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, 1990*c*; 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 $\Omega(\alpha)\,\mathrm{d}\omega$ is the fraction of molecules in an element of solid angle $\mathrm{d}\omega$ inclined at an angle α 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

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

where α_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.

4.5.2.3.1. 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, φ, 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

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

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 φ and z, it can be decomposed into a Fourier series as

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

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

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

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

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

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

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

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

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

where

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

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, ψ, 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

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

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.*

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

and the inverse transform is

$$g_{nl}(r) = \int_0^{\infty} G_{nl}(R) J_n(2\pi Rr) 2\pi R\, \mathrm{d}R. \qquad (4.5.2.11)$$

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

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

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

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

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

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 $G_{nl}(R)$ 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

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

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 nth Bessel order and its associated value of m, and using the selection rule shows that

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

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.*

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

using equation (4.5.2.12) shows that

$$I_l(R) = \sum_n |G_{nl}(R)|^2, \qquad (4.5.2.17)$$

where, as usual, the sum is over the values of n that satisfy the helix selection rule. On the diffraction pattern, reciprocal space (R, ψ, 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 lth layer line can be approximated by

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

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 (ρ, σ) (where σ is the angle with the Z axis) in (R, Z) space. Assuming a Gaussian orientation density function [equation (4.5.2.1)], if α_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)

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

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

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

and σ_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