from the known portion. In such single-site search procedures, single atoms are placed at all possible positions in a crystal, using a search grid of the same fineness as for the stored Patterson function, preferably about one-third of the resolution of the Patterson map. Solutions are gauged to be acceptable if all expected vectors due to symmetry-related atoms are observed above a specified threshold in the Patterson.

Systematic computerized Patterson search procedures for orienting and positioning known molecular fragments were also developed in the early 1960s. These hierarchical procedures rely on first using the 'self'-vectors which depend only on the orientation of a molecular fragment. A search for the position of the fragment relative to the crystal symmetry elements then uses the cross-vectors between molecules (see Sections 2.3.6 and 2.3.7). Nordman constructed a weighted point representation of the predicted vector set for a fragment (Nordman & Nakatsu, 1963; Nordman, 1966) and successfully solved the structure of a number of complex alkaloids. Huber (1965) used the convolution molecule method of Hoppe (1957*a*) in three dimensions to solve a number of natural-product structures, including steroids. Various program systems have used Patterson search methods operating in real space to solve complex structures (Braun *et al*., 1969; Egert, 1983).

Others have used reciprocal-space procedures for locating known fragments. Tollin & Cochran (1964) developed a procedure for determining the orientation of planar groups by searching for origin-containing planes of high density in the Patterson. General procedures using reciprocal-space representations for determining rotation and translation parameters have been developed and will be described in Sections 2.3.6 and 2.3.7, respectively.

Although as many functions have been used to detect solutions in these Patterson search procedures as there are programs, most rely on some variation of the sum, product and minimum functions (Section 2.3.2.4). The quality of the stored Patterson density representation also varies widely, but it is now common to use 16 or more bits for single density values. Treatment of vector overlap is handled differently by different investigators and the choice will depend on the degree of overlapping (Nordman & Schilling, 1970; Nordman, 1972). General Gaussian multiplicity corrections can be employed to treat coincidental overlap of independent vectors in general positions and overlap which occurs for symmetric peaks in the vicinity of a special position or mirror plane in the Patterson (G. Kamer, S. Ramakumar & P. Argos, unpublished results; Rossmann *et al*., 1972).

2.3.3. Isomorphous replacement difference Pattersons

2.3.3.1. *Introduction*

One of the initial stages in the application of the isomorphous replacement method is the determination of heavy-atom positions. Indeed, this step of a structure determination can often be the most challenging. Not only may the number of heavy-atom sites be unknown, and have incomplete substitution, but the various isomorphous compounds may also lack isomorphism. To compound these problems, the error in the measurement of the isomorphous difference in structure amplitudes is often comparable to the differences themselves. Clearly, therefore, the ease with which a particular problem can be solved is closely correlated with the quality of the data-measuring procedure.

The isomorphous replacement method was used incidentally by Bragg in the solution of NaCl and KCl. It was later formalized by J. M. Robertson in the analysis of phthalocyanine where the coordination centre could be Pt, Ni and other metals (Robertson, 1935, 1936; Robertson & Woodward, 1937). In this and similar cases, there was no difficulty in finding the heavy-atom positions. Not only were the heavy atoms frequently in special positions, but they dominated the total scattering effect. It was not until Perutz and his colleagues (Green *et al*., 1954; Bragg & Perutz, 1954) applied the technique to the solution of haemoglobin, a protein of 68 000 Da, that it was necessary to consider methods for detecting heavy atoms. The effect of a single heavy atom, even uranium, can only have a very marginal effect on the structure amplitudes of a crystalline macromolecule. Hence, techniques had to be developed which were dependent on the difference of the isomorphous structure amplitudes rather than on the solution of the Patterson of the heavy-atom-derivative compound on its own.

2.3.3.2. *Finding heavy atoms with centrosymmetric projections*

Phases in a centrosymmetric projection will be 0 or π if the origin is chosen at the centre of symmetry. Hence, the native structure factor, $\mathbf{F}_N$, and the heavy-atom-derivative structure factor, $\mathbf{F}_{NH}$, will be collinear. It follows that the structure amplitude, $|\mathbf{F}_H|$, of the heavy atoms alone in the cell will be given by

$$|\mathbf{F}_H| = |(|\mathbf{F}_{NH}| \pm |\mathbf{F}_N|)| + \varepsilon,$$

where ε is the error on the parenthetic sum or difference. Three different cases may arise (Fig. 2.3.3.1). Since the situation shown in Fig. 2.3.3.1(*c*) is rare, in general

$$|\mathbf{F}_H|^2 \simeq (|\mathbf{F}_{NH}| - |\mathbf{F}_N|)^2. \qquad (2.3.3.1)$$

Thus, a Patterson computed with the square of the differences between the native and derivative structure amplitudes of a centrosymmetric projection will approximate to a Patterson of the heavy atoms alone.

The approximation (2.3.3.1) is valid if the heavy-atom substitution is small enough to make $|\mathbf{F}_H| \ll |\mathbf{F}_{NH}|$ for most reflections, but sufficiently large to make $\varepsilon \ll (|\mathbf{F}_{NH}| - |\mathbf{F}_N|)^2$. It is also assumed that the native and heavy-atom-derivative data have been placed on the same relative scale. Hence, the relation (2.3.3.1) should be re-written as

$$|\mathbf{F}_H|^2 \simeq (|\mathbf{F}_{NH}| - k|\mathbf{F}_N|)^2,$$

where k is an experimentally determined scale factor (see Section 2.3.3.7). Uncertainty in the determination of k will contribute further to ε, albeit in a systematic manner.

Centrosymmetric projections were used extensively for the determination of heavy-atom sites in early work on proteins such as haemoglobin (Green *et al*., 1954), myoglobin (Bluhm *et al*.,

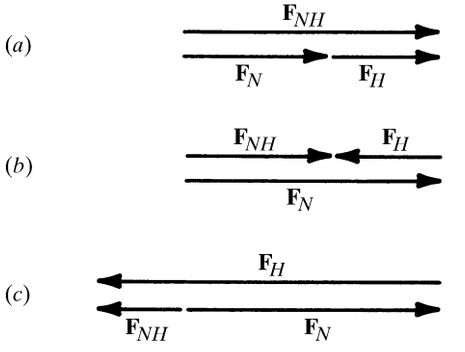


Fig. 2.3.3.1. Three different cases which can occur in the relation of the native, $\mathbf{F}_N$, and heavy-atom derivative, $\mathbf{F}_{NH}$, structure factors for centrosymmetric reflections. $\mathbf{F}_N$ is assumed to have a phase of 0, although analogous diagrams could be drawn when $\mathbf{F}_N$ has a phase of π. The crossover situation in (*c*) is clearly rare if the heavy-atom substitution is small compared to the native molecule, and can in general be neglected.

1958) and lysozyme (Poljak, 1963). However, with the advent of faster data-collecting techniques, low-resolution (*e.g.* a 5 Å limit) three-dimensional data are to be preferred for calculating difference Pattersons. For noncentrosymmetric reflections, the approximation (2.3.3.1) is still valid but less exact (Section 2.3.3.3). However, the larger number of three-dimensional differences compared to projection differences will enhance the signal of the real Patterson peaks relative to the noise. If there are N terms in the Patterson synthesis, then the peak-to-noise ratio will be proportionally $\sqrt{N}$ and $1/\varepsilon$. With the subscripts 2 and 3 representing two- and three-dimensional syntheses, respectively, the latter will be more powerful than the former whenever

$$\frac{\sqrt{N_3}}{\varepsilon_3} > \frac{\sqrt{N_2}}{\varepsilon_2}.$$

Now, as $\varepsilon_3 \simeq \sqrt{2}\varepsilon_2$, it follows that N_3 must be greater than $2N_2$ if the three-dimensional noncentrosymmetric computation is to be more powerful. This condition must almost invariably be true.

2.3.3.3. *Finding heavy atoms with three-dimensional methods*

A Patterson of a native bio-macromolecular structure (coefficients F_N^2) can be considered as being, at least approximately, a vector map of all the light atoms (carbons, nitrogens, oxygens, some sulfurs, and also phosphorus for nucleic acids) other than hydrogen atoms. These interactions will be designated as *LL*. Similarly, a Patterson of the heavy-atom derivative will contain $HH + HL + LL$ interactions, where H represents the heavy atoms. Thus, a true difference Patterson, with coefficients $F_{NH}^2 - F_N^2$, will contain only the interactions $HH + HL$. In general, the carpet of *HL* vectors completely dominates the *HH* vectors except for very small proteins such as insulin (Adams *et al.*, 1969). Therefore, it would be preferable to compute a Patterson containing only *HH* interactions in order to interpret the map in terms of specific heavy-atom sites.

Blow (1958) and Rossmann (1960) showed that a Patterson with $(|\mathbf{F}_{NH}| - |\mathbf{F}_N|)^2$ coefficients approximated to a Patterson containing only *HH* vectors. If the phase angle between $\mathbf{F}_N$ and $\mathbf{F}_{NH}$ is φ (Fig. 2.3.3.2), then

$$|\mathbf{F}_H|^2 = |\mathbf{F}_N|^2 + |\mathbf{F}_{NH}|^2 - 2|\mathbf{F}_N||\mathbf{F}_{NH}|\cos\varphi.$$

In general, however, $|\mathbf{F}_H| \ll |\mathbf{F}_N|$. Hence, φ is small and

$$|\mathbf{F}_H|^2 \simeq (|\mathbf{F}_{NH}| - |\mathbf{F}_N|)^2,$$

which is the same relation as (2.3.3.1) for centrosymmetric approximations. Since the direction of $\mathbf{F}_H$ is random compared to $\mathbf{F}_N$, the root-mean-square projected length of $\mathbf{F}_H$ onto $\mathbf{F}_N$ will be $\mathbf{F}_H/\sqrt{2}$. Thus it follows that a better approximation is

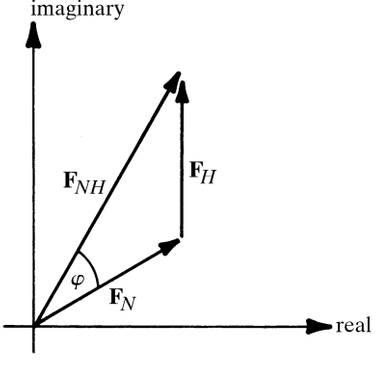


Fig. 2.3.3.2. Vector triangle showing the relationship between $\mathbf{F}_N$, $\mathbf{F}_{NH}$ and $\mathbf{F}_H$, where $\mathbf{F}_{NH} = \mathbf{F}_N + \mathbf{F}_H$.

$$|\mathbf{F}_H|^2 \simeq \sqrt{2}(|\mathbf{F}_{NH}| - |\mathbf{F}_N|)^2, \qquad (2.3.3.2)$$

which accounts for the assumption (Section 2.3.3.2) that $\varepsilon_3 = \sqrt{2}\varepsilon_2$. The almost universal method for the initial determination of major heavy-atom sites in an isomorphous derivative utilizes a Patterson with $(|\mathbf{F}_{NH}| - |\mathbf{F}_N|)^2$ coefficients. Approximation (2.3.3.2) is also the basis for the refinement of heavy-atom parameters in a single isomorphous replacement pair (Rossmann, 1960; Cullis *et al.*, 1962; Terwilliger & Eisenberg, 1983).

2.3.3.4. *Correlation functions*

In the most general case of a triclinic space group, it will be necessary to select an origin arbitrarily, usually coincident with a heavy atom. All other heavy atoms (and subsequently also the macromolecular atoms) will be referred to this reference atom. However, the choice of an origin will be independent in the interpretation of each derivative's difference Patterson. It will then be necessary to correlate the various, arbitrarily chosen, origins. The same problem occurs in space groups lacking symmetry axes perpendicular to the primary rotation axis (*e.g.* $P2_1$, $P6$ *etc.*), although only one coordinate, namely parallel to the unique rotation axis, will require correlation. This problem gave rise to some concern in the 1950s. Bragg (1958), Blow (1958), Perutz (1956), Hoppe (1959) and Bodo *et al.* (1959) developed a variety of techniques, none of which were entirely satisfactory. Rossmann (1960) proposed the $(\mathbf{F}_{NH1} - \mathbf{F}_{NH2})^2$ synthesis and applied it successfully to the heavy-atom determination of horse haemoglobin. This function gives positive peaks $(H1 \cdot H1)$ at the end of vectors between the heavy-atom sites in the first compound, positive peaks $(H2 \cdot H2)$ between the sites in the second compound, and negative peaks between sites in the first and second compound (Fig. 2.3.3.3). It is thus the negative peaks which provide the necessary correlation. The function is unique in that it is a Patterson containing significant information in both positive and negative peaks. Steinrauf (1963) suggested using the coefficients $(|\mathbf{F}_{NH1}| - |\mathbf{F}_N|) \cdot (|\mathbf{F}_{NH2}| - |\mathbf{F}_N|)$ in order to eliminate the positive $H1 \cdot H1$ and $H2 \cdot H2$ vectors.

Although the problem of correlation was a serious concern in the early structural determination of proteins during the late 1950s and early 1960s, the problem has now been by-passed. Blow &

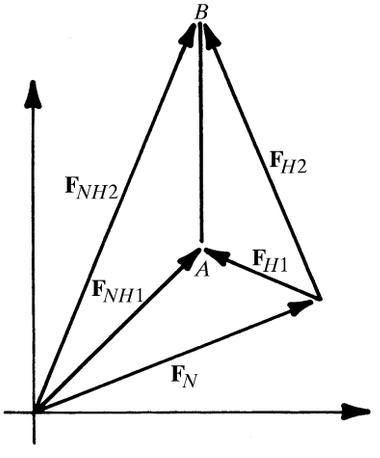


Fig. 2.3.3.3. A Patterson with coefficients $(\mathbf{F}_{NH1} - \mathbf{F}_{NH2})^2$ will be equivalent to a Patterson whose coefficients are $(AB)^2$. However, $AB = -\mathbf{F}_{H1} + \mathbf{F}_{H2}$. Thus, a Patterson with $(AB)^2$ coefficients is equivalent to having negative atomic substitutions in compound 1 and positive substitutions in compound 2, or *vice versa*. Therefore, the Patterson will contain positive peaks for vectors of the type $H1 \cdot H1$ and $H2 \cdot H2$, but negative vector peaks for vectors of type $H1 \cdot H2$.

Rossmann (1961) and Kartha (1961) independently showed that it was possible to compute usable phases from a single isomorphous replacement (SIR) derivative. This contradicted the previously accepted notion that it was necessary to have at least two isomorphous derivatives to be able to determine a noncentrosymmetric reflection's phase (Harker, 1956). Hence, currently, the procedure used to correlate origins in different derivatives is to compute SIR phases from the first compound and apply them to a difference electron-density map of the second heavy-atom derivative. Thus, the origin of the second derivative will be referred to the arbitrarily chosen origin of the first compound. More important, however, the interpretation of such a 'feedback' difference Fourier is easier than that of a difference Patterson. Hence, once one heavy-atom derivative has been solved for its heavy-atom sites, the solution of other derivatives is almost assured. This concept is examined more closely in the following section.

2.3.3.5. *Interpretation of isomorphous difference Pattersons*

Difference Pattersons have usually been manually interpreted in terms of point atoms. In more complex situations, such as crystalline viruses, a systematic approach may be necessary to analyse the Patterson. That is especially true when the structure contains noncrystallographic symmetry (Argos & Rossmann, 1976). Such methods are in principle dependent on the comparison of the observed Patterson, $P_1(\mathbf{x})$, with a calculated Patterson, $P_2(\mathbf{x})$. A criterion, C_P, based on the sum of the Patterson densities at all test vectors within the unit-cell volume V, would be

$$C_P = \int_V P_1(\mathbf{x}) \cdot P_2(\mathbf{x})\, \mathrm{d}\mathbf{x}.$$

C_P can be evaluated for all reasonable heavy-atom distributions. Each different set of trial sites corresponds to a different P_2 Patterson. It is then easily shown that

$$C_P = \sum_{\mathbf{h}} \Delta_{\mathbf{h}}^2 E_{\mathbf{h}}^2,$$

where the sum is taken over all $\mathbf{h}$ reflections in reciprocal space, $\Delta_{\mathbf{h}}^2$ are the observed differences and $E_{\mathbf{h}}$ are the structure factors of the trial point Patterson. (The symbol E is used here because of its close relation to normalized structure factors.)

Let there be n noncrystallographic asymmetric units within the crystallographic asymmetric unit and m crystallographic asymmetric units within the crystal unit cell. Then there are L symmetry-related heavy-atom sites where $L = nm$. Let the scattering contribution of the ith site have a_i and b_i real and imaginary structure-factor components with respect to an arbitrary origin. Hence, for reflection $\mathbf{h}$

$$E_{\mathbf{h}}^2 \left(\sum_L a_{\mathbf{h}i}\right)^2 + \left(\sum_L b_{\mathbf{h}i}\right)^2 = L + \sum_{i\neq j}^{N}\sum^{N} (a_{\mathbf{h}i}a_{\mathbf{h}j} + b_{\mathbf{h}i}b_{\mathbf{h}j}).$$

Therefore,

$$C_P = \sum_{\mathbf{h}} \Delta_{\mathbf{h}}^2 \left[L + 2 \sum_{i\neq j}\sum (a_{\mathbf{h}i}a_{\mathbf{h}j} + b_{\mathbf{h}i}b_{\mathbf{h}j}) \right].$$

But $\sum_{\mathbf{h}} \Delta_{\mathbf{h}}^2$ must be independent of the number, L, of heavy-atom sites per cell. Thus the criterion can be re-written as

$$C'_P = \sum_{\mathbf{h}} \Delta_{\mathbf{h}}^2 \left[\sum_{i\neq j}\sum (a_{\mathbf{h}i}a_{\mathbf{h}j} + b_{\mathbf{h}i}b_{\mathbf{h}j}) \right]. \quad (2.3.3.3)$$

More generally, if some sites have already been tentatively determined, and if these sites give rise to the structure-factor components $A_{\mathbf{h}}$ and $B_{\mathbf{h}}$, then

$$E_{\mathbf{h}}^2 = \left(A_{\mathbf{h}} + \sum_N a_{\mathbf{h}i}\right)^2 + \left(B_{\mathbf{h}} + \sum_N b_{\mathbf{h}i}\right)^2. \quad (2.3.3.4)$$

Following the same procedure as above, it follows that

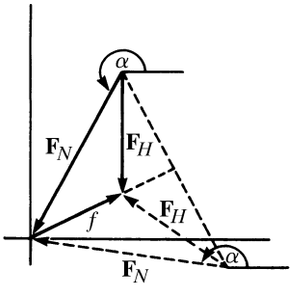
$$C'_P = \sum_{\mathbf{h}} \Delta_{\mathbf{h}}^2 \left[(A_{\mathbf{h}}a_{\mathbf{h}} + B_{\mathbf{h}}b_{\mathbf{h}}) + \sum_{i\neq j}\sum (a_{\mathbf{h}i}a_{\mathbf{h}j} + b_{\mathbf{h}i}b_{\mathbf{h}j}) \right], \quad (2.3.3.5)$$

where $a_{\mathbf{h}} = \sum_{i=1}^{L} a_{\mathbf{h}i}$ and $b_{\mathbf{h}} = \sum_{i=1}^{L} b_{\mathbf{h}i}$.

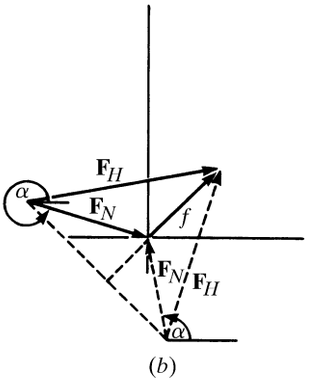
Expression (2.3.3.5) will now be compared with the 'feedback' method (Dickerson *et al*., 1967, 1968) of verifying heavy-atom sites using SIR phasing. Inspection of Fig. 2.3.3.4 shows that the native phase, α, will be determined as $\alpha = \varphi + \pi$ (φ is the structure-factor phase corresponding to the presumed heavy-atom positions) when $|\mathbf{F}_N| > |\mathbf{F}_H|$ and $\alpha = \varphi$ when $|\mathbf{F}_N| \ll |\mathbf{F}_H|$. Thus, an SIR difference electron density, $\Delta\rho(\mathbf{x})$, can be synthesized by the Fourier summation

$$\begin{aligned}\Delta\rho(\mathbf{x}) &= \frac{1}{V}\sum m(|\mathbf{F}_{NH}| - |\mathbf{F}_N|)\cos(2\pi\mathbf{h}\cdot\mathbf{x} - \varphi_{\mathbf{h}}) \\ &\quad \text{from terms with } \Delta_{\mathbf{h}} = |\mathbf{F}_{NH}| - |\mathbf{F}_N| > 0 \\ &\quad + \frac{1}{V}\sum m(|\mathbf{F}_{NH}| - |\mathbf{F}_N|)\cos(2\pi\mathbf{h}\cdot\mathbf{x} - \varphi_{\mathbf{h}} - \pi) \\ &\quad \text{from terms with } \Delta_{\mathbf{h}} < 0 \\ &= \frac{1}{V}\sum m|\Delta_{\mathbf{h}}|\cos(2\pi\mathbf{h}\cdot\mathbf{x} - \varphi_{\mathbf{h}}),\end{aligned}$$

where m is a figure of merit of the phase reliability (Blow & Crick, 1959; Dickerson *et al*., 1961). Now,



(*a*)



(*b*)

Fig. 2.3.3.4. The phase α of the native compound (structure factor $\mathbf{F}_N$) is determined either as being equal to, or 180° out of phase with, the presumed heavy-atom contribution when only a single isomorphous compound is available. In (*a*) is shown the case when $|\mathbf{F}_N| > |\mathbf{F}_{NH}|$ and $\alpha \simeq \varphi + \pi$. In (*b*) is shown the case when $|\mathbf{F}_N| < |\mathbf{F}_{NH}|$ and $\alpha = \varphi$, where φ is the phase of the heavy-atom structure factor $\mathbf{F}_H$.

$$\mathbf{F}_\mathbf{h} = A_\mathbf{h} + iB_\mathbf{h} = F_H \cos\varphi_\mathbf{h} + iF_H \sin\varphi_\mathbf{h},$$

where $A_\mathbf{h}$ and $B_\mathbf{h}$ are the real and imaginary components of the presumed heavy-atom sites. Therefore,

$$\Delta\rho(\mathbf{x}) = \frac{1}{V}\sum \frac{m|\Delta_\mathbf{h}|}{|\mathbf{F}_H|}(A_\mathbf{h}\cos 2\pi\mathbf{h}\cdot\mathbf{x} + B_\mathbf{h}\sin 2\pi\mathbf{h}\cdot\mathbf{x}).$$

If this SIR difference electron-density map shows significant peaks at sites related by noncrystallographic symmetry, then those sites will be at the position of a further set of heavy atoms. Hence, a suitable criterion for finding heavy-atom sites is

$$C_{\rm SIR} = \sum_{j=1}^{n}\Delta\rho(\mathbf{x}_j),$$

or by substitution

$$C_{\rm SIR} = \sum_{j=1}^{n}\frac{1}{V}\sum_\mathbf{h}\frac{m|\Delta_\mathbf{h}|}{|\mathbf{F}_H|}(A_\mathbf{h}\cos 2\pi\mathbf{h}\cdot\mathbf{x}_j + B_\mathbf{h}\sin 2\pi\mathbf{h}\cdot\mathbf{x}_j).$$

But

$$a_\mathbf{h} = \sum_{j=1}^{n}\cos 2\pi\mathbf{h}\cdot\mathbf{x}_j \quad \text{and} \quad b_\mathbf{h} = \sum_{j=1}^{n}\sin 2\pi\mathbf{h}\cdot\mathbf{x}_j.$$

Therefore,

$$C_{\rm SIR} = \frac{1}{V}\sum_\mathbf{h}\frac{m|\Delta_\mathbf{h}|}{|\mathbf{F}_H|}(A_\mathbf{h}a_\mathbf{h} + B_\mathbf{h}b_\mathbf{h}). \qquad (2.3.3.6)$$

This expression is similar to (2.3.3.5) derived by consideration of a Patterson search. It differs from (2.3.3.5) in two respects: the Fourier coefficients are different and expression (2.3.3.6) is lacking a second term. Now the figure of merit m will be small whenever $|\mathbf{F}_H|$ is small as the SIR phase cannot be determined well under those conditions. Hence, effectively, the coefficients are a function of $|\Delta_\mathbf{h}|$, and the coefficients of the functions (2.3.3.5) and (2.3.3.6) are indeed rather similar. The second term in (2.3.3.5) relates to the use of the search atoms in phasing and could be included in (2.3.3.6), provided the actual feedback sites in each of the n electron-density functions tested by $C_{\rm SIR}$ are omitted in turn. Thus, a systematic Patterson search and an SIR difference Fourier search are very similar in character and power.

2.3.3.6. *Direct structure determination from difference Pattersons*

The difference Patterson computed with coefficients $F_{HN}^2 - F_N^2$ contains information on the heavy atoms (*HH* vectors) and the macromolecular structure (*HL* vectors) (Section 2.3.3.3). If the scaling between the $|\mathbf{F}_{HN}|$ and $|\mathbf{F}_N|$ data sets is not perfect there will also be noise. Rossmann (1961*b*) was partially successful in determining the low-resolution horse haemoglobin structure by using a series of superpositions based on the known heavy-atom sites. Nevertheless, Patterson superposition methods have not been used for the structure determination of proteins owing to the successful error treatment of the isomorphous replacement method in reciprocal space. However, it is of some interest here for it gives an alternative insight into SIR phasing.

The deconvolution of an arbitrary molecule, represented as '?', from an $(F_{HN}^2 - F_N^2)$ Patterson, is demonstrated in Fig. 2.3.3.5. The original structure is shown in Fig. 2.3.3.5(*a*) and the corresponding Patterson in Fig. 2.3.3.5(*b*). Superposition with respect to one of the heavy-atom sites is shown in Fig. 2.3.3.5(*c*) and the other in Fig. 2.3.3.5(*d*). Both Figs. 2.3.3.5(*c*) and (*d*) contain a centre of symmetry because the use of only a single *HH* vector implies a centre of symmetry half way between the two sites. The centre is broken on combining information from all three sites (which together lack a centre of symmetry) by superimposing Figs. 2.3.3.5(*c*) and (*d*) to obtain either the original structure (Fig. 2.3.3.5*a*) or its enantiomorph. Thus it is clear, in principle, that there is sufficient information in a single isomorphous derivative data set, when used in conjunction with a native data set, to solve a structure completely. However, the procedure shown in Fig. 2.3.3.5 does not consider the accumulation of error in the selection of individual images when these intersect with another image. In this sense the reciprocal-space isomorphous replacement technique has greater elegance and provides more insight, whereas the alternative view given by the Patterson method was the original stimulus for the discovery of the SIR phasing technique (Blow & Rossmann, 1961).

Other Patterson functions for the deconvolution of SIR data have been proposed by Ramachandran & Raman (1959), as well as others. The principles are similar but the coefficients of the functions are optimized to emphasize various aspects of the signal representing the molecular structure.

2.3.3.7. *Isomorphism and size of the heavy-atom substitution*

It is insufficient to discuss Patterson techniques for locating heavy-atom substitutions without also considering errors of all kinds. First, it must be recognized that most heavy-atom labels are not a single atom but a small compound containing one or more heavy atoms. The compound itself will displace water or ions and locally alter the conformation of the protein or nucleic acid. Hence, a simple Gaussian approximation will suffice to represent individual heavy-atom scatterers responsible for the difference between native and heavy-atom derivatives. Furthermore, the heavy-atom compound often introduces small global structural changes which can be detected only at higher resolution. These problems were considered with some rigour by Crick & Magdoff (1956). In general, lack of isomorphism is exhibited by an increase in the size of the isomorphous differences with increasing resolution (Fig. 2.3.3.6).

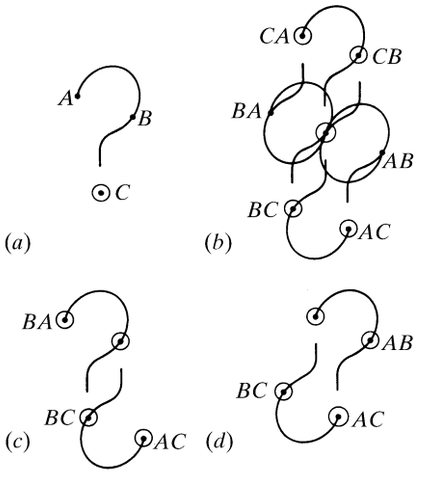


Fig. 2.3.3.5. Let (*a*) be the original structure which contains three heavy atoms *ABC* in a noncentrosymmetric configuration. Then a Fourier summation, with $(F_{NH}^2 - F_N^2)$ coefficients, will give the Patterson shown in (*b*). Displacement of the Patterson by the vector *BC* and selecting the common patterns yields (*c*). Similarly, displacement by *AC* gives (*d*). Finally, superposition of (*c*) on (*d*) gives the original figure or its enantiomorph. This series of steps demonstrates that, in principle, complete structural information is contained in an SIR derivative.

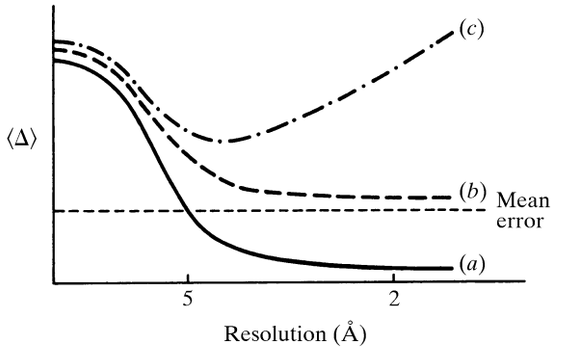


Fig. 2.3.3.6. A plot of mean isomorphous differences as a function of resolution. (*a*) The theoretical size of mean differences following roughly a Gaussian distribution. (*b*) The observed size of differences for a good isomorphous derivative where the smaller higher-order differences have been largely masked by the error of measurement. (*c*) Observed differences where 'lack of isomorphism' dominates beyond approximately 5 Å resolution.

Crick & Magdoff (1956) also derived the approximate expression

$$\sqrt{\frac{2N_H}{N_P}}\cdot\frac{f_H}{f_P}$$

to estimate the r.m.s. fractional change in intensity as a function of heavy-atom substitution. Here, N_H represents the number of heavy atoms attached to a protein (or other large molecule) which contains N_P light atoms. f_H and f_P are the scattering powers of the average heavy and protein atom, respectively. This function was tabulated by Eisenberg (1970) as a function of molecular weight (proportional to N_P). For instance, for a single, fully substituted, Hg atom the formula predicts an r.m.s. intensity change of around 25% in a molecule of 100 000 Da. However, the error of measurement of a reflection intensity is likely to be arround 10% of I, implying perhaps an error of around 14% of I on a difference measurement. Thus, the isomorphous replacement difference measurement for almost half the reflections will be buried in error for this case.

Scaling of the different heavy-atom-derivative data sets onto a common relative scale is clearly important if error is to be reduced. Blundell & Johnson (1976, pp. 333–336) give a careful discussion of this subject. Suffice it to say here only that a linear scale factor is seldom acceptable as the heavy-atom-derivative crystals frequently suffer from greater disorder than the native crystals. The heavy-atom derivative should, in general, have a slightly larger mean value for the structure factors on account of the additional heavy atoms (Green *et al.*, 1954). The usual effect is to make $\sum|\mathbf{F}_{NH}|^2/\sum|\mathbf{F}_N|^2 \simeq 1.05$ (Phillips, 1966).

As the amount of heavy atom is usually unknown in a yet unsolved heavy-atom derivative, it is usual practice either to apply a scale factor of the form $k\exp[-B(\sin\theta/\lambda)^2]$ or, more generally, to use local scaling (Matthews & Czerwinski, 1975). The latter has the advantage of not making any assumption about the physical nature of the relative intensity decay with resolution.

2.3.4. Anomalous dispersion

2.3.4.1. *Introduction*

The physical basis for anomalous dispersion has been well reviewed by Ramaseshan & Abrahams (1975), James (1965), Cromer (1974) and Bijvoet (1954). As the wavelength of radiation approaches the absorption edge of a particular element, then an atom will disperse X-rays in a manner that can be defined by the complex scattering factor

$$f_0 + \Delta f' + i\Delta f'',$$

where f_0 is the scattering factor of the atom without the anomalous absorption and re-scattering effect, $\Delta f'$ is the real correction term (usually negative), and $\Delta f''$ is the imaginary component. The real term $f_0 + \Delta f'$ is often written as f', so that the total scattering factor will be $f' + if''$. Values of $\Delta f'$ and $\Delta f''$ are tabulated in *IT* IV (Cromer, 1974), although their precise values are dependent on the environment of the anomalous scatterer. Unlike f_0, $\Delta f'$ and $\Delta f''$ are almost independent of scattering angle as they are caused by absorption of energy in the innermost electron shells. Thus, the anomalous effect resembles scattering from a point atom.

The structure factor of index $\mathbf{h}$ can now be written as

$$\mathbf{F}_\mathbf{h} = \sum_{j=1}^{N} f_j' \exp(2\pi i\mathbf{h}\cdot\mathbf{x}_j) + i\sum_{j=1}^{N} f_j'' \exp(2\pi i\mathbf{h}\cdot\mathbf{x}_j). \qquad (2.3.4.1)$$

(Note that the only significant contributions to the second term are from those atoms that have a measurable anomalous effect at the chosen wavelength.)

Let us now write the first term as $A + iB$ and the second as $a + ib$. Then, from (2.3.4.1),

$$\mathbf{F} = (A + iB) + i(a + ib) = (A - b) + i(B + a). \qquad (2.3.4.2)$$

Therefore,

$$|\mathbf{F}_\mathbf{h}|^2 = (A - b)^2 + (B + a)^2$$

and similarly

$$|\mathbf{F}_{\bar{\mathbf{h}}}|^2 = (A + b)^2 + (-B + a)^2,$$

demonstrating that Friedel's law breaks down in the presence of anomalous dispersion. However, it is only for noncentrosymmetric reflections that $|\mathbf{F}_\mathbf{h}| \neq |\mathbf{F}_{\bar{\mathbf{h}}}|$.

Now,

$$\rho(\mathbf{x}) = \frac{1}{V}\sum_{\mathbf{h}}^{\text{sphere}} \mathbf{F}_\mathbf{h}\exp(2\pi i\mathbf{h}\cdot\mathbf{x}).$$

Hence, by using (2.3.4.2) and simplifying,

$$\rho(\mathbf{x}) = \frac{2}{V}\sum_{\mathbf{h}}^{\text{hemisphere}} [(A\cos 2\pi\mathbf{h}\cdot\mathbf{x} - B\sin 2\pi\mathbf{h}\cdot\mathbf{x}) + i(a\cos 2\pi\mathbf{h}\cdot\mathbf{x} - b\sin 2\pi\mathbf{h}\cdot\mathbf{x})]. \qquad (2.3.4.3)$$

The first term in (2.3.4.3) is the usual real Fourier expression for electron density, while the second term is an imaginary component due to the anomalous scattering of a few atoms in the cell.

2.3.4.2. *The $P_s(\mathbf{u})$ function*

Expression (2.3.4.3) gives the complex electron density expression in the presence of anomalous scatterers. A variety of Patterson-type functions can be derived from (2.3.4.3) for the determination of a structure. For instance, the $P_s(\mathbf{u})$ function gives vectors between the anomalous atoms and the 'normal' atoms.

From (2.3.4.1) it is easy to show that

$$\begin{aligned}\mathbf{F}_\mathbf{h}\mathbf{F}_\mathbf{h}^* &= |\mathbf{F}_\mathbf{h}|^2 \\ &= \sum_{i,j}(f_i'f_j' + f_i''f_j'')\cos 2\pi\mathbf{h}\cdot(\mathbf{x}_i - \mathbf{x}_j) \\ &\quad + \sum_{i,j}(f_i'f_j'' - f_i''f_j')\sin 2\pi\mathbf{h}\cdot(\mathbf{x}_i - \mathbf{x}_j).\end{aligned}$$

Therefore,

references