities are equal to the sums of the intensities of the contributing structure factors. The observed intensity, denoted by  $I_l(R_{hk})$ , at a projected reciprocal-lattice point on the  $l$ th layer line and with  $R = R_{hk}$  is therefore given by (assuming, for simplicity, a monoclinic system)

$$I_l(R_{hk}) = \sum_{h', k' \in \mathcal{S}(h, k)} |F_{h'k'l}|^2, \quad (4.5.2.24)$$

where  $\mathcal{S}(h, k)$  denotes the set of indices  $(h', k')$  such that  $R_{h'k'} = R_{hk}$ . The number of independent reflections contributing in equation (4.5.2.24) depends on the space-group symmetry of the crystallites, because of either systematic absences or structure factors whose values are related.

The effect of a finite crystallite size is to smear what would otherwise be infinitely sharp reflections into broadened reflections of a finite size. If the average crystallite dimensions normal and parallel to the  $z$  axis are  $l_{\text{lat}}$  (*i.e.* in the 'lateral' direction) and  $l_{\text{axial}}$  (*i.e.* in the 'axial' direction), respectively, the profile of the reflection centred at  $(R_{hk}, Z = l/c)$  can be written as (Fraser *et al.*, 1984; Millane & Arnott, 1986; Millane, 1989c)

$$I(R, Z) = I_l(R_{hk})S(R - R_{hk}, Z - l/c), \quad (4.5.2.25)$$

where the profile function  $S(R, Z)$  can be approximated by

$$S(R, Z) = \exp[-\pi(l_{\text{lat}}^2 R^2 + l_{\text{axial}}^2 Z^2)]. \quad (4.5.2.26)$$

The effect of crystallite disorientation is to smear the reflections given by equation (4.5.2.26) about the origin of the projected reciprocal space. If the effects of disorientation dominate over those of crystallite size, then the profile of a reflection can be approximated by (Fraser *et al.*, 1984; Millane & Arnott, 1986; Millane, 1989c)

$$I(\rho, \sigma) \simeq \frac{I_l(R_{hk})}{2\pi\alpha_0 l_{\text{lat}} l_{\text{axial}} \rho} \times \exp\left(-\left[\frac{(\rho - \rho_{hkl})^2}{2\beta_\rho^2} + \frac{(\sigma - \sigma_{hkl})^2}{2\beta_\sigma^2}\right]\right), \quad (4.5.2.27)$$

where  $(\rho_{hkl}, \sigma_{hkl})$  are the polar coordinates of the reflection,

$$\beta_\rho^2 = \frac{l_{\text{lat}}^2 l_{\text{axial}}^2}{2\pi(l_{\text{lat}}^2 \sin^2 \sigma_{hkl} + l_{\text{axial}}^2 \cos^2 \sigma_{hkl})} \quad (4.5.2.28)$$

and

$$\beta_\sigma^2 = \alpha_0 + \frac{l_{\text{lat}}^2 l_{\text{axial}}^2}{2\pi\rho_{hkl}^2(l_{\text{axial}}^2 \sin^2 \sigma_{hkl} + l_{\text{lat}}^2 \cos^2 \sigma_{hkl})}. \quad (4.5.2.29)$$

Reflections that have similar enough  $(R, Z)$  coordinates overlap severely with each other and are also included in the sum in equation (4.5.2.24). This is quite common in practice because a number of sets of reflections may have similar values of  $R_{hk}$ .

##### 4.5.2.4.3. Random copolymers

*Random copolymers* are made up of a small number of different *kinds* of monomer, whose sequence along the polymer chain is not regular, but is random, or partially random. A particularly interesting class are synthetic polymers such as copolyesters that form a variety of liquid-crystalline phases and have useful mechanical properties (Biswas & Blackwell, 1988a). The structures of these materials have been studied quite extensively using X-ray fibre diffraction analysis. Because the molecules do not have an average  $c$  repeat, their diffraction patterns do not consist of equally spaced layer lines. However, as a result of the small number of distinct spacings associated with the monomers, diffracted intensity is concentrated about layer lines, but these are irregularly spaced (along  $Z$ ) and are *aperiodic*. Since the molecule is not periodic, the basic theory of diffraction by helical molecules described in Section 4.5.2.3.2 does not apply in this case. Cylindrically averaged diffraction from random copolymers is described here. Related approaches have been described independently by Hendricks & Teller (1942) and Blackwell *et al.* (1984). Hendricks & Teller (1942) considered the rather general problem of diffraction by layered structures made up of different kinds of layers, the probability of a layer at a

#### 4. DIFFUSE SCATTERING AND RELATED TOPICS

particular level depending on the layers present at adjacent levels. This is a one-dimensional disordered structure that can be used to describe a random copolymer. Blackwell and co-workers have developed a similar theory in terms of a one-dimensional paracrystalline model (Hosemann & Bagchi, 1962) for diffraction by random copolymers (Blackwell *et al.*, 1984; Biswas & Blackwell, 1988a), and this is the theory described here.

Consider a random copolymer made up of monomer units (residues) of  $N$  different types. Since the disorder is along the length of the polymer, some of the main characteristics of diffraction from such a molecule can be elucidated by studying the diffraction along the meridian of the diffraction pattern. The meridional diffraction is the intensity of the Fourier transform of the polymer chain projected onto the  $z$  axis and averaged over all possible monomer sequences. The diffraction pattern depends on the monomer (molar) compositions, denoted by  $p_i$ , the statistics of the monomer sequence (described by the probability of the different possible monomer pairs in this model) and the Fourier transform of the monomer units. Development of this model shows that the meridional diffracted intensity  $I(Z)$  can be written in the form (Blackwell *et al.*, 1984; Biswas & Blackwell, 1988a; Schneider *et al.*, 1991)

$$I(Z) = \sum_i p_i |F_i(Z)|^2 + 2 \sum_i \sum_{j \neq i} \Re\{F_i(Z)F_j^*(Z)T_{ij}(Z)\}, \quad (4.5.2.30)$$

where the summations are over the different monomer types and  $F_i(Z)$  is the axial Fourier transform of the  $i$ th monomer unit (each referenced to a common origin). The  $T_{ij}(Z)$  are most conveniently described by defining them as the  $ij$ th element of an  $N \times N$  matrix  $\mathbf{T}$ , which is given by

$$\mathbf{T} = \frac{\mathbf{P}\mathbf{M}\mathbf{X}(Z)}{\mathbf{I} - \mathbf{M}\mathbf{X}(Z)}, \quad (4.5.2.31)$$

where  $\mathbf{P}$ ,  $\mathbf{M}$ ,  $\mathbf{X}(Z)$  and  $\mathbf{I}$  are  $N \times N$  matrices.  $\mathbf{I}$  is the unit matrix and  $\mathbf{P}$  is a diagonal matrix with elements  $p_i$ . The elements  $M_{ij}$  of  $\mathbf{M}$  are the probabilities of forming  $ij$  monomer pairs and can be generated for different kinds of random sequence (*e.g.* chemical restrictions on the occurrence of particular monomer pairs, random chains, varying degrees of blockiness, tendency towards alternating sequences *etc.*) (Schneider *et al.*, 1991). The matrix  $\mathbf{X}(Z)$  is diagonal with elements equal to  $\langle \exp(i2\pi Z\zeta_i) \rangle$  where the  $\zeta_i$  are the projected monomer lengths and the average is over all chain conformations.

Equation (4.5.2.30) can be used to calculate the meridional diffraction for a particular random copolymer. The most important result of such a calculation is that intensity maxima are spaced irregularly along the meridian. The positions of the maxima depend on the monomer proportions, the sequence statistics and the projected monomer lengths.

The full cylindrically averaged diffraction pattern, denoted by  $I(R, Z)$ , from a noncrystalline specimen containing oriented random copolymer chains can be calculated by replacing  $F_i(Z)$  in equation (4.5.2.30) by  $F_i(R, Z)$ , *i.e.*

$$I(R, Z) = \sum_i p_i |F_i(R, Z)|^2 + 2 \sum_i \sum_{j \neq i} \Re\{F_i(R, Z)F_j^*(R, Z)T_{ij}(Z)\} \quad (4.5.2.32)$$

(Biswas & Blackwell, 1988a). Note that we write  $I(R, Z)$  rather than  $I_l(Z)$  since the pattern cannot be indexed on the basis of regularly spaced layer lines. The  $F_i(R, Z)$  in equation (4.5.2.32) depend on the chain conformation, since this affects the range of monomer orientations and hence their average diffraction. Chivers & Blackwell (1985) have considered two extreme cases,

one corresponding to fixed conformations between monomers and the other corresponding to completely random conformations between monomers, and have derived expressions for the diffracted intensity in both cases. Equation (4.5.2.32) allows one to calculate the fibre diffraction pattern from an array of parallel random copolymers that exhibit no lateral ordering. The diffraction pattern consists of irregularly spaced layer lines whose spacings (in  $Z$ ) are the same as those described above for the meridional maxima. Measurement of layer-line spacings and intensities and comparison with calculations based on the constituent monomers allows chain conformations to be estimated (Biswas & Blackwell, 1988a).

Diffraction patterns from liquid-crystalline random copolymers often contain sharp Bragg maxima on the layer lines. This indicates that, despite the random sequence and the possible dissimilarity of the component monomers, the chains are able to pack together in a regular way (Biswas & Blackwell, 1988b,c). Expressions that allow calculation of diffraction patterns for arrays of polymers with minimal registration, in which short, non-identical sequences form layers, have been derived (Biswas & Blackwell, 1988b,c). Calculation of diffracted intensities, coupled with molecular-mechanics modelling, allows chain conformations and packing to be investigated (Hofmann *et al.*, 1994).

##### 4.5.2.4.4. Partially crystalline fibres

In this section we address disorder in the packing of the molecules in a polycrystalline fibre. The presence of disorder within the crystallites modifies the intensities of the Bragg reflections, as well as introducing continuous diffraction. The dominant effect, sometimes seen on fibre diffraction patterns (Stroud & Millane, 1995a), is for Bragg reflections to remain at low resolution but to be replaced by continuous diffraction at high resolution. There are two distinct cases to consider. The first is where the distortions at different lattice points in the crystallite are uncorrelated, and the second is where they are correlated.

Disorder within a crystallite in a polycrystalline fibre may consist of (1) deviations in the positions of the molecules (which are treated as rigid bodies) from their positions in a regular lattice, (2) rotations of the molecules about their molecular axes from their rotational positions in an ordered crystal, and (3) random orientations (up or down) of the molecules. The first of these is called *lattice disorder*, and the second and third are components of *substitution disorder*.

Uncorrelated disorder has been treated by a number of authors (Clark & Muus, 1962; Tanaka & Naya, 1969; Fraser & MacRae, 1973; Arnott, 1980). A rather complete model has been described by Millane & Stroud (1991) and Stroud & Millane (1995b), which is presented here. If the lattice and substitution disorders are independent, and the lattice and substitution distortions at different lattice sites are uncorrelated, then the cylindrically averaged layer-line intensities  $I_l(R)$  diffracted from a fibre can be written as a sum of Bragg and diffuse (continuous) intensities (Tanaka & Naya, 1969):

$$I_l(R) = I_l^B(R) + I_l^D(R). \quad (4.5.2.33)$$

The profiles of the Bragg reflections are independent of the position of the reflection in reciprocal space. If the Cartesian components of the lattice distortions are independent, normally distributed, and the  $x$  and  $y$  components have equal variances, cylindrical averaging of the diffracted intensity can be performed analytically.

The lattice disorder consists of distortions of the two-dimensional lattice (in the lateral plane) into three-dimensional space, and in the *absence* of substitution disorder the Bragg component is given by (Stroud & Millane, 1995b)

#### 4.5. POLYMER CRYSTALLOGRAPHY

$$I_l^B(R) = I_l(R_{hk})w_{\text{lattice}}(R_{hk}, l/c) \quad (4.5.2.34)$$

where  $I_l(R_{hk})$  is given by equation (4.5.2.24), the *lattice disorder weight*,  $w_{\text{lattice}}(R, Z)$ , is given by

$$w_{\text{lattice}}(R, Z) = \exp[-4\pi^2(R^2\sigma_{\text{lat}}^2 + Z^2\sigma_{\text{axial}}^2)], \quad (4.5.2.35)$$

where  $\sigma_{\text{lat}}^2$  and  $\sigma_{\text{axial}}^2$  are the variances of the lattice distortions normal ('lateral') and parallel ('axial') to the  $z$  axis, respectively. The diffuse component is given by

$$I_l^D(R) = (1/A_{\text{cryst}})I_l(R)[1 - w_{\text{lattice}}(R, l/c)], \quad (4.5.2.36)$$

where  $I_l(R)$  is given by equation (4.5.2.17) and  $A_{\text{cryst}}$  is the average cross-sectional area of the crystallites. Inspection of equations (4.5.2.34) and (4.5.2.36) shows that the effect of the lattice disorder is to weight the amplitudes of the Bragg reflections down with increasing  $R$  and  $l$ , and to introduce a continuous intensity component whose amplitude increases with  $R$  and  $Z$ . Furthermore, the amplitude of the diffuse component relative to the Bragg component is inversely proportional to  $A_{\text{cryst}}$ , and therefore is not significant unless the crystallites are small.

If substitution disorder is also present, then equation (4.5.2.33) still applies but equations (4.5.2.34) and (4.5.2.36) are replaced by

$$I_l^B(R) = \left| \sum_n w_{nl} G_{nl}^2(R_{hk}) \exp[in(\psi_{hk} + \pi/2)] \right|^2 w_{\text{lattice}}(R_{hk}, l/c) \quad (4.5.2.37)$$

and

$$I_l^D(R) = (1/A_{\text{cryst}}) \sum_n |G_{nl}(R)|^2 [1 - |w_{nl}|^2 w_{\text{lattice}}(R, l/c)], \quad (4.5.2.38)$$

respectively, where the *substitution disorder weight*,  $w_{nl}$ , is given by

$$w_{nl} = \int_0^{c/2\pi} \int_0^z p(\varphi, z) \exp[i(2\pi z l/c - n\varphi)] d\varphi dz. \quad (4.5.2.39)$$

In equation (4.5.2.39),  $p(\varphi, z)$  is the probability density function (p.d.f.) that describes the substitution disorder, *i.e.* the p.d.f. for a molecule being rotated by  $\varphi$  about and translated by  $z$  along the molecular axis relative to its position in the undistorted lattice. Inspection of equations (4.5.2.37) and (4.5.2.38) shows that the substitution disorder weights the different contributing Bessel terms differently. This can lead to quite complicated effects on the diffraction pattern for various kinds of substitution disorder, resulting in different distributions and amplitudes of Bragg and diffuse diffraction over the diffraction pattern (Stroud & Millane, 1995b). If one assumes either uniform or normal distributions for  $\varphi$  and  $z$ , then expressions can be obtained for the  $w_{nl}$  in terms of the variances of the distributions of  $\varphi$  and  $z$  (Stroud & Millane, 1995b). The cases where distortions in  $\varphi$  are correlated with distortions in  $z$  (*e.g.* 'screw disorder'), and directional (up and down) disorder, can also be accommodated. This model has been shown to be capable of predicting diffraction patterns which are in good agreement with those measured from some disordered polycrystalline fibres (Stroud & Millane, 1995a).

We consider now the case of correlated packing disorder. As a result of intermolecular contacts within a polycrystalline

specimen, it is possible that distortions at one lattice site will affect the degree of distortion at neighbouring sites. Coupling between distortions at different lattice sites can be included in the model of disorder by allowing the distortions at different lattice sites to be correlated. The effect of correlated distortions on diffraction patterns is that the diffracted intensity does not separate into Bragg and diffuse components as it does in the case of uncorrelated distortions [equation (4.5.2.33)]. The intensity can be described as being diffuse on the whole diffraction pattern, with (often broad) maxima occurring at some of the reciprocal-lattice points, but with no significant maxima at other reciprocal-lattice points. The widths of the profiles of the maxima generally increase with increasing resolution, whereas the widths of the Bragg maxima resulting from uncorrelated disorder as described above are independent of resolution. A broadening of diffraction maxima with increasing resolution and blending into continuous diffraction is sometimes seen on diffraction patterns from polycrystalline fibres, indicating the presence of correlated disorder (Stroud & Millane, 1996b).

Correlated lattice disorder consists of correlated distortions of the two-dimensional lattice into three-dimensional space. A flexible model of crystalline disorder is that based on the perturbed lattice approach (Welberry *et al.*, 1980). While formulating perturbed lattices with only nearest-neighbour interactions is complicated, a more tractable approach is to base the statistics on an imposed correlation field (de Graaf, 1989; Stroud & Millane, 1996a). This approach has been used to describe cylindrically averaged diffraction from polycrystalline fibres that contain correlated lattice disorder and uncorrelated substitution disorder (Stroud & Millane, 1996a,b), and is presented here.

To develop a flexible and tractable theory for diffraction from crystallites with correlated disorder, it is necessary to formulate the problem in real space. The size and shape of a crystallite in the  $xy$  (lateral) plane is described by a *shape function*  $s_{\text{lat}}(\mathbf{r})$ , where  $\mathbf{r}$  denotes the position vector in real space, which is equal to unity inside the crystallite and zero outside. The autocorrelation of the shape function,  $t(\mathbf{r})$ , is given by

$$t(\mathbf{r}) = \int s(\mathbf{r}') s(\mathbf{r} + \mathbf{r}') d\mathbf{r}'. \quad (4.5.2.40)$$

The correlations between the  $x$  components, and between the  $y$  components, of the distortions at any two lattice sites are taken to be identical. The correlations between distortion vectors are defined in terms of lateral,  $\rho_{\text{lat}}(\mathbf{r})$ , and axial,  $\rho_{\text{axial}}(\mathbf{r})$ , correlation fields such that the correlation coefficients between components of the distortions in the  $x$  (or  $y$ ) and  $z$  directions, respectively, are equal to the correlation field evaluated for  $\mathbf{r}$  equal to the inter-site vector. Various functional forms for the correlation fields are possible, but exponential correlation functions are usually used (Stroud & Millane, 1996a). If  $t(\mathbf{r})$  and the correlation fields are circularly symmetric, then cylindrical averaging of the diffracted intensity can be performed analytically.

For a polycrystalline fibre with correlated lattice disorder and uncorrelated substitution disorder, the diffracted intensity is given by (Stroud & Millane, 1996b)

$$\begin{aligned} I_l(R) &= \sum_{j,k} t(r_{jk}) w_{\text{lat}}(R, r_{jk}) w_{\text{axial}}(l/c, r_{jk}) \\ &\times \sum_{m,n} J_{n-m}(2\pi R r_{jk}) \Re\{w_{ml} w_{nl}^* G_{ml}(R) G_{nl}^*(R)\} \\ &\times \exp[i(m-n)\varphi_{jk}] \end{aligned} \quad (4.5.2.41)$$

where  $r = |\mathbf{r}|$ , the sum over  $(j, k)$  is over all the sites of the undistorted lattice within the region occupied by the autocorrelation function,  $(r_{jk}, \varphi_{jk})$  are the polar coordinates of the

#### 4. DIFFUSE SCATTERING AND RELATED TOPICS

lattice sites, and the lateral and axial lattice disorder weights are given by

$$w_{\text{lat}}(R, r) = \exp(-4\pi^2 R^2 \sigma_{\text{lat}}^2 [1 - \rho_{\text{lat}}(r)]) \quad (4.5.2.42)$$

and

$$w_{\text{axial}}(Z, r) = \exp(-4\pi^2 Z^2 \sigma_{\text{axial}}^2 [1 - \rho_{\text{axial}}(r)]). \quad (4.5.2.43)$$

Equation (4.5.2.41) is an expression for the continuous intensity distribution along the layer lines and does not separate into Bragg and continuous components as in the case of uncorrelated disorder. However, calculations using these expressions show that the continuous intensity is sharply peaked around the projected reciprocal-lattice points at low resolution, the peaks broadening with increasing resolution until they have the character of continuous diffraction at high resolution (Stroud & Millane, 1996a). This is consistent with the character of diffraction patterns from some disordered polycrystalline fibres. A detailed study of the effects of correlated disorder on fibre diffraction patterns, and analysis of such disorder, can be found in Stroud & Millane (1996a) and Stroud & Millane (1996b).

##### 4.5.2.5. Processing diffraction data

Since the diffraction pattern from a fibre is two-dimensional, it can be collected with a single exposure of a stationary specimen. Diffraction data are collected either on film, which is subsequently scanned by a two-dimensional microdensitometer to obtain a digitized representation of the diffracted intensity, or using an electronic area detector (imaging plate, CCD camera, wire detector *etc.*) (Fraser *et al.*, 1976; Namba, Yamashita & Vonderviszt, 1989; Lorenz & Holmes, 1993). We assume here that the diffraction pattern is recorded on a flat film (or detector) that is normal to the incident X-ray beam, although other film geometries are easily accommodated (Fraser *et al.*, 1976). The fibre specimen is usually oriented with its axis normal to the incident X-ray beam, although, as is described below, it is sometimes tilted by a small angle to the normal in order to better access reciprocal space close to the meridian. The diffraction and camera geometry are shown in Fig. 4.5.2.1. Referring to this figure, P and S denote the intersections of the diffracted beam with the sphere of reflection and the film, respectively. The fibre, and therefore reciprocal space, is tilted by an angle  $\beta$  to the normal to the incident beam. The angles  $\mu$  and  $\chi$  define the direction of the diffracted beam and  $\theta$  is the Bragg angle. Cartesian and polar coordinates on the film are denoted by  $(u, v)$  and  $(r, \varphi)$ , respectively, and  $D$  denotes the film-to-specimen distance.

Inspection of Fig. 4.5.2.1 shows that the cylindrical  $(R, \psi, Z)$  and spherical  $(\rho, \psi, \sigma)$  polar coordinates in reciprocal space are related to  $\mu$  and  $\chi$  by

$$\rho = (1/\lambda)[2(1 - \cos \mu - \cos \chi)]^{1/2}, \quad (4.5.2.44)$$

$$Z = (1/\lambda)[\sin \beta(1 - \cos \mu \cos \chi) + \cos \beta \sin \chi]^{1/2}, \quad (4.5.2.45)$$

$$R = (\rho^2 - Z^2)^{1/2}, \quad (4.5.2.46)$$

$$\sin \psi = \frac{\sin \mu \cos \chi}{R\lambda} \quad (4.5.2.47)$$

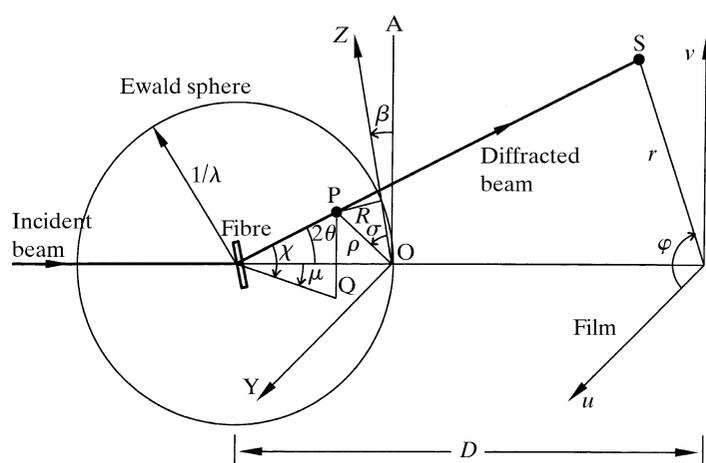


Fig. 4.5.2.1. Fibre diffraction geometry (see text). O is the origin of reciprocal space and  $\mathbf{OA}$  is normal to the incident X-ray beam. Reciprocal space is rotated about  $\mathbf{OY}$  so that the Z axis is inclined at an angle  $\beta$  to  $\mathbf{OA}$ . Q is the projection of P onto the plane containing the incident beam and  $\mathbf{OY}$ .

and

$$\tan \sigma = R/Z. \quad (4.5.2.48)$$

The coordinates on the film are related to  $\mu$  and  $\chi$  by

$$u = D \tan \mu \quad (4.5.2.49)$$

and

$$v = D \cos \mu \tan \chi, \quad (4.5.2.50)$$

and we also have that

$$r = D \tan 2\theta. \quad (4.5.2.51)$$

Use of the above equations allows the reciprocal-space coordinates to be calculated from film-space coordinates, and *vice versa*. The film coordinates  $(u, v)$  represent a relatively undistorted map of reciprocal space  $(R, Z)$ , except near the  $v$  (vertical) axis of the diffraction pattern. The meridian of reciprocal space does not map onto the film. Inspection of Fig. 4.5.2.1 shows that the only point on the meridian that does appear on the film is at  $Z = \lambda^{-1} \sin \beta$ . The region *close* to the meridian that appears on the film can therefore be manipulated by adjusting the fibre tilt.

The film-to-specimen distance can be determined by including with the specimen a crystalline power that gives a diffraction ring of known spacing and adjusting the film-to-specimen distance so that the calculated and observed rings coincide. A nonzero fibre tilt leads to differences between the upper and lower halves of the diffraction pattern, and these differences can be used to determine the tilt. This can be done by either calculating the  $\rho$  and  $\chi$  values for several sets of the same reflection above and below the equator and using the relationship

$$\tan \beta = \frac{\sin \chi_U + \sin \chi_L}{\lambda^2 \rho^2}, \quad (4.5.2.52)$$

where  $\chi_U$  and  $\chi_L$  refer to the upper and lower ( $\chi < 0$ ) reflections (Millane & Arnott, 1986; Lorenz & Holmes, 1993), or by finding the tilt that minimizes the differences between optical densities at the same reciprocal-space coordinates above and below the

#### 4.5. POLYMER CRYSTALLOGRAPHY

equator (Fraser *et al.*, 1976). The optical densities may also be corrected for the effects of film (or detector) nonlinearity (Fraser *et al.*, 1976) and the effects of variable absorption owing to the oblique passage of the beam through the film using expressions given by Fraser *et al.* (1976) and Lorenz & Holmes (1993).

Accurate subtraction of background diffraction is important in order to obtain accurate intensity measurements. One approach to estimating background diffraction is to fit a global background function, usually expanded as a polynomial (Lorenz & Holmes, 1993) or a Fourier–Bessel (Millane & Arnott, 1985) series, to optical densities at a set of points on the diffraction pattern that represent background alone. The background function may or may not be circularly symmetric. The background function is subtracted from the whole diffraction pattern. Another approach, suitable only for Bragg diffraction patterns, is to fit a plane under each reflection, either to the peripheral regions of the reflection or as part of a profile-fitting procedure (Fraser *et al.*, 1976). A different plane is required for each reflection. A third approach, more suitable for continuous diffraction patterns, is to fit a one-dimensional polynomial in angle, for each value of  $r$  on the film, possibly as part of a deconvolution procedure (Makowski, 1978). Recently, Ivanova & Makowski (1998) have described an iterative low-pass filtering technique for estimating the background on diffraction patterns from poorly oriented specimens in which there is little space between the layer lines for sampling the background.

A polarization correction is applied to the diffraction pattern, where for unpolarized X-rays (laboratory sources) the polarization factor  $p$  is given by (Fraser *et al.*, 1976)

$$p = (1 + \cos^2 2\theta)/2. \quad (4.5.2.53)$$

The diffraction pattern is usually mapped into reciprocal space  $(R, Z)$  for subsequent analysis. The mapping is performed by assigning to the intensity  $I(R, Z)$  at position  $(R, Z)$  in reciprocal space the value given by (Fraser *et al.*, 1976)

$$I(R, Z) = \frac{I(u(R, Z); v(R, Z))}{\cos^3 \mu \cos^3 \chi}, \quad (4.5.2.54)$$

where  $I(u; v)$  denotes the intensity on the film. The functions  $u(R, Z)$  and  $v(R, Z)$  can be derived from the equations given above. Note that equation (4.5.2.54) includes, implicitly, the Lorentz factor.

Subsequent processing depends on whether the diffraction pattern is continuous (*i.e.* from a noncrystalline specimen) or Bragg (*i.e.* from a polycrystalline specimen). Diffraction patterns from partially crystalline specimens that contain both components have been analysed using a combination of both approaches (Arnott *et al.*, 1986; Park *et al.*, 1987).

For a diffraction pattern containing continuous diffraction on layer lines, one usually extracts the cylindrically averaged transform  $I_l(R)$  from the intensity  $I(R, Z)$  on the diffraction pattern mapped into reciprocal space. This involves correcting for the effects of coherence length and disorientation expressed by equation (4.5.2.19), and for the overlap of the smeared layer lines that results from their increasing width with increasing  $R$ . The diffracted intensity  $I(\rho, \sigma)$  in polar coordinates in reciprocal space is equal to the sum of the diffraction  $I_l(\rho, \sigma)$  due to each (overlapping) layer line so that

$$I(\rho, \sigma) = \sum_l I_l(\rho, \sigma). \quad (4.5.2.55)$$

Referring to equations (4.5.2.55), (4.5.2.19) and (4.5.2.20) shows that if the smearing due to disorientation dominates over that

due to coherence length, then for fixed  $\rho$ , equation (4.5.2.55) represents a convolution along  $\sigma$  of the layer-line intensities  $I_l(\rho \sin \sigma_i)$  with the Gaussian angular profile in equation (4.5.2.19). By mapping the intensity  $I(R, Z)$  into polar coordinates as  $I(\rho, \sigma)$ , or by simply sampling  $I(R, Z)$  for fixed  $\rho$  and equally spaced samples of  $\sigma$ ,  $I_l(R)$  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  $\sigma$  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  $\alpha_0$  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

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

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

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

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  $l_{\text{lat}}$ ,  $l_{\text{axial}}$  and  $\alpha_0$ , equation (4.5.2.57) can be written as a system of linear equations that can be solved

## 4. DIFFUSE SCATTERING AND RELATED TOPICS

for the intensities  $I_l(R_{hk})$  from the data  $I(R, Z)$  on the diffraction pattern. The parameters  $l_{\text{lat}}$ ,  $l_{\text{axial}}$  and  $\alpha_0$ , 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 (<http://www.ccp13.ac.uk/>) (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  $R_{\text{min}}$  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  $R_{\text{min}}$  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

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

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.*

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

The  $\hat{\phantom{P}}$  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 auto-correlation 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

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

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

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

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 nonsystematic 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  $hk0$  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

#### 4. DIFFUSE SCATTERING AND RELATED TOPICS

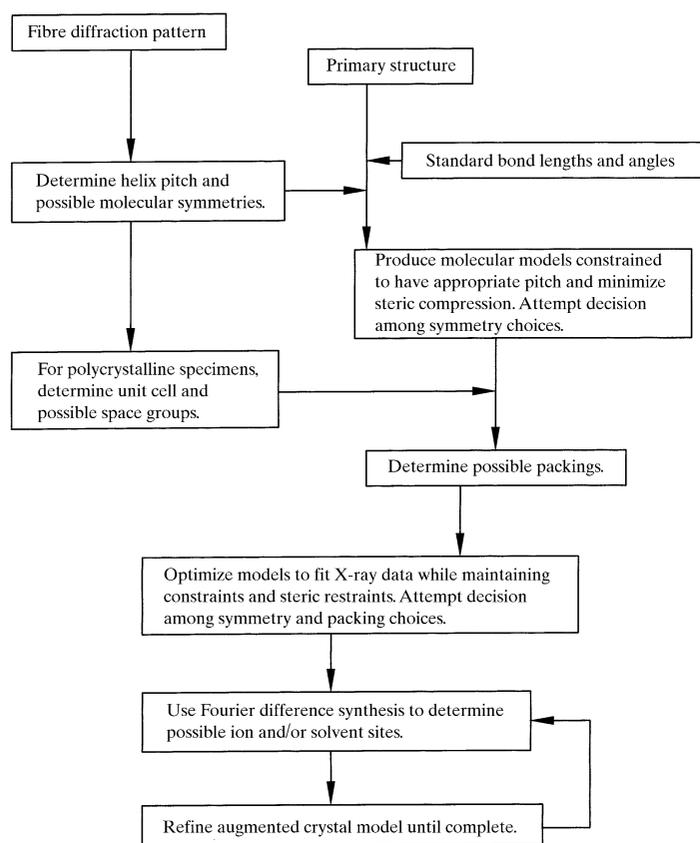


Fig. 4.5.2.2. Flow chart of the molecular-model-building approach to structure determination (Arnett, 1980).

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.

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 & Arnett, 1978; Millane *et al.*, 1985), originally developed by Arnett and co-workers in the early 1960s (Arnett & Wonacott, 1966). This system has been used to determine the structures of a wide variety of polynucleotides, polysaccharides, polyesters and polypeptides (Arnett, 1980; Arnett & Mitra, 1984; Chandrasekaran & Arnett, 1989; Millane, 1990c). 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 & Arnett, 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 overshoot nonbonded interatomic distances. These are incorporated by a quadratic nonbonded potential that is matched to a Buckingham potential (Campbell Smith & Arnett, 1978). A variety of other restraints can also be incorporated.

In the LALS system, the quantity  $\Omega$  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  $\Delta F_m$  between the model and experimental X-ray amplitudes – Bragg and/or continuous. The term  $C$  involves restraints to ensure that overshoot 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 & Arnett, 1978). The  $\omega_m$  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  $\lambda_m$  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  $\Omega$  is minimized using full-matrix nonlinear least squares and singular value decomposition (Campbell Smith & Arnett, 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  $(\Omega_p/\Omega_0)^{1/2}$  and  $(X_p/X_0)^{1/2}$  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

#### 4.5. POLYMER CRYSTALLOGRAPHY

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, 1999a)

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

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, 1999b).

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  $\alpha_c$  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 (1987a) 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, 1987a). 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-

#### 4. DIFFUSE SCATTERING AND RELATED TOPICS

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

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

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  $\Phi$  chosen to minimize the mean-square difference. The values of  $\Phi$  found for each layer line can then be combined to determine  $\varphi_h$  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 ( $r_h, \varphi_h, z_h$ ).

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  $j$ th derivative, so that

$$G_{n,j} = G_n + g_{n,j}, \quad (4.5.2.69)$$

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  $j$ th heavy-atom structure alone. Equation (4.5.2.17) can then be written as

$$I = \sum_n (A_n^2 + B_n^2) \quad (4.5.2.70)$$

for the native and

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

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

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

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 leveling. 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

## 4.5. POLYMER CRYSTALLOGRAPHY

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

$$E = E_c + S \sum_i \sum_i w_{ii} \{ [I_i^o(R_i)]^{1/2} - k [I_i^c(R_i)]^{1/2} \}^2, \quad (4.5.2.73)$$

where  $E_c$  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),  $I_i^o(R_i)$  and  $I_i^c(R_i)$  are the observed and calculated, respectively, cylindrically averaged diffraction intensities sampled at  $R = R_i$ , the  $w_{ii}$  are weights for the observed intensities  $I_i^o(R_i)$  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  $\alpha$ -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<sub>13</sub> helix symmetry with a  $c$  repeat of 216.5 Å, and the high-temperature form (that discussed in the previous paragraph) has 27<sub>5</sub> 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  $\alpha$ -helical polyaniline 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  $\alpha$ -helical polyaniline 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

#### 4. DIFFUSE SCATTERING AND RELATED TOPICS

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

$$R = \frac{\sum_i ||F_i^o - |F_i^c||}{\sum_i |F_i^o|}, \quad (4.5.2.74)$$

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, 1990a). Making use of these statistics shows that the largest likely  $R$  factor,  $R_m$ , for  $m$  components is given by (Stubbs, 1989; Millane, 1989a)

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

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, 1989a). The expression in equation (4.5.2.75) for  $R_m$  can be

written in various approximate forms (Millane, 1990d, 1992a), the simplest being

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

(Millane, 1990d), 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, 1989b)

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

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

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

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, 1989b, 1992b). 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, 1992b):

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

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, 1992b).

Largest likely  $R$  factors in fibre diffraction studies are typically between about 0.3 and 0.5, depending on the particular structure (Millane, 1989b, 1992b; 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

BY 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 2.5.8).

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, epitaxial orientation, *in situ* polymerization on a substrate, in a Langmuir–Blodgett layer, *in situ* polymerization within a thin layer and polymerization in dilute solution. If these preparations do not provide sufficient information, then data can also be collected from microfibrils. Thin cast films have also been examined after stretching.

Self-seeding (Blundell *et al.*, 1966) has been one of the most important techniques for growing single chain-folded lamellae. The technique is very simple. A dilute suspension of the polymer is made in a poor solvent. The temperature is raised to cause total solubilization of the macromolecule and then lowered to room temperature to crystallize ill-formed particles (mostly dendrites). The temperature is then elevated again until the suspension just clears, leaving small seeds of the polymer crystals behind. Upon lowering to a suitable temperature above ambient, which is then fixed, isothermal crystallization of well formed lamellae is allowed to occur over time. When the crystallization procedure is complete, the suspension can be cooled again to room temperature and the lamellae harvested. These lamellae are typically less than 10 nm thick, with lateral dimensions between 1.0 and 10.0  $\mu\text{m}$ .

Epitaxial orientation techniques, to give alternative projections of the chain packing, have become increasingly important in recent years. While inorganic substrates have been described (Mauritz *et al.*, 1978), the use of organic layers for this purpose (Wittmann & Lotz, 1990; Lotz & Wittmann, 1993) has been more promising because these substrates are less easily contaminated by adsorbed gases and water vapour, and because the nucleation is anisotropic. Often the crystallization can be carried out from a cooled co-melt, *i.e.* a dilute solution of the polymer in the organic small molecule. When the liquidus curve of the eutectic phase diagram is crossed, the diluent crystals form first. Since these have a surface lattice spacing closely resembling that of the polymer-chain packing, the polymer chains can be directed to lie *along* the substrate surface, rather than normal to it, as the solidus line of the phase diagram is crossed. The substrate can then be removed by some suitable technique (sublimation, selective solvation) to permit the investigation of the oriented film. Variations of this procedure include crystallization of polymer-chain segments from the vapour phase onto a substrate (Wittmann & Lotz, 1985) and *in situ* crystallization of monomers that have first been epitaxially oriented on a suitable substrate (Rickert *et al.*, 1979).

A number of other possibilities for crystal growth also exist. Langmuir troughs have been used to orient monomers that may have hydrophilic moieties. If the monomers contain triple bonds that can be cross-linked, then a polymer film can be formed, *e.g.*,