weights, k_j, are based on the Buckingham energy function for non-bonded contacts and empirical variances for hydrogen bonds. Finally, the fourth term imposes constraints (G_h, with Lagrange multipliers λ_h) for helix connectivity and ring closure, as in a furanose or pyranose, and it vanishes when all such constraints are satisfied. During the refinement, the structure factors are calculated with either the conventional atomic scattering factor f or with a solvent-corrected atomic scattering factor f_w (Fraser *et al.*, 1978; Chandrasekaran & Radha, 1992) given by the function

$$f_w(D) = f(D) - v\sigma_s \exp(-\pi v^{2/3} D^2), \qquad (19.5.7.2)$$

where $D = (R^2 + Z^2)^{1/2}$, σ_s is the electron density of the solvent and v is the excluded volume of the atom. If the van der Waals radius of water is taken as 2 Å, σ_s for water is 0.2984 e Å^{-3}. Equation (19.5.7.2) allows for the solvent contribution to the diffracted intensity and is particularly useful in studying hydrated fibres in which structured and amorphous water can account for up to 50% of the total mass.

19.5.7.3. *Data-to-parameter ratio*

The total number of data used in this refinement process is $M + I + J$, where M, I and J are, respectively, the number of observations in the first three terms of equation (19.5.7.1). If P is the number of parameters refined and H is the number of independent constraints in the last term, then the number of degrees of freedom of the system is $P - H$. The effective number of data is given by $D = (M + I + J) - (P - H)$. The data-to-parameter ratio (D/P), a measure of the dependability of the final results, must be greater than one for meaningful refinement. D/P is typically in the range 3 to 11 in the analysis of polynucleotide and polysaccharide structures. This ratio is comparable to those commonly reported for single-crystal structures, confirming that fibre-diffraction analysis of polymers, despite the limited number of X-ray data, can yield reliable results.

19.5.7.4. *Initial models: large unit cells*

For large macromolecular aggregates, such as viruses and cytoskeletal filaments, initial models cannot usually be devised using the primary structure of the molecule alone. The largely α-helical filamentous bacteriophages form a rare class of exceptions (Makowski *et al.*, 1980). Molecular-replacement methods, in which initial models are constructed from single-crystal structure determinations of the separated components of the aggregate or from known related structures, can be useful, but because of the limited number of data in a fibre pattern such models can sometimes be difficult to refine.

Multi-dimensional isomorphous replacement (MDIR), an extension of the isomorphous-replacement method of protein crystallography, has been useful in studying helical viruses (Stubbs & Diamond, 1975; Namba & Stubbs, 1985). The dimensions are the real and imaginary parts of the various overlapping structure factors at a given point in the diffraction pattern. Information about both the phases of the structure factors and the relative magnitudes of the overlapping structure factors is obtained from heavy-atom derivatives of the virus; at least twice as many heavy-atom derivatives as the number of significant **G** terms in equation (19.5.3.7) are required. If the structure of a related aggregate is known, MDIR can be combined with molecular replacement (Namba & Stubbs, 1987*a*; Wang & Stubbs, 1994); in this case, fewer derivatives are required.

Layer-line splitting (Franklin & Klug, 1955) arises when the helical symmetry of the scattering particles is close to, but not exactly, integral. For example, tobacco mosaic virus (TMV) has 49.02 subunits in three turns of the viral helix. In this case, the **G** terms in each layer line do not fall at exactly the same Z values in the diffraction pattern. The resulting shifts in the positions of the layer lines can be measured for the native aggregate and, in favourable cases, for heavy-atom derivatives, and used to provide additional phase information (Stubbs & Makowski, 1982). Information from electron microscopy (Beese *et al.*, 1987) and neutron scattering (Nambudripad *et al.*, 1991) has also been used.

19.5.7.5. *Refinement: large unit cells*

Refinement of fibre structures having large unit cells has many parallels to refinement in protein crystallography. Refinement in real space, especially the solvent-flattening approach, has been widely used to improve electron-density maps and is particularly valuable in structure determination of noncrystalline fibres. Since helical aggregates have finite radii, **g** terms [equation (19.5.3.6)] can be set to zero outside a maximum radius and back-transformed to obtain refined estimates of the phases of the **G** terms. More detailed solvent-flattening algorithms can also be used (Namba & Stubbs, 1985).

Molecular models can be refined by methods conceptually related to those of *LALS*. The principal difference is that bond lengths and angles are not kept fixed, but are restrained to remain close to standard values. The restrained least-squares method (Hendrickson, 1985), widely used in protein crystallography, has been adapted (Stubbs *et al.*, 1986) for fibre diffraction and used to refine a number of filamentous virus structures (Namba *et al.*, 1989; Nambudripad *et al.*, 1991). Although effective, the radius of convergence of this method is less than desired, probably because of the limited number of data available from fibre diffraction (Wang & Stubbs, 1993).

Molecular-dynamics methods have been used to increase the radius of convergence of refinement (Wang & Stubbs, 1993). The program *X-PLOR* (Brünger *et al.*, 1987) has been adapted for fibre diffraction and can handle data from both crystalline and noncrystalline fibres. A potential-energy function of the form

$$\Omega = E + S \sum_l \sum_i w_{li} \{ [I_o(R_i)]^{1/2} - [I_c(R_i)]^{1/2} \}^2 \qquad (19.5.7.3)$$

is minimized. The first term, E, is an empirical energy function that accounts for distortions in bond lengths, bond angles and conformation angles, and for non-bonded, electrostatic and hydrogen-bonding interactions. The second term accounts for the differences between the observed and calculated X-ray intensities at specific values of R_i on every layer line l; w_{li} is the weight for each observation and S is a normalizing factor. In the most effective use of this method, simulated annealing, the process of heating the structure to a temperature of 3000 to 4000 K is simulated, then the structure is cooled ('annealed') in small increments. At high temperatures, energy barriers between the starting model and structures of lower potential can be overcome; in this way, the radius of convergence of the refinement is increased.

19.5.7.6. *Difference Fourier methods*

As in crystallography, difference maps are used during refinement to correct errors and to identify missing fragments of the model and, in the final stages of refinement, to identify solvent molecules and associated ions.

In crystalline fibre diffraction, the most common difference maps use calculated phases with amplitudes of either $F_o - F_c$ or $2F_o - F_c$. In both cases, weighting the coefficients on the basis of the observed and calculated structure amplitudes has been used to minimize the root-mean-square error in the electron-density maps. Reflections superposed by cylindrical averaging do, however, present problems. One solution is to divide the observed intensity

references