

## 2.4. ISOMORPHOUS REPLACEMENT AND ANOMALOUS SCATTERING

by simply adding  $f'$  to the normal scattering factor of the anomalous scatterers.

The effects of the imaginary component of the dispersion correction are, however, more complex. These effects could lead to serious errors in positional parameters when the space group is polar, if data in the entire diffraction sphere are not used (Ueki *et al.*, 1966; Cruickshank & McDonald, 1967). For example, accessible data in a hemisphere are normally used for X-ray analysis when the space group is  $P1$ . If the hemisphere has say  $h$  positive, the  $x$  coordinates of all the atoms would be in error when the structure contains anomalous scatterers. The situation in other polar space groups has been discussed by Cruickshank & McDonald (1967). In general, in the presence of anomalous scattering, it is desirable to collect data for the complete sphere, if accurate structural parameters are required (Srinivasan, 1972).

Methods have been derived to correct for dispersion effects in observed data from centrosymmetric and noncentrosymmetric crystals (Patterson, 1963). The methods are empirical and depend upon the refined parameters at the stage at which corrections are applied. This is obviously an unsatisfactory situation and it has been suggested that the measured structure factors of Bijvoet equivalents should instead be treated as independent observations in structure refinement (Ibers & Hamilton, 1964). The effect of dispersion corrections needs to be taken into account to arrive at the correct scale and temperature factors also (Wilson, 1975; Gilli & Cruickshank, 1973).

#### 2.4.4. Isomorphous replacement and anomalous scattering in protein crystallography

##### 2.4.4.1. Protein heavy-atom derivatives

Perhaps the most spectacular applications of isomorphous replacement and anomalous-scattering methods have been in the structure solution of large biological macromolecules, primarily proteins. Since its first successful application on myoglobin and haemoglobin, the isomorphous replacement method, which is often used in conjunction with the anomalous-scattering method, has been employed in the solution of scores of proteins. The application of this method involves the preparation of protein heavy-atom derivatives, *i.e.* the attachment of heavy atoms like mercury, uranium and lead, or chemical groups containing them, to protein crystals in a coherent manner without changing the conformation of the molecules and their crystal packing. This is only rarely possible in ordinary crystals as the molecules in them are closely packed. Protein crystals, however, contain large solvent regions and isomorphous derivatives can be prepared by replacing the disordered solvent molecules by heavy-atom-containing groups without disturbing the original arrangement of protein molecules.

##### 2.4.4.2. Determination of heavy-atom parameters

For any given reflection, the structure factor of the native protein crystal ( $\mathbf{F}_N$ ), that of a heavy-atom derivative ( $\mathbf{F}_{NH}$ ), and the contribution of the heavy atoms in that derivative ( $\mathbf{F}_H$ ) are related by the equation

$$\mathbf{F}_{NH} = \mathbf{F}_N + \mathbf{F}_H. \quad (2.4.4.1)$$

The value of  $\mathbf{F}_H$  depends not only on the positional and thermal parameters of the heavy atoms, but also on their occupancy factors, because, at a given position, the heavy atom may not often be present in all the unit cells. For example, if the heavy atom is present at a given position in only half the unit cells in the crystal, then the occupancy factor of the site is said to be 0.5.

For the successful determination of the heavy-atom parameters, as also for the subsequent phase determination, the data sets from

the native and the derivative crystals should have the same relative scale. The different data sets should also have the same overall temperature factor. Different scaling procedures have been suggested (Blundell & Johnson, 1976) and, among them, the following procedure, based on Wilson's (1942) statistics, appears to be the most feasible in the early stages of structure analysis.

Assuming that the data from the native and the derivative crystals obey Wilson's statistics, we have, for any range of  $\sin^2 \theta / \lambda^2$ ,

$$\ln \left\{ \frac{\sum f_{Nj}^2}{\langle F_N^2 \rangle} \right\} = \ln K_N + 2B_N \frac{\sin^2 \theta}{\lambda^2} \quad (2.4.4.2)$$

and

$$\ln \left\{ \frac{\sum f_{Nj}^2 + \sum f_{Hj}^2}{\langle F_{NH}^2 \rangle} \right\} = \ln K_{NH} + 2B_{NH} \frac{\sin^2 \theta}{\lambda^2}, \quad (2.4.4.3)$$

where  $f_{Nj}$  and  $f_{Hj}$  refer to the atomic scattering factors of protein atoms and heavy atoms, respectively.  $K_N$  and  $K_{NH}$  are the scale factors to be applied to the intensities from the native and the derivative crystals, respectively, and  $B_N$  and  $B_{NH}$  the temperature factors of the respective structure factors. Normally one would be able to derive the absolute scale factor and the temperature factor for both the data sets from (2.4.4.2) and (2.4.4.3) using the well known Wilson plot. The data from protein crystals, however, do not follow Wilson's statistics as protein molecules contain highly non-random features. Therefore, in practice, it is difficult to fit a straight line through the points in a Wilson plot, thus rendering the parameters derived from it unreliable. (2.4.4.2) and (2.4.4.3) can, however, be used in a different way. From the two equations we obtain

$$\begin{aligned} \ln \left\{ \frac{\sum f_{Nj}^2 + \sum f_{Hj}^2}{\sum f_{Nj}^2} \cdot \frac{\langle F_N^2 \rangle}{\langle F_{NH}^2 \rangle} \right\} \\ = \ln \left( \frac{K_{NH}}{K_N} \right) + 2(B_{NH} - B_N) \frac{\sin^2 \theta}{\lambda^2}. \end{aligned} \quad (2.4.4.4)$$

The effects of structural non-randomness in the crystals obviously cancel out in (2.4.4.4). When the left-hand side of (2.4.4.4) is plotted against  $(\sin^2 \theta) / \lambda^2$ , it is called a comparison or difference Wilson plot. Such plots yield the ratio between the scales of the derivative and the native data, and the additional temperature factor of the derivative data. Initially, the number and the occupancy factors of heavy-atom sites are unknown, and are roughly estimated from intensity differences to evaluate  $\sum f_{Hj}^2$ . These estimates usually undergo considerable revision in the course of the determination and the refinement of heavy-atom parameters.

At first, heavy-atom positions are most often determined by Patterson syntheses of one type or another. Such syntheses are discussed in some detail elsewhere in Chapter 2.3. They are therefore discussed here only briefly.

Equation (2.4.2.6) holds when the data are centric.  $F_H$  is usually small compared to  $F_N$  and  $F_{NH}$ , and the minus sign is then relevant on the left-hand side of (2.4.2.6). Thus the difference between the magnitudes of  $\mathbf{F}_{NH}$  and  $\mathbf{F}_N$ , which can be obtained experimentally, normally gives a correct estimate of the magnitude of  $\mathbf{F}_H$  for most reflections. Then a Patterson synthesis with  $(F_{NH} - F_N)^2$  as coefficients corresponds to the distribution of vectors between heavy atoms, when the data are centric. But proteins are made up of L-amino acids and hence cannot crystallize in centrosymmetric space groups. However, many proteins crystallize in space groups with centrosymmetric projections. The centric data corresponding to these projections can then be used for determining heavy-atom positions through a Patterson synthesis of the type outlined above.