considered individually when calculating such composite intensities.

19.5.4. Fibre preparation

Natural fibre specimens may require only the selection of fibres or regions of fibres in which the polymers are well oriented, but many other fibres must be made in the laboratory, and orientation can often be improved by a wide variety of laboratory procedures.

Orientation often requires controlled conditions of relative humidity and temperature during the preparation of the fibre; in many cases, these conditions must be maintained during data collection. In some cases, tension must be applied to the fibres; in an increasing number of cases, magnetic fields have been found to improve orientation. Fibres may be drawn directly from concentrated polymer solutions or made by stretching gels, using weights to stretch strips of polymer films cast on Teflon blocks or applying radial heating while forming polymer films (Arnott, Guss *et al.*, 1974; Chandrasekaran, Radha, Lee & Zhang, 1994). A drop of concentrated polymer solution may simply be dried while suspended between two supports. Magnetic fields have dramatically improved the orientation in dried fibres of polymers having significant dipole moments (Torbet, 1987).

Oriented sols, generally enclosed in glass capillaries, are usually made using shearing forces, either by moving the sol in the capillary (Gregory & Holmes, 1965) or by centrifugation (Cohen *et al.*, 1971). Again, magnetic fields can greatly improve orientation, sometimes in combination with centrifugation (Yamashita, Suzuki, & Namba, 1998).

Any of these stretching or orienting processes might facilitate the growth of long microcrystallites along the fibre axis. Crystallization in general and lateral organization in particular are achieved primarily by careful choice of solution conditions, including solvent, pH, additives, relative humidity and temperature. In both crystalline and noncrystalline specimens, annealing processes are often important to both crystallization and orientation.

19.5.5. Data collection

Fibre-diffraction data have generally been collected using laboratory X-ray sources and photographic film. However, synchrotron sources are increasingly being used (Shotton *et al.*, 1998), taking advantage of reduced exposure time, the potential for time-resolved studies and the fact that many fibres (or the well oriented regions of fibres) are too small for laboratory data collection. Imaging-plate systems and charge-coupled device (CCD) cameras are replacing film as detectors (Yamashita *et al.*, 1995; Okuyama *et al.*, 1996; Shotton *et al.*, 1998). Pinhole cameras, mirror-monochromator optics and double-mirror optics are used in different applications. Diffraction by most fibres is inherently weak, and very long repeat spacings often require long distances between the specimen and the detector, so fibre cameras are often flushed with helium to reduce air scatter. Constant relative humidity is often required and is achieved by bubbling the helium stream through a saturated salt solution followed by a salt trap.

The X-ray beam commonly strikes a stationary fibre perpendicular to the fibre axis. Because of the cylindrical averaging of the data, this procedure allows most of the diffraction pattern to be collected in a single exposure. There is, however, a 'blind region' around the meridian, where the Ewald sphere does not intersect the diffraction pattern (Fraser *et al.*, 1976). Data in this region are collected by tilting the fibre.

19.5.6. Data processing

19.5.6.1. *Coordinate transformation*

Data must be transformed from detector space into reciprocal space (Fraser *et al.*, 1976). Transformation of coordinates requires determination of the origin of the diffraction pattern in detector space, the fibre tilt angle β, the twist angle (often called in-plane tilt, the inclination of the projection of Z along the beam to the detector coordinate system) and the specimen-to-detector distance. It may also require determination of the detector mis-setting angles (the deviation of the normal to the detector plane from the beam). All of these parameters can be determined by comparing equivalent reflections in the diffraction pattern.

Most data-processing programs determine the transformation parameters by some form of minimization of the deviation from equivalence in the positions of well resolved equivalent reflections. The tilt was traditionally determined by comparing the apparent Z values of equivalent reflections, but the apparent value of R for near-meridional reflections is much more sensitive to tilt. The minimization set should therefore include some near-meridional reflections if the tilt value is to be determined accurately. The helical repeat distance and, for polycrystalline fibres, the unit-cell parameters must also be determined at this time, but helical repeat distance and specimen-to-detector distance are so highly correlated that it is not often practical to refine both.

19.5.6.2. *Intensity correction*

Data intensities must be corrected for geometric and polarization effects (Fraser *et al.*, 1976; Millane & Arnott, 1986). The geometric correction has two components: a factor due to the geometry of the intersection between the diffraction pattern in reciprocal space and the sphere of reflection, and a factor due to the angle of incidence of the diffracted beam on the detector. The first factor is analogous to the Lorentz factor in crystallography, which arises because of the time taken for a reflection from a moving sample to pass through the Ewald sphere. The geometric correction can be applied to each data point as a single correction (Fraser *et al.*, 1976); this is the simpler procedure for diffraction from noncrystalline fibres. For crystalline fibres, it is often convenient to apply Lorentz and polarization corrections to each data point, to integrate the intensities within each reflection, and then to apply the remaining geometric corrections (Millane & Arnott, 1986). The Lorentz correction is

$$1/L = 2\pi \sin\theta[\cos^2\theta\cos^2\beta - (\cos\sigma - \sin\theta\sin\beta)^2]^{1/2}, \quad (19.5.6.1)$$

where θ is the Bragg angle and $\tan\sigma = R/Z$ (Millane & Arnott, 1986). The polarization correction is

$$p = (1 + \cos^2 2\theta)/2. \quad (19.5.6.2)$$

Intensities should be divided by Lp. Intensities may also be corrected for nonlinearity of detector response and for absorption by the specimen and by detector components.

19.5.6.3. *Background subtraction*

The background can be very high in fibre-diffraction data because of long exposure times and scattering from amorphous material. Because of specimen disorientation, fibre-diffraction data often contain large regions where there is no space between layer lines, so local-background-fitting methods are rarely useful. The background may be determined by fitting an analytical function to intensities at points between reflections (Millane & Arnott, 1985; Lorenz & Holmes, 1993), or by fitting a function that includes both signal and background components to the reflection data. This type of profile fitting has been described for individual reflections (Fraser *et al.*, 1976), for data in concentric rings about the centre of the

diffraction pattern (Makowski, 1978) and for entire data sets (Yamashita *et al.*, 1995; Ivanova & Makowski, 1998).

19.5.6.4. *Integration of crystalline fibre data*

The variation of reflection shape in detector space can be determined using a few sharp reflections and taking into account parameters related to crystallite size and disorientation in the specimen (Millane & Arnott, 1986). This allows the integration boundary of a reflection to be determined. Sometimes, the boundary encompasses two or more reflections too close to separate; such reflections are considered to constitute a composite reflection.

19.5.6.5. *Integration of continuous data*

In diffraction from noncrystalline fibres, intensity is a function of R on each layer line. Angular deconvolution (Makowski, 1978; Namba & Stubbs, 1985; Yamashita *et al.*, 1995) or profile fitting (Millane & Arnott, 1986) corrects for disorientation and overlap between adjacent layer lines and may also incorporate background subtraction. The intensity determined in this way should be corrected for geometric and other effects if this has not been done previously (Section 19.5.6.2; Namba & Stubbs, 1985; Millane & Arnott, 1986).

19.5.7. Determination of structures

If the amplitude and phase of each diffracted wave are known, structure determination is, in principle, straightforward (Section 19.5.3.4). In practice, however, the phase problem for fibres is more acute than for single crystals because of the limited resolution of the data, and because the diffracted intensities overlap as a result of disorientation and cylindrical averaging. Patterson methods (MacGillavry & Bruins, 1948; Stubbs, 1987) have sometimes been useful, but the cylindrically averaged Patterson function is usually too complicated for detailed interpretation. Phasing by heavy-atom methods is not practical for polymers with small unit cells because of the difficulties in incorporating heavy atoms into the structures. Structures having small unit cells are instead determined by constructing initial models based on chemical information and the observed helical parameters. Extensions of the isomorphous-replacement method (Namba & Stubbs, 1985) have been useful in determining structures, such as those of helical viruses, in which the unit cells are much larger. In all cases, refinement and evaluation of the model structures are essential. A flow chart of the sequential steps in the determination and refinement of fibre structures with small unit cells is shown in Fig. 19.5.7.1.

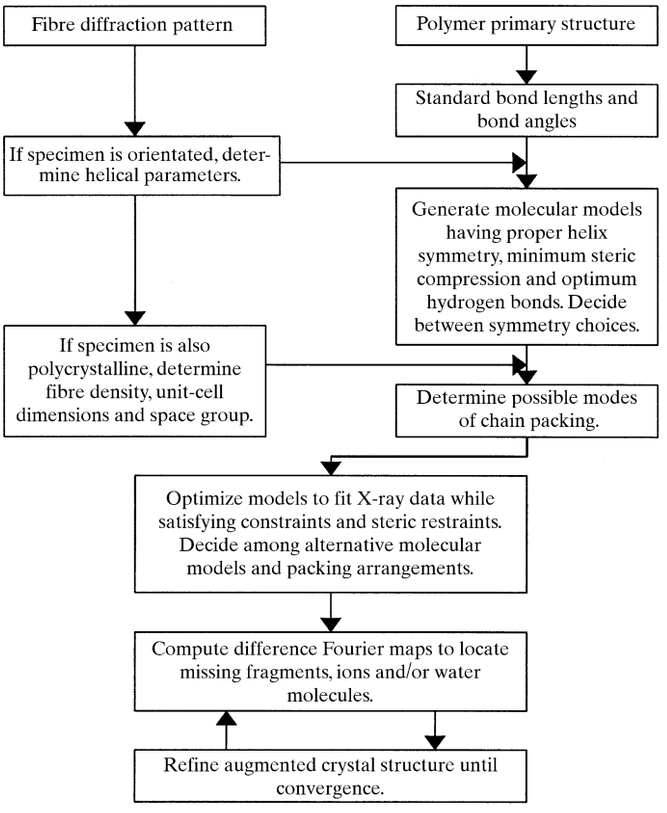

Fig. 19.5.7.1. Flow chart of the principal steps in the determination and refinement of fibre structures with small unit cells.

19.5.7.1. *Initial models: small unit cells*

For many biopolymers, especially polypeptides, polynucleotides and polysaccharides, the repeating unit is a monomer or a small oligomer and the unit-cell dimensions are in the range 10 to 50 Å. Such unit cells can accommodate one or more polymer helices, packed in an organized fashion.

An initial model is constructed from the primary structure of the repeating unit, using bond lengths, bond angles and some conformation angles derived from surveys of accurate single-crystal analyses. The model must satisfy the observed helical parameters and have reasonable intra- and inter-chain non-bonded, hydrogen-bonded and polar interactions.

This preliminary model provides an approximate solution to the phase problem and a starting point for refinement. Since there is no assurance that the refined model represents the true structure, however, stereochemically plausible alternatives must be carefully considered, refined and objectively adjudicated. Alternatives can include both right- and left-handed helices, single helices, and multistranded helices with parallel and antiparallel strands. The next stage involves the packing arrangement in the unit cell. If two or more helices are present, their positions, orientations and relative polarities must be varied in refinement.

19.5.7.2. *Refinement: small unit cells*

The widely used linked-atom least-squares (*LALS*) technique (Arnott & Wonacott, 1966; Smith & Arnott, 1978) and the variable virtual bond (*PS*79) method (Zugenmaier & Sarko, 1980) were developed for fibre structures. They are similar in principle to the least-squares refinement procedure for crystalline proteins (Hendrickson, 1985), although bond lengths and bond angles are usually kept fixed in the fibre refinements. The function minimized by the *LALS* program is of the form

$$\Omega = \sum_m w_m \Delta F_m^2 + \sum_i e_i \Delta\theta_i^2 + \sum_j k_j \Delta c_j^2 + \sum_n \lambda_n G_n. \quad (19.5.7.1)$$

The first term on the right-hand side is the weighted sum of the squares of the differences, ΔF_m, between observed and calculated X-ray structure amplitudes of Bragg reflections or continuous diffraction. Either or both types of data can be used as necessary. The weights, w_m, are inversely proportional to the estimated variance of the data. The second term minimizes the differences, $\Delta\theta_i$, between the expected (standard) values of conformation and bond angles and those in the model; the weights, e_i, are based on empirically determined variances. The third term is designed to take care of non-bonded interactions and thus keep the model free from steric compression. It includes the deviations from target values of both intra- and inter-chain hydrogen bonds and the differences between acceptable and calculated non-bonded distances for those contacts that are smaller than the acceptable limiting values. The

references