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

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

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 $\sin^2\theta/\lambda^2$,

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

and

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

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

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

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.

The situation is more complex for three-dimensional acentric data. It has been shown (Rossmann, 1961) that

$$(F_{NH} - F_N)^2 \simeq F_H^2 \cos^2(\alpha_{NH} - \alpha_H) \qquad (2.4.4.5)$$

when F_H is small compared to F_{NH} and F_N. Patterson synthesis with $(F_{NH} - F_N)^2$ as coefficients would, therefore, give an approximation to the heavy-atom vector distribution. An isomorphous difference Patterson synthesis of this type has been used extensively in protein crystallography to determine heavy-atom positions. The properties of this synthesis have been extensively studied (Ramachandran & Srinivasan, 1970; Rossmann, 1960; Phillips, 1966; Dodson & Vijayan, 1971) and it has been shown that this Patterson synthesis would provide a good approximation to the heavy-atom vector distribution even when F_H is large compared to F_N (Dodson & Vijayan, 1971).

As indicated earlier (see Section 2.4.3.1), heavy atoms are always anomalous scatterers, and the structure factors of any given reflection and its Friedel equivalent from a heavy-atom derivative have unequal magnitudes. If these structure factors are denoted by $\mathbf{F}_{NH}(+)$ and $\mathbf{F}_{NH}(-)$ and the real component of the heavy-atom contributions (including the real component of the dispersion correction) by $\mathbf{F}_H$, then it can be shown (Kartha & Parthasarathy, 1965) that

$$\left(\frac{k}{2}\right)^2 [F_{NH}(+) - F_{NH}(-)]^2 = F_H^2 \sin^2(\alpha_{NH} - \alpha_H), \qquad (2.4.4.6)$$

where $k = (f_H + f'_H)/f''_H$. Here it has been assumed that all the anomalous scatterers are of the same type with atomic scattering factor f_H and dispersion-correction terms f'_H and f''_H. A Patterson synthesis with the left-hand side of (2.4.4.6) as coefficients would also yield the vector distribution corresponding to the heavy-atom positions (Rossmann, 1961; Kartha & Parthasarathy, 1965). However, $F_{NH}(+) - F_{NH}(-)$ is a small difference between two large quantities and is liable to be in considerable error. Patterson syntheses of this type are therefore rarely used to determine heavy-atom positions.

It is interesting to note (Kartha & Parthasarathy, 1965) that addition of (2.4.4.5) and (2.4.4.6) readily leads to

$$(F_{NH} - F_N)^2 + \left(\frac{k}{2}\right)^2 [F_{NH}(+) - F_{NH}(-)]^2 \simeq F_H^2. \qquad (2.4.4.7)$$

Thus, the magnitude of the heavy-atom contribution can be estimated if intensities of Friedel equivalents have been measured from the derivative crystal. F_{NH} is then not readily available, but to a good approximation

$$F_{NH} = [F_{NH}(+) + F_{NH}(-)]/2. \qquad (2.4.4.8)$$

A different and more accurate expression for estimating F_H^2 from isomorphous and anomalous differences was derived by Matthews (1966). According to a still more accurate expression derived by Singh & Ramaseshan (1966),

$$\begin{aligned} F_H^2 &= F_{NH}^2 + F_N^2 - 2F_{NH}F_N \cos(\alpha_N - \alpha_{NH}) \\ &= F_{NH}^2 + F_N^2 \pm 2F_{NH}F_N \\ &\quad \times (1 - \{k[F_{NH}(+) - F_{NH}(-)]/2F_N\}^2)^{1/2}. \end{aligned} \qquad (2.4.4.9)$$

The lower estimate in (2.4.4.9) is relevant when $|\alpha_N - \alpha_{NH}| < 90^\circ$ and the upper estimate is relevant when $|\alpha_N - \alpha_{NH}| > 90^\circ$. The lower and the upper estimates may be referred to as F_{HLE} and F_{HUE}, respectively. It can be readily shown (Dodson & Vijayan, 1971) that the lower estimate would represent the correct value of F_H for a vast majority of reflections. Thus, a Patterson synthesis with F_{HLE}^2 as coefficients would yield the vector distribution of heavy atoms in the derivative. Such a synthesis would normally be superior to those with the left-hand sides of (2.4.4.5) and (2.4.4.6) as coefficients. However, when the level of heavy-atom substitution is low, the anomalous differences are also low and susceptible to large percentage errors. In such a situation, a synthesis with $(F_{NH} - F_N)^2$ as coefficients is likely to yield better results than that with F_{HLE}^2 as coefficients (Vijayan, 1981).

Direct methods employing different methodologies have also been used successfully for the determination of heavy-atom positions (Navia & Sigler, 1974). These methods, developed primarily for the analysis of smaller structures, have not yet been successful in *a priori* analysis of protein structures. The very size of protein structures makes the probability relations used in these methods weak. In addition, data from protein crystals do not normally extend to high enough angles to permit resolution of individual atoms in the structure and the feasibility of using many of the currently popular direct-method procedures in such a situation has been a topic of much discussion. The heavy atoms in protein derivative crystals, however, are small in number and are normally situated far apart from one another. They are thus expected to be resolved even when low-resolution X-ray data are used. In most applications, the magnitudes of the differences between F_{NH} and F_N are formally considered as the 'observed structure factors' of the heavy-atom distribution and conventional direct-method procedures are then applied to them.

Once the heavy-atom parameters in one or more derivatives have been determined, approximate protein phase angles, α_N's, can be derived using methods described later. These phase angles can then be readily used to determine the heavy-atom parameters in a new derivative employing a difference Fourier synthesis with coefficients

$$(F_{NH} - F_N)\exp(i\alpha_N). \qquad (2.4.4.10)$$

Such syntheses are also used to confirm and to improve upon the information on heavy-atom parameters obtained through Patterson or direct methods. They are obviously very powerful when centric data corresponding to centrosymmetric projections are used. The synthesis yields satisfactory results even when the data are acentric although the difference Fourier technique becomes progressively less powerful as the level of heavy-atom substitution increases (Dodson & Vijayan, 1971).

While the positional parameters of heavy atoms can be determined with a reasonable degree of confidence using the above-mentioned methods, the corresponding temperature and occupancy factors cannot. Rough estimates of the latter are usually made from the strength and the size of appropriate peaks in difference syntheses. The estimated values are then refined, along with the positional parameters, using the techniques outlined below.

2.4.4.3. *Refinement of heavy-atom parameters*

The least-squares method with different types of minimization functions is used for refining the heavy-atom parameters, including the occupancy factors. The most widely used method (Dickerson *et al.*, 1961; Muirhead *et al.*, 1967; Dickerson *et al.*, 1968) involves the minimization of the function

$$\varphi = \sum w(F_{NH} - |\mathbf{F}_N + \mathbf{F}_H|)^2, \qquad (2.4.4.11)$$

where the summation is over all the reflections and w is the weight factor associated with each reflection. Here F_{NH} is the observed magnitude of the structure factor for the particular derivative and $\mathbf{F}_N + \mathbf{F}_H$ is the calculated structure factor. The latter obviously depends upon the protein phase angle α_N, and the magnitude and the phase angle of $\mathbf{F}_H$ which are in turn dependent on the heavy-atom parameters. Let us assume that we have three derivatives A, B and C, and that we have already determined the heavy-atom