equally among the superposed reflections. This is a reasonable approach in the initial stages of structure analysis, when the reliability of the model is uncertain, and has the advantage of minimizing bias toward the model. Alternatively, the observed intensity may be split in the same ratio as the calculated intensity. This approach, although biased, is more effective for locating solvent molecules and ions in an otherwise well determined structure. Difference Fourier maps have played a significant role in determining the molecular structures and packing arrangements in unit cells mediated by water molecules and cations of several polynucleotide (Chandrasekaran *et al.*, 1995, 1997) and polysaccharide helices (Winter *et al.*, 1975; Chandrasekaran *et al.*, 1988, 1998; Chandrasekaran, Radha & Lee, 1994).

In noncrystalline fibre diffraction, the superposition of intensities due to cylindrical averaging is more serious and must be taken into account. Namba & Stubbs (1987*b*) have shown that the coefficients yielding the most accurate electron-density maps of the full structure have amplitudes of $NG_o - (N-1)G_c$, where N is the number of significant terms in equation (19.5.3.7) (the number of superposed intensities), and the observed intensity is divided in the ratio of the calculated intensity. For filamentous viruses at moderate resolution, N is typically in the range four to six. As in crystallography and crystalline fibre diffraction, maps calculated from amplitudes of $F_o - F_c$ have low noise levels and are most useful for checking the accuracy of final models and for locating solvent molecules.

19.5.7.7. *Evaluation*

As in crystallography, fibre structures are evaluated by statistical measures, such as R values, and by the examination of difference maps. Fibre-diffraction R values are inherently lower than those expected in crystallography, particularly when large numbers of intensities have been superposed by cylindrical averaging (Stubbs, 1989). The largest likely R value for noncrystalline TMV at 3 Å resolution is about 0.31 and for polycrystalline DNA at 3 Å resolution it is about 0.41, both significantly less than the value of 0.59 to be expected from noncentric single-crystal analyses (Millane, 1989).

Comparison of R values alone is not necessarily a reliable way to discriminate between competing models. Such discrimination is often required for structures with small unit cells, for which alternative models are routinely refined (Sections 19.5.7.1 and 19.5.7.2). The relative merits of any pair of competing models can be assessed on the basis of several types of statistics (Arnott, 1980) using Hamilton's significance test (Hamilton, 1965), which considers not only residuals but also numbers of degrees of freedom (Section 19.5.7.3). Such a test is essential. There are many examples in the literature where R values have been lowered by the simple process of increasing the number of degrees of freedom; a decreased R value obtained in this way may or may not have any significance.

Difference Fourier maps have been used to evaluate crystalline fibre diffraction analyses for many years, for example, to reject the controversial Hoogsteen base pairing in double-stranded DNA (Arnott *et al.*, 1965), and later to discriminate between 10- and 11-fold double helices of RNA (Arnott *et al.*, 1967). Difference maps have been essential in the refinement of fibre structures with large unit cells (Namba *et al.*, 1989; Wang & Stubbs, 1994), both to identify errors in early models and to confirm that the final structures contained no major errors or omissions.

19.5.8. Structures determined by X-ray fibre diffraction

The α-helix of several synthetic polypeptides (Pauling & Corey, 1951), the double helix of DNA (Watson & Crick, 1953), the ribbon structure of cellulose (Meyer & Misch, 1937) and the low-resolution structure of tobacco mosaic virus (Barrett *et al.*, 1971) were early examples of structures determined by fibre diffraction. Early workers also examined a number of fibrous proteins (Bailey *et al.*, 1943). In the past 50 years, developments in theory and practice and the availability of fast computers have made it possible to determine and refine about 200 biological polymer structures of varying complexities. The largest repeating units in polypeptides, polynucleotides and polysaccharides solved to date correspond to a tripeptide, a tetranucleotide and a hexasaccharide, respectively.

19.5.8.1. *Polypeptides*

The structural details of the α-helix and β-sheet, the principal secondary-structure elements of proteins, have emerged from the analysis of synthetic polypeptides (Pauling & Corey, 1951, 1953). Analysis of noncrystalline fibre-diffraction patterns led to the triple-helical coiled-coil model of collagen (Ramachandran & Kartha, 1955; Rich & Crick, 1955). Recent studies on the organization of β-sheets in peptides of up to about 45 residues are providing an understanding of the molecular details of amyloid fibrils, related to Alzheimer's disease (Inouye *et al.*, 1993; Malinchik *et al.*, 1998).

19.5.8.2. *Polynucleotides*

The molecular structures of a series of DNA and RNA helices have been determined and refined using data from polycrystalline fibres (Arnott *et al.*, 1969; Chandrasekaran & Arnott, 1989). These include the canonical A, B and C forms of DNA, corresponding, respectively, to 11-, 10- and 9.3-fold right-handed antiparallel Watson–Crick base-paired helices. Structural differences between the three have been attributed to changes in furanose puckerings and helical parameters: the A form has C3-*endo*, but B and C have C2-*endo* or analogous C3-*exo* puckers. All RNA duplexes are members of the A family. Later important structures included the sixfold single helix of poly (C) (Arnott *et al.*, 1976), a compact eightfold double helix for poly d(AT) and poly d(IC) (Arnott *et al.*, 1983), and the left-handed Z-DNA for poly d(GC) (Arnott *et al.*, 1980). Difference Fourier syntheses were instrumental in locating a spine of water molecules in the minor groove and a series of sodium ions and water molecules that bridge the phosphate groups of adjacent DNA molecules in the tenfold helices of poly (dA)·poly (dT) (Chandrasekaran *et al.*, 1995), poly (dA)·poly (dU) and poly d(AI)·poly d(CT) (Chandrasekaran *et al.*, 1997). Data from noncrystalline fibres have been used to determine, among others, the structures of DNA·RNA hybrid duplexes (Arnott *et al.*, 1986), a DNA triple-stranded helix (Chandrasekaran *et al.*, 2000*a*) and two RNA triple-stranded helices (Chandrasekaran *et al.*, 2000*b*,*c*). In each case mentioned, the best model was clearly preferred statistically (Hamilton, 1965) and had an R value between 0.2 and 0.3 to about 3 Å resolution.

19.5.8.3. *Polysaccharides*

Among the three-dimensional structures determined for industrially useful and biologically important polysaccharides are the gel-forming calcium *i*-carrageenan (Arnott, Scott *et al.*, 1974), sodium pectate (Walkinshaw & Arnott, 1981), gellan (Chandrasekaran *et al.*, 1988) and welan (Chandrasekaran, Radha & Lee, 1994), and a series of distinct helical forms of the glycosaminoglycan hyaluronan (Arnott & Mitra, 1984). The conformations of these molecules are delicately controlled by ions, such as sodium, potassium and calcium. The repeating units range from a simple monosaccharide to a branched pentasaccharide.

Specific interactions among the polysaccharides and their associated small molecules can be correlated with their observed properties. A number of neutral polysaccharides, such as cellulose, chitin and mannan, are twofold ribbon-like helices, which aggregate

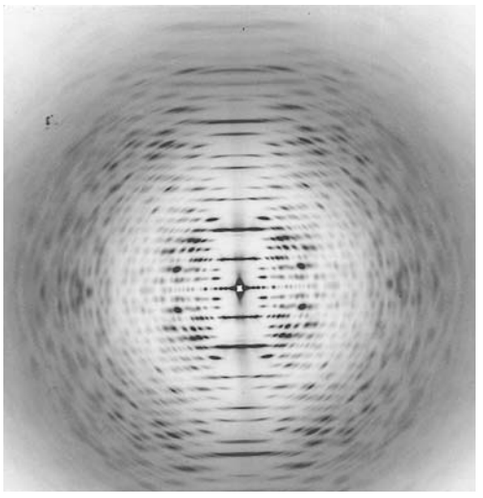

Fig. 19.5.8.1. X-ray diffraction pattern from an oriented sol of the U2 strain of tobacco mosaic virus.

and are hence water insoluble. The *A* and *B* forms of amylose, the main constituents of starch granules, are sixfold left-handed parallel double helices. Derivatization of amylose leads to the formation of single helices (Chandrasekaran, 1997). The water-soluble galactomannan derives its high viscosity in aqueous solution from intermolecular side-chain interactions (Chandrasekaran *et al.*, 1998).

19.5.8.4. *Helical viruses and bacteriophages*

The largest repeating units in structures determined by fibre diffraction are those of several members of the tobamovirus family, including tobacco mosaic virus (Namba *et al.*, 1989), cucumber green mottle mosaic virus (Wang & Stubbs, 1994) and ribgrass mosaic virus (Wang *et al.*, 1997). These viruses are rod-shaped, 3000 Å long and about 180 Å in diameter. Oriented sols yield exceptionally good diffraction patterns (Fig. 19.5.8.1). The asymmetric unit consists of a protein subunit of approximate molecular weight 18 000 Da and three nucleotides of RNA. The coat proteins are folded like globular proteins and are about 40% α-helical, with small regions of β-sheet. All of the amino acids, all three nucleotides, and in some cases water molecules and calcium ions, are seen in the electron-density maps. The TMV structure was determined by MDIR; the remaining structures were determined by molecular replacement from TMV or by a combination of molecular replacement and isomorphous replacement. All of the structures were refined by restrained least-squares or molecular-dynamics methods to *R* values of less than 0.10 at resolutions between 2.9 and 3.5 Å.

Several filamentous bacteriophage structures, including fd, Pf1 and related strains, have been determined and refined. Filamentous bacteriophages are flexible viruses, about 60 Å in diameter and 10 000 to 20 000 Å in length. Several thousand copies of a coat protein of about 50 residues wrap around a central single-stranded circular DNA. The DNA does not appear to be sufficiently ordered to appear in electron-density maps. The coat-protein molecules have an unusually simple structure, being almost entirely α-helical (Marvin *et al.*, 1974). Model-building approaches have therefore been used, sometimes supplemented by isomorphous replacement (Bryan *et al.*, 1983). Neutron scattering from bacteriophages with selectively deuterated amino-acid residues has also been used to assist model building (Nambudripad *et al.*, 1991). Both restrained least-squares (Nambudripad *et al.*, 1991) and molecular-dynamics (Gonzalez *et al.*, 1995) refinement methods have been used. Although there is not complete agreement about the structure, the coat protein clearly forms two α-helical layers, possibly with a short intervening peptide loop (Nambudripad *et al.*, 1991).

19.5.8.5. *Other large assemblies*

Low-resolution X-ray fibre-diffraction data have been successfully used to model the structural details of a number of complex assemblies. For example, the structure of the F-actin helix at 8 Å resolution has been described by combining the single-crystal structure of the G-actin monomer with fibre-diffraction data (Holmes *et al.*, 1990). This structure, in turn, has been used to model the muscle thin filament, composed of F-actin monomers and tropomyosin, at about 25 Å resolution, both in the resting and activated states, and hence to understand the movement of tropomyosin in muscle function (Squire *et al.*, 1993). The structure of the microtubule has been determined at 18 Å resolution using information from electron microscopy and fibre diffraction (Beese *et al.*, 1987). A similar but more sophisticated approach was used for bacterial flagellar filaments at 9 Å resolution (Yamashita, Hasegawa *et al.*, 1998); the diffraction patterns obtained from these filaments are of such high quality that prospects for a complete molecular structure are excellent.

references