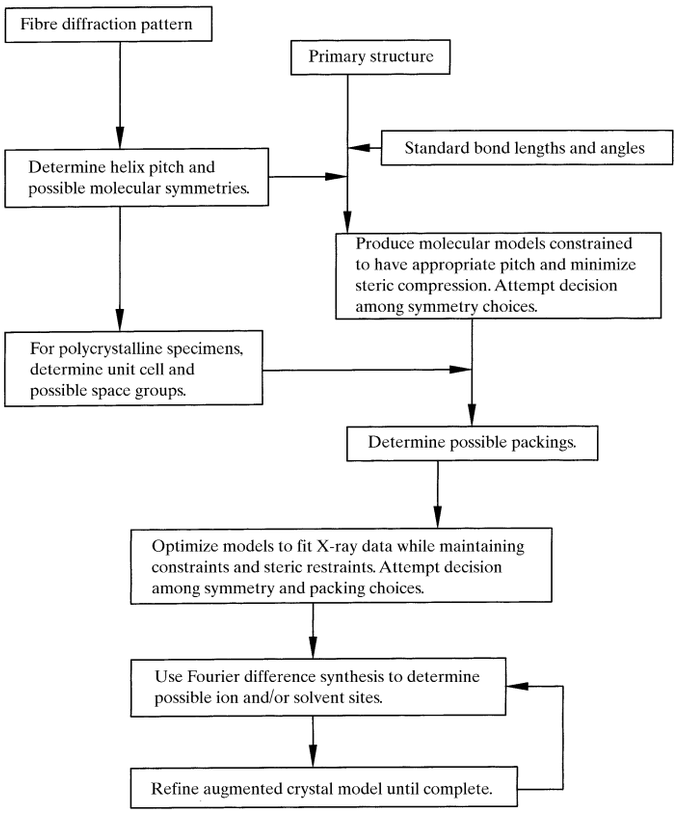

Fig. 4.5.2.2. Flow chart of the molecular-model-building approach to structure determination (Arnott, 1980).

values of v. For example, helix symmetries u_v and u_{u-v}, which correspond to the left- and right-handed helices, cannot be distinguished on the basis of the overall intensity distribution alone. Other examples of different kinds of molecular model may include single, double or multiple helices, parallel or antiparallel double helices, different juxtapositions of chains within multiple helices and different conformational domains within the molecule. For polycrystalline systems, in addition to different kinds of molecular structures, there are often different kinds of possible packing arrangements within the unit cell. There may be a number of possible packings which correspond to different arrangements within the crystallographic asymmetric unit, and there may be more than one space group that needs to be considered.

Despite the apparent large number of potential starting models implied by the above discussion, in practice the number of feasible models is usually quite small, and many of these are often eliminated at an early stage. Definition and refinement of helical polymers [steps (1) and (2) above] are carried out using computer programs, the most popular and versatile being the *linked-atom least-squares* (*LALS*) system (Campbell Smith & Arnott, 1978; Millane *et al.*, 1985), originally developed by Arnott and co-workers in the early 1960s (Arnott & Wonacott, 1966). This system has been used to determine the structures of a wide variety of polynucleotides, polysaccharides, polyesters and polypeptides (Arnott, 1980; Arnott & Mitra, 1984; Chandrasekaran & Arnott, 1989; Millane, 1990*c*). Other refinement systems exist (Zugenmaier & Sarko, 1980; Iannelli, 1994), but the principles are essentially the same and the following discussion is in terms of the *LALS* system. The atomic coordinates are defined, using a linked-atom description, in terms of bond lengths, bond angles and conformation (torsion) angles (Campbell Smith & Arnott, 1978). Stereochemical constraints are imposed, and the number of parameters reduced, by fixing the bond lengths, often (but not always) the bond angles, and possibly some of the conformation angles. The molecular conformation is then defined by the remaining parameters. For polycrystalline systems, there are usually additional variable parameters that define the packing of the molecule(s) in the unit cell. A further source of stereochemical data is the requirement that a model exhibit no overshort nonbonded interatomic distances. These are incorporated by a quadratic nonbonded potential that is matched to a Buckingham potential (Campbell Smith & Arnott, 1978). A variety of other restraints can also be incorporated.

In the *LALS* system, the quantity Ω given by

$$\Omega = \sum_m \omega_m \Delta F_m^2 + \sum_m k_m \Delta d_m^2 + \sum_m \lambda_m G_m = X + C + L \tag{4.5.2.62}$$

is minimized by varying a set of chosen parameters consisting of conformation angles, possibly bond angles, and packing parameters. The term X involves the differences ΔF_m between the model and experimental X-ray amplitudes – Bragg and/or continuous. The term C involves restraints to ensure that overshort nonbonded interatomic distances are driven beyond acceptable minimum values, that conformations are within desired domains, that hydrogen-bond and coordination geometries are close to the expected configurations, and a variety of other relationships are satisfied (Campbell Smith & Arnott, 1978). The ω_m and k_m are weights that are inversely proportional to the estimated variances of the data. The term L involves constraints which are relationships that are to be satisfied exactly ($G_m = 0$) and the λ_m are Lagrange multipliers. Constraints are used, for example, to ensure connectivity from one helix pitch to the next and to ensure that chemical ring systems are closed. The cost function Ω is minimized using full-matrix nonlinear least squares and singular value decomposition (Campbell Smith & Arnott, 1978).

Structure determination usually involves first using equation (4.5.2.62) with the terms C and L only, to establish the stereochemical viability of each kind of possible molecular model and packing arrangement. It is worth emphasizing that it is usually advantageous if the specimen is polycrystalline, even though the continuous diffraction contains, in principle, more information than the Bragg reflections (since the latter are sampled). This is because the molecule in a noncrystalline specimen must be refined in steric isolation, whereas for a polycrystalline specimen it is refined while packed in the crystal lattice. The extra information provided by the intermolecular contacts can often help to eliminate incorrect models. This can be particularly significant if the molecule has flexible sidechains. The initial models that survive the steric optimization are then optimized also against the X-ray data, by further refinement with X included in equation (4.5.2.62). The ratios $(\Omega_P/\Omega_Q)^{1/2}$ and $(X_P/X_Q)^{1/2}$ can be used in Hamilton's test (Hamilton, 1965) to evaluate the differences between models P and Q. On the basis of these statistical tests, one can decide if one model is superior to the others at an acceptable confidence level. In the final stages of refinement, bond angles may be varied in a 'stiffly elastic' fashion from their mean values if there are sufficient data to justify the increase in the number of degrees of freedom.

If sufficient X-ray data are available, it is sometimes possible to locate additional ordered molecules such as counterions or solvent molecules by difference Fourier synthesis as described in Section 4.5.2.6.5. Their positions can then be co-refined with the polymer structure while hydrogen bonds and coordination geometries are optimized. The resulting structure can then be used to compute improved phases to search for additional molecules. Since the signal-to-noise ratio in fibre difference syntheses is usually low, difference maps must be interpreted with caution. The assignment of counterions or solvent molecules to peaks in the difference synthesis must be supported by plausible