

4. DIFFUSE SCATTERING AND RELATED TOPICS

$f(r, \varphi, z)$ 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 φ and z , it can be decomposed into a Fourier series as

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

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

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

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, 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

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

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 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

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

4.5. POLYMER CRYSTALLOGRAPHY

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

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

where, as usual, the sum is over the values of n that satisfy the helix selection rule. On the diffraction pattern, reciprocal space (R, 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 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 (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)

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

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 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 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.