(1987). Owing to errors, the triangle formed by F_P, F_{PH} and F_H fails to close. The *lack of closure error* ε is a function of the calculated phase angle φ_P:

$$\varepsilon(\varphi_P) = |F_{PH}|_{\rm obs} - |F_{PH}|_{\rm calc}.$$

Once an initial set of heavy-atom positions has been found, it is necessary to refine their parameters (x, y, z, occupancy and thermal parameters). This can be achieved through the minimization of

$$\sum_{\mathbf{S}} \varepsilon^2/E,$$

where E is the estimated error ($\simeq \langle(|F_{PH}|_{\rm obs} - |F_{PH}|_{\rm calc})^2\rangle$) (Rossmann, 1960; Terwilliger & Eisenberg, 1983). This procedure is safest for noncentrosymmetric reflections (φ restricted to 0 or π) if enough are present. Phase refinement is generally monitored by three factors:

$$R_{\rm Cullis} = \sum\||F_{PH} + F_P| - |F_H|_{\rm calc}|/\sum|F_{PH} - F_P|$$

for noncentrosymmetric reflections only; acceptable values are between 0.4 and 0.6;

$$R_{\rm Kraut} = \sum\||F_{PH}|_{\rm obs} - |F_{PH}|_{\rm calc}|/\sum|F_{PH}|_{\rm obs},$$

which is useful for monitoring convergence; and the

$$\text{phasing power} = \sum|F_H|_{\rm calc}/\sum\||F_{PH}|_{\rm obs} - |F_{PH}|_{\rm calc}|,$$

which should be greater than 1 (if less than 1, then the phase triangle cannot be closed *via* F_H).

The resulting phase probability is given by

$$P(\varphi_P) = \exp\{-\varepsilon^2(\varphi_P)/2E^2\}.$$

The phases have a minimum error when the *best phase* $\varphi_{\rm best}$, *i.e.* the centroid of the phase distribution,

$$\varphi_{\rm best} = \int \varphi_P P(\varphi_P)\ {\rm d}\varphi_P,$$

is used instead of the most probable phase. The quality of the phases is indicated by the *figure of merit* m, where

$$m = \int P(\varphi_P)\exp(i\varphi_P)\ {\rm d}\varphi_P \Big/ \int P(\varphi_P)\ {\rm d}\varphi_P.$$

A value of 1 for m indicates no phase error, a value of 0.5 represents a phase error of about 60°, while a value of 0 means that all phases are equally probable.

The *best Fourier* is calculated from

$$\rho_{\rm best}(\mathbf{r}) = (1/V)\sum_{\mathbf{S}} m|F_P(\mathbf{S})|\exp\{i\varphi_{P\rm best}(\mathbf{S})\},$$

where the electron density should have minimal errors.

12.2.4.2. *Automated search procedures*

If the derivative shows a high degree of substitution, then the Harker sections become more difficult to interpret. Furthermore, Terwilliger *et al.* (1987) have shown that the intrinsic noise in the difference Patterson map increases with increasing heavy-atom substitution. It is at this stage that automated procedures are invaluable.

One such automated procedure is implemented in *PROTEIN* (Steigemann, 1991). The unit cell is scanned for possible heavy-atom sites; for each search point (x, y, z), all possible Harker vectors are calculated, and the difference-Patterson-map values at these points are summed or multiplied. As the origin peak dominates the Patterson function, this region is set to zero. The resulting correlation map should contain peaks at all possible heavy-atom positions. The peak list can then be used to find a set of consistent heavy-atom locations through a subsequent search for difference vectors (cross vectors) between putative sites. It should be possible to locate all major and minor heavy-atom sites through repetition of this procedure. A similar strategy is adopted in the program *HEAVY* (Terwilliger *et al.*, 1987), but sets of heavy-atom sites are ranked according to the probability that the peaks are not random. The program *SOLVE* (Terwilliger & Berendzen, 1999) takes this process a stage further, where potential heavy-atom structures are solved and refined to generate an (interpretable) electron density in an automated fashion.

The search method can also be applied in reciprocal space, where the Fourier transform of the trial heavy-atom structure is calculated, and the resulting $F_{H\rm calc}$ is compared to the measured differences between derivative and native structure-factor amplitudes (Rossmann *et al.*, 1986). In the programme *XtalView* (McRee, 1998), the correlation coefficient between $|F_H|$ and $|F_{PH} - F_P|$ is calculated, whilst a correlation between F_H^2 and $(F_{PH} - F_P)^2$ is used by Badger & Athay (1998). Dumas (1994*b*) calculates the correlation between $|F_{H\rm calc}|^2$ and $|F_{H\rm estimated}|^2$, based on the estimated lack of isomorphism.

Vagin & Teplyakov (1998) have reported a heavy-atom search based on a reciprocal-space translation function. In this case, low-resolution peaks are not removed but weighted down using a Gaussian function. Potential solutions are ranked not only according to their translation-function height, but also through their phasing power, which appears to be a stronger selection criterion.

All these searches are based upon the sequential identification of heavy-atom sites and their incorporation in a heavy-atom partial structure. Problems arise when bogus sites influence the search for further heavy-atom positions. In an attempt to overcome this problem, the heavy-atom search has been reprogrammed using a genetic algorithm, with the Patterson minimum function as a selection criterion (Chang & Lewis, 1994). This approach has the potential to reveal all heavy-atom positions in one calculation, and tests on model data have shown it to be faster than traditional sequential searches.

12.2.5. Special complications

12.2.5.1. *Lack of isomorphism*

This problem is by far the most common in protein crystallography. An isomorphous derivative is one in which the crystalline arrangement has not been disturbed by derivatization. An early study of Crick & Magdoff (1956) proposed a rule of thumb that a change in any of the cell dimensions by more than around 5% would result in a lack of isomorphism that would defeat any attempt to locate the heavy-atom positions or extract useful phase information. Lack of isomorphism can, however, be more subtle; sometimes a natural variation in the native crystal form may occur, resulting in poor merging statistics of data obtained from different crystals. Coupling this variation with commonly observed structural changes upon heavy-atom binding can provide a considerable barrier to obtaining satisfactory phases. Dumas (1994*a*) has provided a theoretical consideration of this problem.

One practical approach is to collect native and derivative data sets from the same crystal, a technique that has been successful in the structure determination of cyclohydrolase (Nar *et al.*, 1995), proteosome (Löwe *et al.*, 1995) and a number of other proteins. Nonisomorphism can be used, however. In the structure solution of carbamoyl sarcosine hydrolase (Romao *et al.*, 1992), derivatives fell into two (related) crystalline classes. By judicious use of two 'native' crystal forms, heavy-atom positions could be obtained in each of the two classes. Phasing and resultant averaging between the two classes provided an interpretable electron density. In the case of ascorbate oxidase (Messerschmidt *et al.*, 1989), multiple isomorphous replacement failed to provide an interpretable density. It was possible, however, to place the initial density into a second

crystal form, which in turn provided phases of sufficient quality to determine heavy-atom sites in derivatives of the second form. Phase-combination and density-modification techniques in the two crystal forms allowed the solution of the structure.

12.2.5.2. *Space-group problems*

Although the macromolecular crystallographer is rarely confronted with the problems facing their small-molecule colleagues with regard to determining the correct space group, the simplified heavy-atom structure may often throw some surprises. Certain pseudosymmetries may become 'exact' for the heavy-atom difference Patterson map. Thus, cross peaks between different heavy atoms may occur on a Harker section (or 'pseudo-Harker section'), complicating interpretation of the Patterson map. Such was the case with azurin (Adman *et al.*, 1978; Nar *et al.*, 1991), where the heavy-atom structure gave rise to a pseudo-homometric Patterson function, *i.e.* one in which two possible (nonequivalent) choices were available for the heavy-atom structure, only one of which was correct. This arose from a pseudo-centring of the lattice that became almost exact for the heavy-atom structure.

In the case of human NC1 (Stubbs *et al.*, 1990), all heavy-atom derivatives appeared to lie on or near the crystallographic twofold axis. This resulted in a partially centrosymmetric heavy-atom structure that failed to deliver sufficient phase information for noncentrosymmetric reflections. To check for problems with the native data set, anomalous difference Patterson maps $\{\text{coefficients } [F_{PH}(\mathbf{S}) - F_{PH}(-\mathbf{S})]^2\}$ were calculated. Coincidence of the peaks obtained from conventional and anomalous Patterson syntheses showed that the heavy-atom positions were correct, but unfortunately did not lead to a structure solution.

12.2.5.3. *High levels of substitution; noncrystallographic symmetry*

Most problematic are the cases where many heavy atoms have become incorporated in the asymmetric unit. Not only does this cause difficulties in the scaling of derivative to native data, but also the large number of peaks results in ambiguities in the solution of the Patterson function. In such cases, it may be necessary to obtain primary phase information from a different source (such as, for example, another low-substitution-site derivative). One important subclass of high-level substitution is when the native asymmetric unit contains several copies of a single molecule (noncrystallographic symmetry or NCS).

A major problem in locating complex noncrystallographic axes is that the geometrical relationship between NCS peaks in the Patterson map is nontrivial. Under certain conditions, NCS results in a recognizable local symmetry within the Patterson map (Stubbs *et al.*, 1996). In many cases, however, these conditions (that the NCS axes of crystallographic symmetry-related molecules are parallel) are not fulfilled. Under such circumstances, all heavy-atom sites (including all crystallographic symmetry-related positions) must be checked carefully with the rotation function in order to pinpoint the NCS axis. This is relatively trivial for low-order NCS (twofold, threefold), but becomes increasingly complicated for higher orders. It should also always be borne in mind that the heavy-atom positions might not necessarily follow the NCS constraints due to crystal packing. If there is reason to suspect that sites are related by local symmetry, then the orientation of this axis can be used in the initial Harker searches; in practice, however, such searches are extremely sensitive to the correct orientation of the axis.

In the case of high-order NCS (such as, *e.g.*, with icosahedral virus structures or symmetric macromolecular complexes), an alternative approach to the usual initial Harker-vector search can be provided by the self-rotation function. Knowledge of the orientation of the NCS axis (from the rotation function) can be used to determine the relative positions of heavy atoms to the NCS axis (Argos & Rossmann, 1976; Arnold *et al.*, 1987; Tong & Rossmann, 1993). The orientation can be refined and the resulting peaks can be used as input in a subsequent translation search of the Harker sections.

References

12.1

Abraham, D. J., Phillips, S. E. V. & Kennedy, P. E. (1983). *Methylphenylmercury: a novel heavy atom reagent for protein crystallography. J. Mol. Biol.* **170**, 249–252.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol.* **112**, 534–552.

Blake, C. C. F. (1968). *The preparation of isomorphous derivatives.* In *Advances in protein chemistry,* Vol. 23, pp. 59–120. New York and London: Academic Press.

Blundell, T. L. (1968). Unpublished results.

Blundell, T. L. & Jenkins, J. A. (1977). *The binding of heavy metals to proteins. Chem. Soc. Rev. (London)*, **6**, 139–171.

Blundell, T. L. & Johnson, L. N. (1976). *Protein crystallography.* New York: Academic Press.

Budisa, N., Karnbrock, W., Steinbacher, S., Humm, A., Prade, L., Neuefeind, T., Moroder, L. & Huber, R. (1997). *Bioincorporation of telluromethionine into proteins: a promising new approach for X-ray structure analysis of proteins. J. Mol. Biol.* **271**, 1–8.

Carvin, D. G. A. (1986). Unpublished results.

Carvin, D. G. A., Islam, S. A., Sternberg, M. J. E. & Blundell, T. L. (1991). *Isomorphous replacement and anomalous scattering*. Warrington: Daresbury Laboratory.

Ceska, T. A., Sayers, J. R., Stier, G. & Suck, D. (1996). *A helical arch allowing single-stranded DNA to thread through t5 5′-exonuclease. Nature (London)*, **382**, 90–93.

Chen, W. & Bahl, O. P. (1991). *Recombinant carbohydrate and selenomethionyl variants of human choriogonadotropin. J. Biol. Chem.* **266**, 8192–8197.

Doublié, S. (1997). *Preparation of selenomethionyl proteins for phase determination. Methods Enzymol.* **276**, 523–530.

Ely, K. R., Girling, R. L., Schiffer, M., Cunningham, D. E. & Edmundson, A. B. (1973). *Preparation and properties of crystals of a Bence–Jones dimer with mercury inserted into the interchain disulphide bond. Biochemistry*, **12**, 4233–4237.

Gilliland, G. L., Tung, M., Blakeslee, D. M. & Ladner, J. E. (1994). *Biological Macromolecular Crystallization Database, version 3.0: new features, data and the NASA archive for protein crystal growth data. Acta Cryst.* D**50**, 408–413.

Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *The method of isomorphous replacement for protein crystallography. Proc. R. Soc. London Ser. A*, **225**, 287–307.

Hendrickson, W. A. (1985). *Analysis of protein structure from diffraction measurement at multiple wavelengths. Trans. Am. Crystallogr. Assoc.* **21**, 11–21.