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 (1957a) 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 further to

determining the orientation of planar groups by searching for

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 \( \pi \) if the origin is chosen at the centre of symmetry. Hence, the native structure factor, \( F_N \), and the heavy-atom-derivative structure factor, \( F_{NH} \), will be collinear. It follows that the structure amplitude, \( |F_H| \), of the heavy atoms alone in the cell will be given by

\[
|F_H| \approx |(F_{NH} \pm F_N)| = |F_N| + \varepsilon,
\]

where \( \varepsilon \) 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

\[
|F_H|^2 \approx (|F_{NH}| - |F_N|)^2.
\]

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 \( |F_H| \ll |F_{NH}| \) for most reflections, but sufficiently large to make \( \varepsilon \ll (|F_{NH}| - |F_N|)^2 \). 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

\[
|F_H|^2 \approx (|F_{NH}| - k|F_N|)^2,
\]

where \( k \) is an experimentally determined scale factor (see Section 2.3.3.7). Uncertainty in the determination of \( k \) will contribute further to \( \varepsilon \), 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., 1954, 1956, 1969; Egert, 1983).
2.3. PATTERSON AND MOLECULAR-REPLACEMENT TECHNIQUES

In order to interpret the map in terms of specific heavy-atom sites, such as insulin (Adams et al., 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

\[
\frac{\sqrt{N_2}}{\varepsilon_3} > \frac{\sqrt{N_3}}{\varepsilon_2}.
\]

Now, as \( \varepsilon_3 \approx \sqrt{2}\varepsilon_2 \), 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 \( F_h^3 \)) 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 \( L \). 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 \( F_{NH} - F_h^3 \), will contain only the interactions \( HH \). 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 \( (|F_{NH} - |F_N|)|^2 \) coefficients approximates to a Patterson containing only \( HH \) vectors. If the phase angle between \( F_N \) and \( F_{NH} \) is \( \varphi \) (Fig. 2.3.3.2), then

\[
|F_H|^2 = |F_N|^2 + |F_{NH}|^2 - 2|F_N||F_{NH}|\cos \varphi.
\]

In general, however, \( |F_H| \ll |F_N| \). Hence, \( \varphi \) is small and

\[
|F_H|^2 \approx (|F_{NH} - |F_N|)|^2,
\]

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

\[
|F_H|^2 \approx \sqrt{2}(|F_{NH} - |F_N|)|^2,
\]

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 \( (|F_{NH} - |F_N|)|^2 \) 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 \( (|F_{NH} - |F_N|)|^2 \) synthesis and applied it successfully to the heavy-atom determination of horse haemoglobin. This function gives positive peaks \((H \cdot H1)\) 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 \( (|F_{NH1} - |F_N|)| \cdot (|F_{NH2} - |F_N|)| \) 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 &