Fourier Techniques in X-Ray Structure Determination

7

7.1 Introduction

We have reached the stage where we can consider how to attack the solving of a crystal structure. After the earliest trial and error determinations in the 1920s with very simple and highly symmetrical structures, it was found that the application of Fourier series, initially in one dimension, led to the electron density function, in which peak maxima in the electron density corresponded to atomic positions. As we have seen in the previous chapters, it is necessary to have the phases of the structure factors for a Fourier synthesis to be carried out meaningfully. One way in which phase information may be obtained is through the Patterson function of vector density, a function of interatomic vectors in the crystal structure.

In this chapter, we examine the application of Fourier series to crystal structure analysis, together with some of its developments. However, in any structure analysis there are certain preliminary investigations that can throw light on the problem in hand. One such investigation, which we have already discussed, leads to the unit-cell dimensions and space group of the crystal. We study next some example crystal structures in order to show how other, readily available information can be used to assist in the structure solution process.

7.2 Analysis of the Unit-Cell Contents

The density D_m of the crystals under examination may be measured by suspending them in a liquid or liquid mixture. The composition of the liquid is altered until the crystals neither rise nor fall; then the density of the liquid, equal to D_m, is measured with a pyknometer. Many organic materials can be suspended in aqueous sodium bromide. At flotation equilibrium, the refractive index of the solution may be measured, and the density determined by interpolation of a graph of the density of the solution as a function of its refractive index. The flotation procedure is best carried out in a thermostat. It may still happen, however, that the demarcation between sinking and floating is a little ill defined. Inclusion of air or solvent in the crystal will lead to a smaller apparent density, and the flotation position corresponding to a maximum value for the density measured should be most appropriate.

If the crystal unit cell contains a number Z of chemical species, each of relative molar mass M_r, then the following relationship holds:

$$D_m = ZM_r m_u / V_c \tag{7.1}$$

M. Ladd and R. Palmer, *Structure Determination by X-ray Crystallography: Analysis by X-rays and Neutrons*, DOI 10.1007/978-1-4614-3954-7_7,

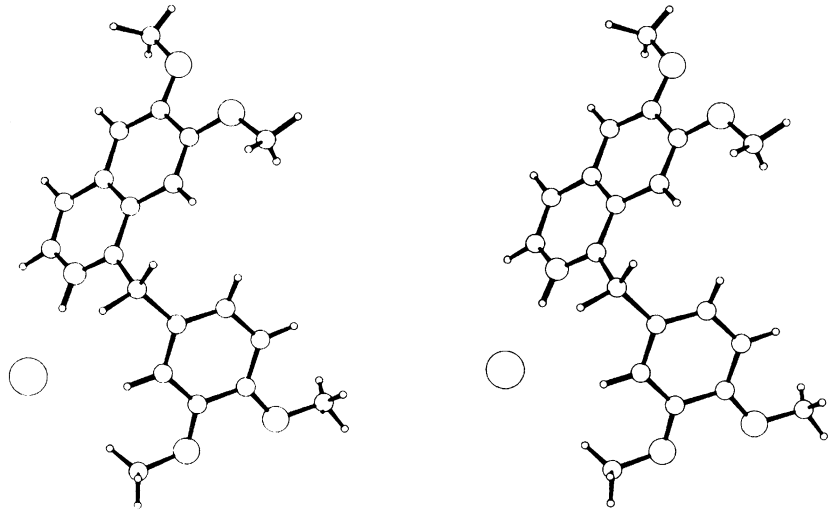


Fig. 7.1 Stereoview of the molecular conformation of papaverine hydrochloride; the *circles*, in order of decreasing size, represent Cl, O, N, C, and H

where m_u is the atomic mass unit. If the volume V_c of the unit cell is in Å^3 and the density in g cm^{-3}, (7.1) may be written as

$$D_m = 1.6605\, ZM_r/V_c \tag{7.2}$$

If several measurements are made, the standard deviation $\sigma(D_m)$ can be deduced. It is useful to calculate the density D_c from the unit-cell volume and the (integral) value of Z. A significant discrepancy between D_m and D_c should be examined, as it might point to an error in the unit-cell dimensions or to solvent of crystallization not included in M_r at that stage.

7.2.1 Papaverine Hydrochloride, $C_{20}H_{21}NO_4$·HCl

Crystal Data
System: monoclinic
Unit-cell dimensions: $a = 13.059$ Å, $b = 15.620$ Å, $c = 9.130$ Å, $\beta = 92.13°$
V_c: 1861.1 Å^3
D_m: 1.33 g cm^{-3}
M_r: 375.85
Z: 4 to the nearest integer (3.97 from (7.2))
Unit-cell contents: 80C, 88H, 4N, 16O, 4C1 atoms
Absent spectra: $h0l$: l odd; $0k0$: k odd
Space group: $P2_1/c$

All atoms are in general equivalent positions. The molecular conformation, obtained by a complete structure analysis [1], is shown in Fig. 7.1.

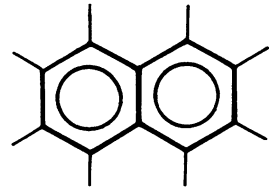


Fig. 7.2 Naphthalene molecular structure; the 9,10 carbon–carbon bond lies on a center of symmetry

7.2.2 Naphthalene, $C_{10}H_8$

Crystal Data
System: monoclinic
Unit-cell dimensions: $a = 8.658$ Å, $b = 6.003$ Å, $c = 8.235$ Å, $\beta = 122.92°$
V_c: 359.28 Å^3
D_m: 1.152 g cm^{-3}
M_r: 128.17
Z: 2 to the nearest integer (1.94 from (7.2))
Unit-cell contents: 20C, 16H atoms
Absent spectra: $h0l$: $l = 2n + 1$; $0k0$: $l = 2n + 1$
Space group: $P2_1/c$

7.2.3 Molecular Symmetry

In papaverine hydrochloride, the four molecules in the unit cell occupy a set of general positions; each atom at coordinates x_j, y_j, z_j, $(j = 1, 2, \ldots, 48)$ is repeated by the space-group symmetry so as to build up the crystal structure. There are, therefore, 48 atoms, including hydrogen, in the asymmetric unit to be located by the structure analysis.

Naphthalene is not quite so straightforward. With two molecules per unit cell, there are 20 carbon atoms and 16 hydrogen atoms that may be distributed in 4 equivalent-position sets of five and four atoms, respectively, in the unit cell. This means that in order to solve the structure, we have to locate five carbon atoms and four hydrogen atoms. This number is only half that expected: since Z is 2, each atom is related by one of the symmetry elements of the space group to a second atom of the same type in the same molecule, so as to generate $C_{10}H_8$ from C_5H_4. There are three different symmetry elements to consider: the 2_1 axis, the c-glide plane, and the center of symmetry. The screw axis and glide plane are discounted because they involve translational symmetry, which would generate an infinite molecule with translational repeats. We must, therefore, conclude that the atom pairs are related by a center of symmetry, which in turn means that the molecule of naphthalene is centrosymmetric.

The symmetry analysis for naphthalene has served two very useful purposes: it has halved the work of the subsequent structure analysis, and shown that the molecules in the crystal exhibit a certain minimum symmetry ($\bar{1}$). This result is, of course, in agreement with chemical knowledge, which ordinarily we are quite entitled to use. The conventional notion that naphthalene should have *mmm* symmetry, Fig. 7.2, is not supported directly, although the crystal structure analysis shows that this symmetry holds within experimental error.

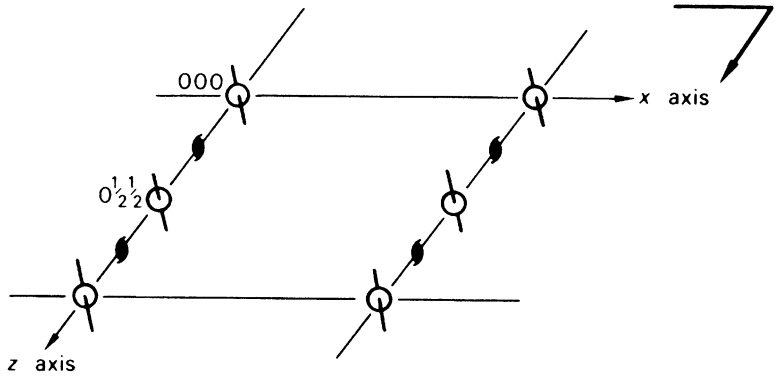


Fig. 7.3 Grouping of one of the special positions sets in $P2_1/c$; the arrangement of molecules (symmetry $\bar{1}$) with their centers at 0, 0, 0 and 0, $\frac{1}{2}$, $\frac{1}{2}$ is shown

7.2.4 Special Positions

The molecules of naphthalene lie on special positions in $P2_1/c$, Fig. 7.3. Special position sites correspond in symmetry to one of the 32 crystallographic point groups and, in subsequent examples, we shall see that both atoms and molecules can occupy special positions.

Glide planes and screw axes do not usually accommodate atoms or molecules; an atom lying exactly on a translational symmetry element would introduce a pseudo-half-axial translation, thus creating special reflection conditions which, depending on the atomic number, may be observable among the X-ray data (see Problem 7.1).

Although the molecules of naphthalene are in special positions, they are subject to the space-group symmetry inherent in the general positions: if one molecule is located at 0, 0, 0, then the second molecule is at 0, $\frac{1}{2}$, $\frac{1}{2}$; see Sect. 2.7.5. This set may be determined by substituting $x = y = z = 0$ into the set of general positions. The structure of naphthalene [2] is shown in Fig. 7.4. The reader may like to consider the three other possible sets of special positions that could be used to represent this structure, and then show from the structure factor equation that $|F(hkl)|$ is invariant with respect to each set of special positions.

7.2.5 Nickel Tungstate, $NiWO_4$

Crystal Data
System: monoclinic
Unit-cell dimensions: $a = 4.60$ Å, $b = 5.66$ Å, $c = 4.91$ Å, $\beta = 90.1°$
V_c: 127.84 Å^3
D_m: 7.964 g cm^{-3}
M_r: 306.81
Z: 2 to the nearest integer (2.00 from (7.2))
Unit-cell contents: 2Ni, 2W, 8O atoms
Absent spectra: $h0l$: $l = 2n + 1$
Possible space groups: Pc or $P2/c$

We shall use space group $P2/c$, since the structure was determined successfully only with this space group [3].

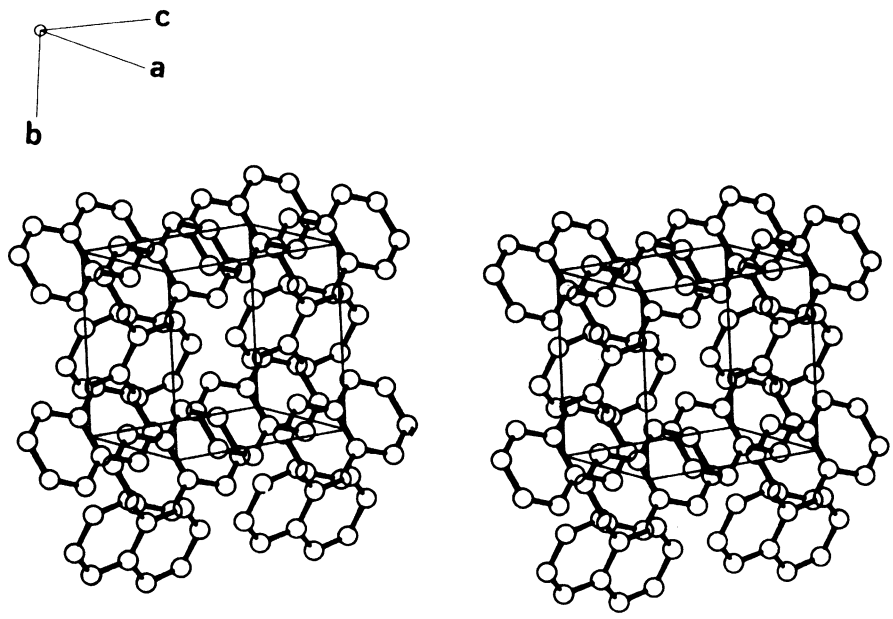


Fig. 7.4 Stereoview of the crystal structure of naphthalene; for clarity, H atoms are not shown

The general equivalent positions in $P2/c$ are

$$\pm \{x, y, z; \quad x, \bar{y}, \tfrac{1}{2} + z\}$$

but in order to study $NiWO_4$ further, we must consider the possible special positions for this space group; they are located on either the twofold axes or the centers of symmetry. The reader should make a drawing for space group $P2/c$, using the coordinates listed above and inserting the symmetry elements.

Special Positions on Twofold Axes
The twofold axes lie along the lines $\left[0, y, \frac{1}{4}\right]$, $\left[\frac{1}{2}, y, \frac{1}{4}\right]$, $\left[0, y, \frac{3}{4}\right]$, and $\left[\frac{1}{2}, y, \frac{3}{4}\right]$. The equivalent positions generated by the space-group symmetry show that the special position sets are

$$\pm \{0, y, \tfrac{1}{4}\} \quad \text{or} \quad \pm \{\tfrac{1}{2}, y, \tfrac{1}{4}\}$$

and each set satisfies $P2/c$ symmetry by accommodating in the unit cell two structural entities with symmetry 2.

Special Positions on Centers of Symmetry
If we repeat the above analysis for the eight centers of symmetry in the space group, we will develop four special position sets:

$$\begin{array}{cc} 0,0,0 & 0,0,\frac{1}{2} \\ \frac{1}{2},0,0 & \frac{1}{2},0,\frac{1}{2} \\ 0,\frac{1}{2},0 & 0,\frac{1}{2},\frac{1}{2} \\ \frac{1}{2},\frac{1}{2},0 & \frac{1}{2},\frac{1}{2},\frac{1}{2} \end{array}$$

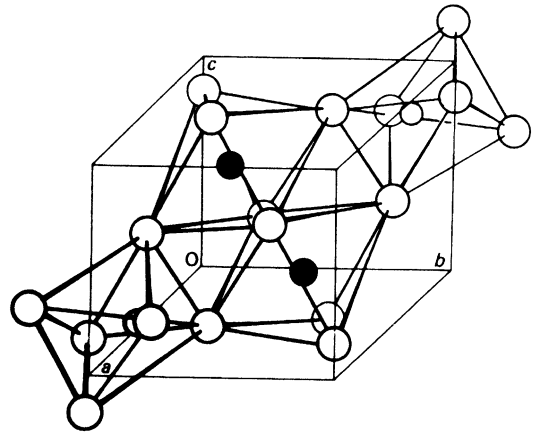


Fig. 7.5 Structure of $NiWO_4$, showing the WO_6 and NiO_6 octahedra: large open circles O, small open circles W, small black circles Ni

The Ni and W atoms must lie on special positions, with either 2 or $\bar{1}$ symmetry. Nothing can be said about the position of the oxygen atoms, and without further detailed analysis we cannot define this structure further. However, to complete the picture, we list the atomic parameters for this structure, and illustrate it in Fig. 7.5:

$$
\begin{array}{lll}
2\text{Ni} & \pm\{\frac{1}{2}, 0.653, \frac{1}{4}\} & \\
2\text{W} & \pm\{0, 0.180, \frac{1}{4}\} & \\
4\text{O} & \pm\{0.22, 0.11, 0.96; & 0.22, 0.89, 0.46\} \\
4\text{O}' & \pm\{0.26, 0.38, 0.39; & 0.26, 0.62, 0.89\}
\end{array}
$$

The heavy atoms (W and Ni) were found to occupy the four twofold axes in pairs. This conclusion, although not uniquely derivable from the symmetry analysis alone, was at least partially indicated by it. Once again, a pencil and paper operation saved considerable effort in the subsequent detailed structure analysis by pointing to the proper course of action.

In these few examples, we have shown the value of a symmetry analysis in the early stages of a structure determination. The procedure may be regarded as a routine to be carried out before the more detailed calculations required in the elucidation of the atomic parameters.

7.3 Interpretation of Electron Density Distributions

We have discussed different forms of the electron density equation in Sect. 6.3ff and we now make use of that theory in studying the distribution of electron density in crystal structures.

Electron density is concentrated in the vicinity of atoms, and rises to peaks at electron density maxima, which correspond to atomic positions within the limits of experimental error, and fall to relatively low values between the peaks. The wavelengths of X-rays used in crystal structure analysis are too long to reveal the intimate electronic structure of atoms themselves, which appear, therefore, somewhat blurred in the calculated electron density function. In general, the more complete and accurate the experimental F_o data, the better will be the atomic resolution and the more precise the final structure model.

7.3.1 Peak Heights and Weights

To a first approximation, the heights of the peaks in an electron density distribution of a crystal are proportional to the corresponding atomic numbers. The hydrogen atom, at the extreme low end of the atomic numbers, is rarely resolved in electron density maps; its small electron density merges into the background density that arises from errors in both the data and the structure model. However, hydrogen atoms can be detected by a difference-Fourier technique, as discussed later, Sect. 8.4.5, and by neutron diffraction; see also Sects. 11.1, 11.4.1, and 11.5.

A better measure of the electron content of a given atom may be obtained from an integrated peak weight, in which the absolute values of $\rho(xyz)$ are summed over the volume occupied by the atom. This technique makes some allowance for the variation of individual atomic temperature factors, high values of which tend to decrease peak heights for a given electron content.

7.3.2 Computation and Display of Electron Density Distributions

Assuming for the moment that phases are available, the electron density function may be calculated over a grid of chosen values of x, y, and z. For this purpose, the unit cell is divided into a selected number of equal divisions, in a manner similar to that employed in the synthesis of the square-wave function, Sect. 6.2.1. Intervals corresponding to about 0.3 Å are satisfactory for most electron density maps. The symmetry of $\rho(xyz)$ is that of the space group of the crystal under investigation. Consequently, a summation over a volume either equal to or just greater than that of the asymmetric unit is adequate.

In order to facilitate the interpretation of $\rho(xyz)$, it is essential to present the distribution of the numerical values in such a way that the geometric relationships between the peaks are easily inspected. This feature is afforded by first calculating the electron density in sections, each corresponding to a constant value of x, y, or z. Each section consists of a field of figures arranged on a grid, which may be true to scale for preference, and can be contoured by lines passing through points of equal electron density, interpolating as necessary, Fig. 7.6. The grading of the contour intervals is selected to produce a reasonable number of contours around the higher-density areas. The contouring should be carried out with care; this exercise leads to fairly precise peak positions and a desirable familiarity with the problem. Sophisticated map-plotting and peak-searching facilities are available, but they should be treated with caution by the beginner.

The contoured sections may be transferred to a transparent medium, such as thin perspex or clear acetate sheets, which are then stacked at the requisite spatial intervals and viewed over a diffuse light source. The diagram in Fig. 1.7 is a photograph of such a display, extending through 17 sections.

An alternative method of displaying the results of an electron density calculation is by means of a ball-and-stick model. An example of this form of representation is shown in Fig. 7.7.

In the analysis of small molecules, it is not considered necessary always to plot and contour the electron density function, although it can be done through some program packages, such as WinGX. There are many graphics programs available that recognize the highest peaks and carry out a geometrical interpretation in terms of their coordinates: Platon for small molecules, and O, Turbo Frodo or Coot for proteins (see Appendix D).

7.3.3 Projections

The use of two-dimensional studies in crystallography is fairly restrictive but, nevertheless, worthy of mention because of the relative ease of calculation and preparation of Fourier maps. For example, the function

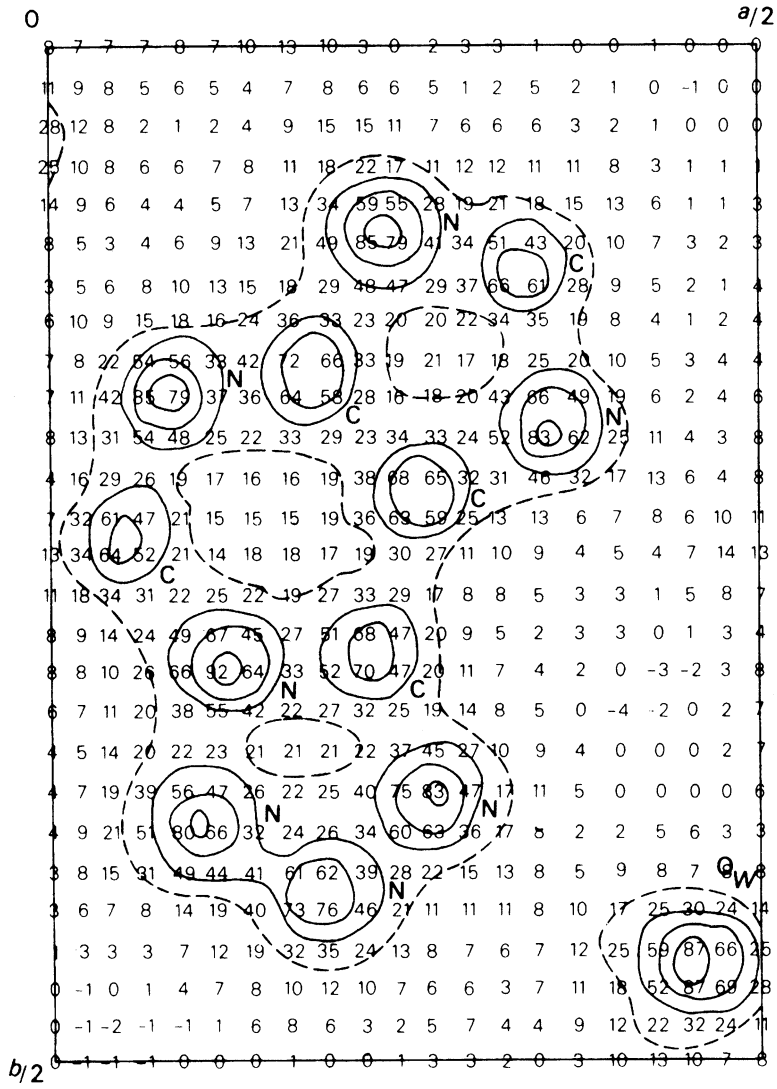


Fig. 7.6 Two-dimensional electron density projection $\rho(xy)$ for azidopurine monohydrate, $C_5H_3N_7{\cdot}H_2O$, calculated from the data of Glusker et al. [3a]. The isolated peak O_w in the *lower right-hand* region of the map represents the oxygen atom of the water molecule. Hydrogen-atom positions are not obtained in this electron density synthesis. The field figures are 10 $\rho(xy)$ $Å^{-2}$ contoured at intervals of 20 units

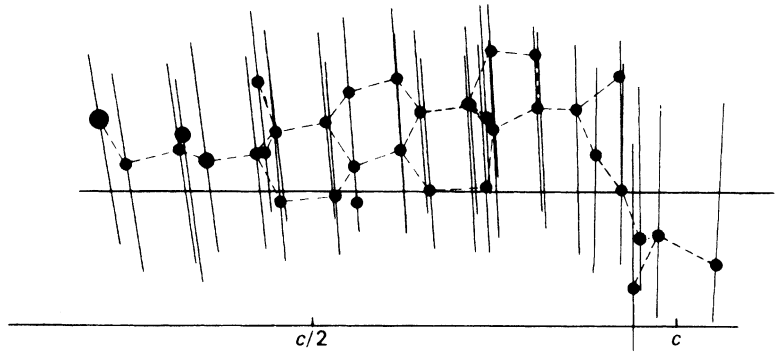


Fig. 7.7 Three-dimensional model of euphenyl iodoacetate; see also Figs. 1.7 and 1.8

$$\rho(xz) = \frac{2}{A} \sum_h \sum_l |\mathrm{F}(h0l)| \cos[2\pi(hx + lz) - \phi(h0l)] \tag{7.3}$$

is calculated with the data from only one level of the reciprocal lattice, the zero level, perpendicular to b, and plotted over the area A of the a, c plane, or the asymmetric portion thereof. The simplification in the calculations is offset, however, by a corresponding complexity in the interpretation of the maps, arising from the superposition of peaks in projection on to the given plane, although this effect is not as severe as in one dimension. Equation (7.3) corresponds to the projection of the electron density along the b axis: it is essential to appreciate the difference between the meaning of $\rho(xz)$ and $\rho(x0z)$; the latter represents the section of the three-dimensional electron density function at $y = 0$. Equations for projections along other principal axes may be written by analogy with (7.3).

Even simple atomic arrangements may appear distorted in projections, with individual molecules overlapping to some degree, but we would not wish to discourage their consideration. We shall restrict their use to examples illustrating various aspects of structure analysis. Practice in the calculation and interpretation of Fourier series is provided by Problems 7.11 and 7.12, and also the exercises with the XRAY system; see Sect. 13.4.

7.4 Methods of Solving the Phase Problem

At this point, it may be convenient for the reader to revise some of the ideas presented in Sect. 6.9ff since a general overview of the present topic was presented there.

The set of $F_o(hkl)$ data constitutes the starting point of an X-ray structure determination. The approximate number of symmetry-independent reflections measurable may be calculated in the following manner.

7.4.1 Number of Reflections in the Data Set

The radius of the limiting sphere is 2 RU (reciprocal lattice units), and its volume is therefore 33.510 RU^3, taking the reciprocal lattice constant κ as λ, Sect. 2.4. The number of reciprocal lattice points within the limiting sphere is approximately equal to the number of times the reciprocal unit-cell volume V^* will fit into 33.510; since, for this application, $V^* = \lambda/V$, this number is $33.510\, V_c/\lambda^3$. The number of symmetry-independent reflections observable, N_{max}, in a given experiment in which θ_{max} represents the practical upper limit of θ is given by

$$N_{max} = 33.510\, V_c \sin^3\theta_{max}/\lambda^3 Gm \tag{7.4}$$

where G is the unit-cell translation constant, Table 3.2, for non-zero reflections and m is the number of symmetry-equivalent reflections, or the number of general equivalent points in the appropriate Laue group. For zones and rows, m may take different values from that for hkl, and a number of systematic absences within the sphere of radius 2 sin θ_{max} may have to be subtracted.

As an example, consider an orthorhombic crystal of space group $Cmm2$, with unit-cell dimensions $a = 9.00$ Å, $b = 10.00$ Å, and $c = 11.00$ Å. For Cu $K\alpha$ radiation ($\lambda = 1.5418$ A) and θ_{max} of 85° ($d_{min} = 0.77$ Å), N_{max} is $(33.510 \times 9 \times 10 \times 11 \times \sin^3 85)/(1.5418^3 \times 2 \times 8) = 559$. If Mo $K\alpha$ radiation (with $\theta_{max} \approx 27°$ and $\bar{\lambda} = 0.71073$ Å) had been used instead of Cu $K\alpha$, the number would have been 5709. Say the structure contains 15 atoms in the asymmetric unit. In the structure analysis, each

atom would be determined by three positional parameters (x_j, y_j, z_j) and, say, one isotropic thermal vibration parameter, which, with an overall scale factor, totals 61 variables. Even with Cu $K\alpha$ radiation, there are nine reflections per variable, a situation which, from a mathematical point of view, is considerably overdetermined. This feature is important, since the experimental intensity measurements contain random errors which cannot be eliminated, and a preponderance of data is needed to ensure good precision in the structural parameters. We shall meet this situation again in Chap. 8 but such a degree of over-determination is not usually possible with macromolecules, as we explain in Chap. 10.

7.4.2 The Patterson Function

Although the connection between Fourier theory and X-ray diffraction was recorded first in 1915, it was not until about 1930 that very much practical use was made of it. Before the advent of computing facilities, the calculation of even a Fourier projection, involved considerable time and effort. Add to this the phase problem, which necessitated many such calculations, and it is easy to understand that X-ray analysts were not anxious to become involved with extensive Fourier calculations; many early structure analyses were based on two projections.

In 1934, Patterson reported a new Fourier series which could be calculated directly from the experimental intensity data. However, because phase information is not required in the Patterson series, the result cannot be interpreted as a set of atomic positions, but rather as a collection of interatomic vectors all taken to a common origin, Sect. 6.9.2. Patterson was led to the formulation of his series from considerations of an earlier theory of Debye on the scattering of X-rays by liquids—a much more difficult problem.

Patterson functions are of considerable importance in X-ray structure analysis, and their application will be considered in some detail. We will study first a one-dimensional function.

One-Dimensional Patterson Function

The electron density at any fractional coordinate x is $\rho(x)$, and that at the point $(x + u)$ is $\rho(x + u)$. The average product of these two electron densities in a repeat of length a, for a given value of u, is

$$A(u) = \int_0^1 \rho(x)\rho(x+u)\,\mathrm{d}x \tag{7.5}$$

where the upper limit of integration corresponds to the use of fractional coordinates. Using (6.21) in a form appropriate to a one-dimensional unit repeat, we obtain

$$A(u) = \int_0^1 \frac{1}{a^2} \sum_h |\mathrm{F}|(h)\mathrm{e}^{-\mathrm{i}2\pi hx} \sum_{h'} |\mathrm{F}|(h')\mathrm{e}^{-\mathrm{i}2\pi h'(x+u)}\,\mathrm{d}x \tag{7.6}$$

where $|\mathrm{F}|$ is an amplitude (phase-free) term. The index h' lies within the same range as h, but is used to effect distinction between the Fourier series for $\rho(x)$ and $\rho(x + u)$. Separating the parts dependent upon x, and remembering that the integral of a sum is the sum of the integrals of the separate terms, we may write

$$A(u) = \frac{1}{a^2} \sum_h \sum_{h'} |\mathrm{F}|(h)|\mathrm{F}|(h')\mathrm{e}^{-\mathrm{i}2\pi h'u} \int_0^1 \mathrm{e}^{-\mathrm{i}2\pi(h+h')x}\,\mathrm{d}x \tag{7.7}$$

Considering the integral

$$\int_0^1 \mathrm{e}^{-\mathrm{i}2\pi(h+h')x}\mathrm{d}x = \frac{\mathrm{e}^{-\mathrm{i}2\pi(h+h')x}}{-\mathrm{i}2\pi(h+h')}\Bigg]_0^1 \tag{7.8}$$

$\mathrm{e}^{-\mathrm{i}2\pi(h+h')}$ is unity, since h and h' are integral, from de Moivre's theorem, Sect. 3.2.3, and the integral is, in general, zero. However, for the particular value of h' equal to $-h$, it becomes indeterminate and we must consider making this substitution before integration. Thus,

$$\int_0^1 \mathrm{d}x = 1 \tag{7.9}$$

Hence, from (7.7), for non-zero values of $A(u)$, where $h' = -h$,

$$A(u) = \frac{1}{a^2}\sum_h \sum_{-h} |\mathrm{F}|(h)|\mathrm{F}|(-h)\mathrm{e}^{\mathrm{i}2\pi hu} \tag{7.10}$$

Equation (7.10) is not really a double summation, since h and $-h$ cover the same field of the function. From Fig. 3.23, we see that the magnitudes $|\mathrm{F}|(h)$ and $|\mathrm{F}|(-h)$ are equal; hence, (7.10) becomes

$$A(u) = \frac{1}{a^2}\sum_h |\mathrm{F}(h)|^2\mathrm{e}^{\mathrm{i}2\pi hu} \tag{7.11}$$

where the index h ranges from $-\infty$ to ∞. Taking h from 0 to ∞, (7.11) may be written as

$$A(u) = \frac{1}{a^2}\sum_h \left(|\mathrm{F}(h)|^2\mathrm{e}^{\mathrm{i}2\pi hu} + |\mathrm{F}(h)|^2\mathrm{e}^{-\mathrm{i}2\pi hu}\right) \tag{7.12}$$

Applying de Moivre's theorem, we obtain

$$A(u) = \frac{2}{a^2}\sum_h |\mathrm{F}(h)|^2 \cos 2\pi hu \tag{7.13}$$

The corresponding Patterson function $P(u)$ is usually defined as

$$P(u) = \frac{2}{a}\sum_h |\mathrm{F}(h)|^2 \cos 2\pi hu \tag{7.14}$$

a small difference from the averaging function $A(u)$.

The practical evaluation of $P(u)$ proceeds through (7.14), but its physical interpretation is best considered in terms of (7.5), neglecting the small difference between $P(u)$ and $A(u)$.

Figure 7.8a shows one unit cell of a one-dimensional structure containing two different atoms A and B situated at fractional coordinates x_A and x_B, respectively. Equation (7.5) represents the value of the electron density product $\rho(x)\rho(x + u)$, for any given value of u, averaged over the repeat period of the unit cell. The average will be zero if one end of the vector **u** always lies in a zero region of electron density, small if both ends of the vector encounter low electron densities, large if the electron density

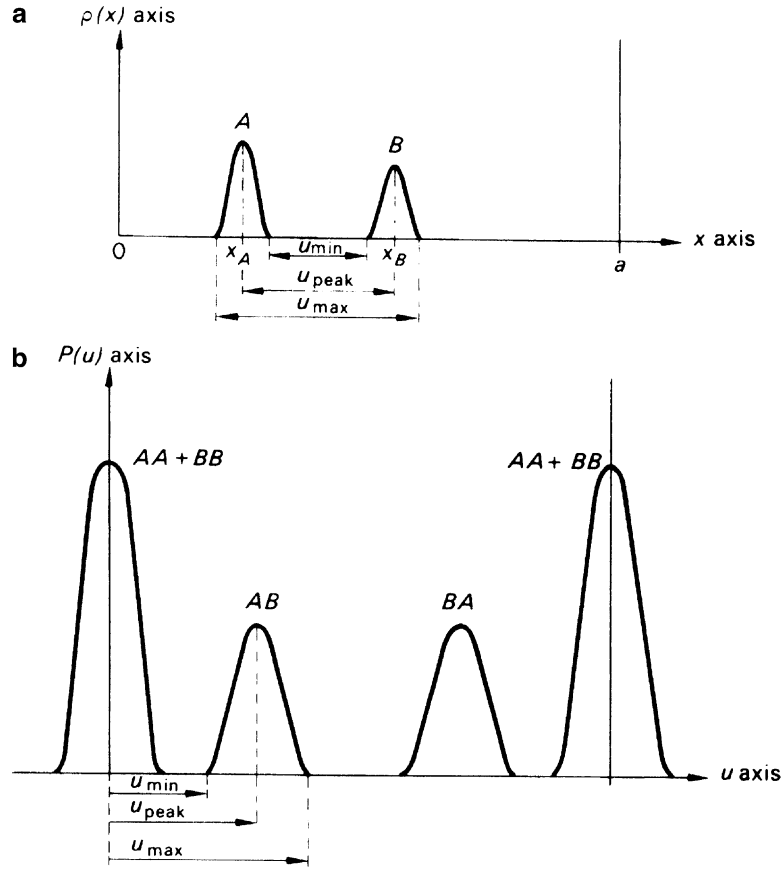


Fig. 7.8 Development of a one-dimensional Patterson function (**b**) for a two-atom structure (**a**). Note the centrosymmetry of the Patterson function that is lacking in the $\rho(x)$ function

products are large, and a *maximum* where u is of such a length that it spans two atomic positions in the unit cell.

For values of u less than u_{min} in Fig. 7.8a, no peak will arise from the pair of atoms. As u is increased, however, both ends of the vector will come simultaneously under the electron density peaks, and from (7.5) a finite value of $A(u)$, or $P(u)$, will be obtained. The integration can be simulated by sliding a vector of a given magnitude u along the x axis, evaluating the average product $\rho(x)$ $\rho(x + u)$ for all sampling intervals between zero and unit fractional repeat; this process is carried out for all fractional values of u between 0 and 1. The graph of $P(u)$ as a function of u is similar in appearance to an electron density function, but we must be careful not to interpret it in this way.

As we proceed through the values of u, we encounter u_{peak}, the interatomic vector A–B, which gives rise to the maximum value of $P(u)$, labeled AB in Fig. 7.8b. As u increases to u_{max}, the electron density product falls to zero and $P(u)$ decreases correspondingly. Since we are concerned with interatomic *vectors*, negative values of u are equally important; $-AB$ is marked off on the negative side of the origin, or at BA within the given unit cell.

If we consider next *very* small values of u, both ends of such vectors will lie inside one and the same electron density peak, and $P(u)$ will be large. In the limit as $u \rightarrow 0$, the product involves that of the electron density maximum with itself, which is a local maximum for each atom, and a very large peak at the origin ($u = 0$) is to be expected. Thus the Patterson function is represented as a map of interatomic vectors, including null vectors, all taken to the origin, Sect. 6.9.2.

The reader should confirm from Fig. 7.8, using tracing paper, that the positions of the peaks in Patterson space can be plotted graphically by placing each atom of the structure $\rho(x)$ in turn at the origin of the Patterson map, in parallel orientation, and marking the positions of the other atoms on to the Patterson unit cell. Because of the centrosymmetry of the Patterson function, implicit in (7.14), it is not strictly necessary to plot vectors lying outside one-half of the unit cell.

Three-Dimensional Patterson Function

If we replace $\rho(x)$ and $\rho(x + u)$ in (7.5) by the three-dimensional analogs $\rho(xyz)$ and $\rho(x + u, y + v, z + w)$ and integrate over a unit fractional volume, we can derive the three-dimensional Patterson function:

$$P(uvw) = \frac{2}{V_c} \sum_h \sum_k \sum_l |\mathrm{F}(hkl)|^2 \cos 2\pi(hu + kv + lw) \tag{7.15}$$

where the summations range, in the most general case, over one half of experimental reciprocal space. This equation should be compared with (7.14) and (6.29) in Chap. 6: it is a Fourier series with zero phases and $|\mathrm{F}|^2$ as coefficients. Since $|\mathrm{F}|^2$ is $\mathrm{F} \cdot \mathrm{F}^*$, we see from (6.28) that (7.15) represents the convolution of the electron density $\rho(\mathbf{r})$ with its inversion in the origin, that is, with $\rho(-\mathbf{r})$. In practice, (7.15) may be handled like the corresponding electron density equation, with u, v, w replacing x, y, z. Both series explore the same field but their interpretation is different, as we shall see. The roving vector is now specified by three coordinates, u, v, and w, and $P(uvw)$ is a maximum where the corresponding vector spans two atoms in the crystal.

7.4.3 Positions and Weights of Peaks in the Patterson Function

The positions of the peaks in $P(uvw)$ may be plotted in three dimensions by placing each atom of the unit cell of a structure in turn at the origin of Patterson space, in parallel orientation, and mapping the positions of all other atoms on to the Patterson unit cell. Examples of this process are illustrated graphically in Fig. 7.9; for simplicity the origin peak is not shown in Fig. 7.9d. In Fig. 7.9a, all atoms and their translation equivalents produce vector peaks lying on the points of a lattice that is identical in shape and size to the crystal lattice. For example, atom 1 at x, y, z and its translation equivalent, $1'$, at x, $1 + y, z$ give rise to a vector ending at 0, 1, 0 on the Patterson map. Peaks of this nature accumulate at the corners of the Patterson unit cell in exactly the same way as those of the origin peak, $P(000)$. From (7.15), we can derive the height of the origin peak:

$$P(000) = \frac{2}{V_c} \sum_{h=0} \sum_k^{\infty} \sum_{l=-\infty} |\mathrm{F_o}(hkl)|^2 \tag{7.16}$$

In general, (7.16) is equivalent to a superposition at the origin of all N products like $\rho(xyz)\rho(xyz)$, where N is the number of atoms in the unit cell. Since $\rho(xyz)$, is proportional to the atomic number Z_j of the jth atom, Sect. 7.3, we have

$$P(000) \propto \sum_{j=1}^{N} Z_j^2 \tag{7.17}$$

A single vector interaction between two atoms j and k, Fig. 7.9b, will have a Patterson peak of height proportional to $Z_j Z_k$. Hence, the height $H(j, k)$ of this peak will be given by

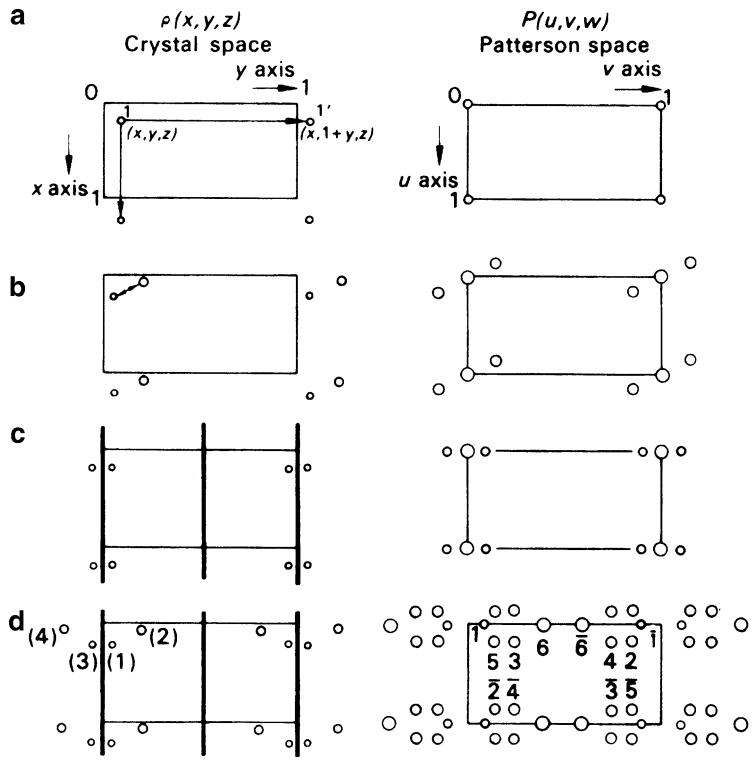


Fig. 7.9 Effects of symmetry-related and symmetry-independent atoms on the Patterson function. The weights of the peaks are approximately proportional to the diameters of the circles: (**a**) $P1(N = 1)$. (**b**) $P1(N = 2)$: two atoms per unit cell produce $(2^2 - 2)$ non-origin peaks. (**c**) $Pm(N = 2)$: two non-origin peaks, but with coordinates $\pm\{0, 2y, 0\}$. (**d**) $Pm(N = 4)$: 12 non-origin peaks per unit cell; for clarity the origin peak has not been drawn. The *Patterson* space group is $P\bar{1}$ in (**a**) and (**b**) and $P2/m$ in (**c**) and (**d**)

$$H(j,k) \approx P(000) Z_j Z_k / \sum_{j=1}^{N} Z_j^2 \tag{7.18}$$

where $P(000)$ is calculated from (7.16). This equation can serve as a useful guide, but overlapping vectors may give rise to misleading indications. The reservations on peak heights already mentioned in Sect. 7.3 apply also to Patterson peaks. It should be remembered that the correct geometrical interpretation of Patterson peaks is of far greater significance than is an adherence to (7.18).

In a structure with N atoms per unit cell, each atom forms a vector with the remaining $N - 1$ atoms. There are, thus, $N(N - 1)$ non-origin peaks per unit cell. From (7.15), substitution of $-u$, $-v$, $-w$ for u, v, w, respectively, leaves $P(uvw)$ unaltered, which is a statement of the centrosymmetry of the Patterson function.

The Patterson unit cell is of the same size and shape as the crystal unit cell, but it has to accommodate N^2 rather than N peaks and is, therefore, correspondingly overcrowded. Thus, peaks in Patterson space tend to overlap when there are many atoms in the unit cell, a feature which introduces difficulties into the process of unraveling the function in terms of the correct distribution of atoms in the crystal.

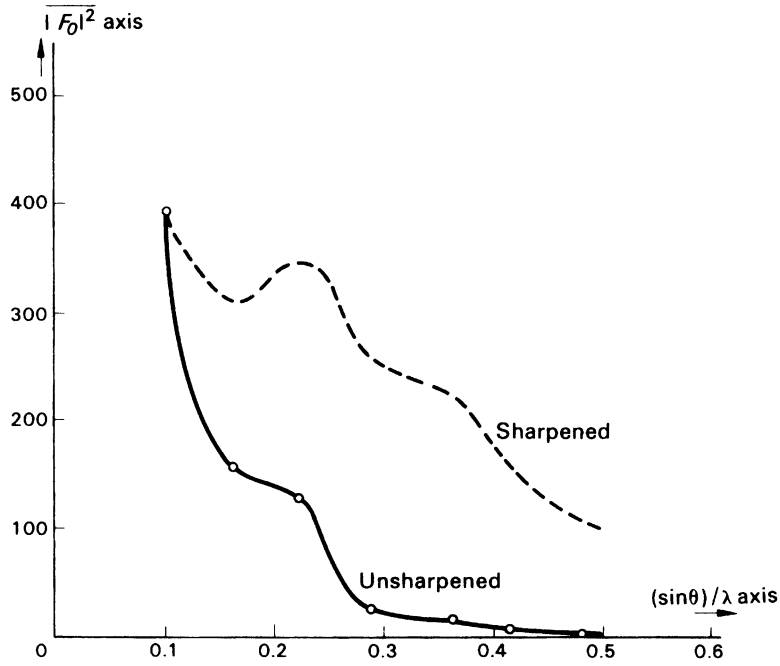


Fig. 7.10 Effect of sharpening on the radial decrease of the local average intensity $\overline{F_o{}^2}$

7.4.4 Sharpened Patterson Function

In a conceptual point atom, the electrons would be concentrated at a point. The atomic scattering factor curves, Fig. 4.6, would be parallel to the abscissa and f would be equal to the atomic number for all values of $\sin\theta/\lambda$ and at all temperatures. The electron density for a crystal composed of point atoms would show a much higher degree of resolution than does that for a real crystal. Put another way, the broad peaks representing real atoms, Fig. 6.4, would be replaced by peaks of very narrow breadth in the point-atom crystal.

A plot of the mean value of $F_o{}^2$ against $\sin\theta/\lambda$ for a typical set of data is shown in Fig. 7.10. The radial decrease in $\overline{F_o{}^2}$ can be reduced by modifying $F_o{}^2$ by a function which increases as $\sin\theta/\lambda$ increases. The coefficients for a *sharpened* Patterson synthesis may be calculated by the following equation. Sharpening can be effected also through the use of |E| values, Sect. 8.2.1.

$$\left\{ F_{o,\mathrm{mod}}{}^2 = \frac{F_o{}^2}{\exp[-2B(\sin^2\theta)/\lambda^2]\left\{\sum_{j=1}^{N} f_j\right\}^2} \right\}_{hkl} \tag{7.19}$$

where N is the number of atoms in the unit cell and B is an overall isotropic temperature factor, Sect. 4.1.8.

The effect of sharpening on a Patterson synthesis is illustrated in Fig. 7.17d, the Harker section $\left(u, \frac{1}{2}, w\right)$ for papaverine hydrochloride. It should be compared with Fig. 7.17b; the increased resolution is very apparent.

Over-sharpening of Patterson coefficients may lead to spurious peaks because of series termination errors, Sect. 6.9.1, particularly where heavy atoms are present, and the technique should not be

Table 7.1 Vectors generated by two independent atoms and their symmetry equivalents in space group Pm[a]

		Subtraction of coordinates		Reduced to one unit cell		
Atom pair	Analytical form of vector	u	v	u	v	Point in Fig. 7.9d
(1), (3)	$\pm\{0, 2y_1, 0\}$	0	0.10	0	0.10	1
		0	−0.10	0	0.90	$\bar{1}$
(1), (2)	$\pm\{x_1 - x_2, y_1 - y_2, z_1 - z_2\}$	0.15	−0.15	0.15	0.85	2
		−0.15	0.15	0.85	0.15	$\bar{2}$
(1), (4)	$\pm\{x_1 - x_2, y_1 + y_2, z_1 - z_2\}$	0.15	0.25	0.15	0.25	3
		−0.15	−0.25	0.85	0.75	$\bar{3}$
(2), (3)	$\pm\{x_1 - x_2, -y_1 - y_2, z_1 - z_2\}$	0.15	−0.25	0.15	0.75	4
		−0.15	0.25	0.85	0.25	$\bar{4}$
(3), (4)	$\pm\{x_1 - x_2, -y_1 + y_2, z_1 + z_2\}$	0.15	0.15	0.15	0.15	5
		−0.15	−0.15	0.85	0.85	$\bar{5}$
(2), (4)	$\pm\{0, 2y_2, 0\}$	0	0.40	0	0.40	6
		0	−0.40	0	0.60	$\bar{6}$

[a]The coordinates of the four atoms in two sets of general positions are x, y, z; $x, \bar{y}, z$ with $x_1 = 0.20$, $y_1 = 0.05$, $x_2 = 0.05$, $y_2 = 0.20$, and $z_1 = z_2 = 0.00$

applied without care. Sometimes the coefficients can be further modified to advantage by multiplication by a function such as $\exp(-m \sin^3 \theta)$, where m is chosen by trial, but might be about 5. This function has the effect of decreasing the magnitude of the $\overline{F_o{}^2}$ curve at high θ-values. Many other sharpening functions have been proposed, but we shall not dwell on this subject. It is often helpful to calculate both the normal and sharpened Patterson functions for comparison. Practice can be gained through exercises with the XRAY program system.

7.4.5 Symmetry of the Patterson Function for a Crystal of Space Group *Pm*

An inspection of Fig. 7.9c, d shows that the peaks on the line $[0, v, 0]$ arise from atom pairs related by the m planes. The vector interactions for case (d) are listed in Table 7.1, and may be easily verified by the reader; the values $z_1 = z_2 = 0.0$ were chosen for convenience only.

The m planes in Pm are carried over into Patterson space, and relate the following pairs of peaks in the vector set:

$$1, \bar{1}; \quad 2, 5; \quad \bar{2}, \bar{5}; \quad 3, 4; \quad \bar{3}, \bar{4}; \quad 6, \bar{6} \tag{7.20}$$

Furthermore, the presence of a center of symmetry in the diffraction pattern generates $2/m$ symmetry in the Patterson map, which corresponds to the Laue symmetry of all monoclinic crystals. Evidently, the symmetry of the diffraction pattern is impressed on to the Patterson function by the use of F_o^2 coefficients in the Patterson series. As a consequence, the Patterson synthesis is computed in the primitive space group corresponding to the Laue symmetry of a crystal, and this situation is similar for all space groups.

We can detect the presence of the twofold axis parallel to b in Fig. 7.9d through vector peaks such as 5, $\bar{2}$ and 3, $\bar{4}$. Finally, the symmetry-related pairs of atoms in the crystal, 1, 3 and 2, 4, give rise to vectors along the line $[0, v, 0]$ corresponding to the peaks 1, 6, $\bar{6}$, and $\bar{1}$ in Patterson space. The presence of a large number of peaks along an axis in a three-dimensional Patterson map may be used as evidence for a mirror plane perpendicular to that axis in the crystal. This feature is important because an m plane does not give rise to systematic absences in the diffraction pattern, Table 3.7. The existence of peaks, arising from symmetry-related atoms in certain regions of Patterson space was noted first by Harker in 1936. The line

[0, v, 0] for Pm is called a Harker line; *planes* containing peaks arising from pairs of symmetry-related atoms are called Harker sections. We shall consider some examples below.

7.4.6 Vector Interactions in Other Space Groups

We shall consider atoms in general positions in a number of space groups that should be now familiar.

Space Group $P\bar{1}$
General positions: $x, y, z; \bar{x}, \bar{y}, \bar{z}$.
Vectors: $\pm\{2x, 2y, 2z\}$.
Harker peaks lie in general positions in Patterson space.

Space Group $P2$
General positions: $x, y, z; \bar{x}, y, \bar{z}$.
Vectors: $\pm\{2x, 0, 2z\}$.
Harker section: $(u, 0, w)$.

It may be noted that for complex structures, not all of the peaks on Harker sections are necessarily true Harker peaks. If in this structure there are two atoms not related by symmetry, which, by chance, have the same or very nearly the same y coordinates, the vector between them will produce a peak on the Harker section.

Space Group $P2/m$
Vectors:

$\pm\{2x, 0, 2z\}$	Double weight	Type 1
$\pm\{0, 2y, 0\}$	Double weight	Type 2
$\pm\{2x, 2y, 2z\}$	Single weight	Type 3
$\pm\{2x, 2\bar{y}, 2z\}$	Single weight	Type 4

Harker section: $(u, 0, w)$.
Harker line: [0, v, 0].

Vector type 1 arises in two ways, once from the pair $x, y, z; \bar{x}, y, \bar{z}$, and once from the pair $x, \bar{y}, z; \bar{x}, \bar{y}, \bar{z}$. These two interactions give rise to identical vectors, which therefore superimpose in Patterson space and form a double-weight peak. Similar comments apply to type 2, but the centrosymmetrically related atoms give rise to single-weight peaks, types 3 and 4. Figure 7.11 illustrates these vectors, as seen along the z axis. The reader may now consider how the Patterson function might be used to differentiate among space groups $P2$, Pm, and $P2/m$. Statistical methods, discussed in Chap. 4, are often employed to verify the results obtained from a study of the vector distribution.

7.4.7 Examples of the Use of the Patterson Function in Solving the Phase Problem

In this section, we shall consider how the Patterson function was used in the solution of three quite different structures.

Bisdiphenylmethyldiselenide, $(C_6H_5)_2CHSe_2\ CH(C_6H_5)_2$
Crystals of this compound form yellow needles, with straight extinction under crossed Polaroids for all directions parallel to the needle axis, and oblique extinction on the section normal to the needle

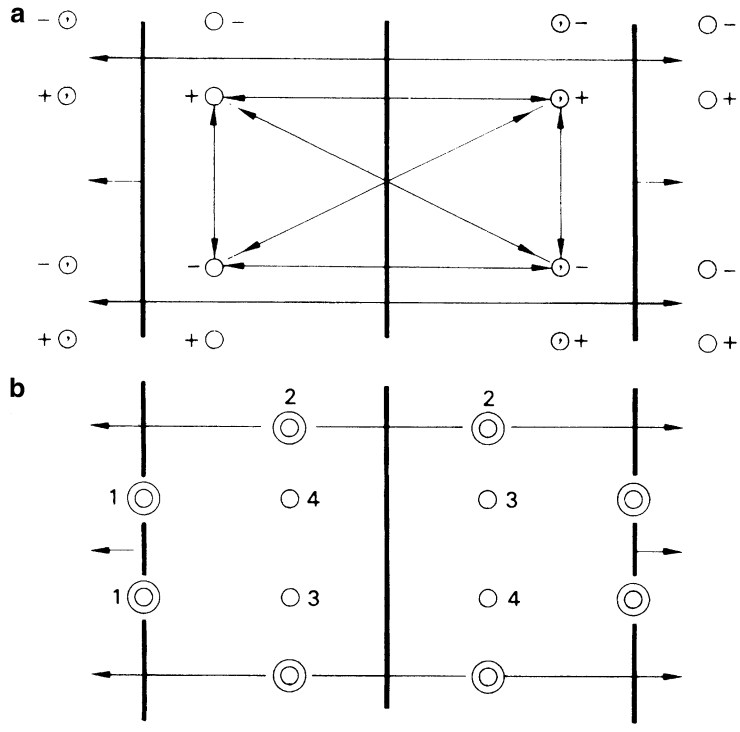


Fig. 7.11 (**a**) Vectors between symmetry-related atoms in general equivalent positions in space group $P2/m$. Coordinates like $\bar{x}$ have been treated as $(1 - x)$ in drawing the vectors. (**b**) One unit cell of the Patterson function: the twofold axes intersect the m planes in centers of symmetry. Note the single-weight and double-weight peaks, and their relation to the space-group symmetry

axis. Photographs taken with the crystal oscillating about its needle axis show only a horizontal m line, while zero- and upper-layer Weissenberg [4] photographs show only symmetry 2. The crystals are therefore monoclinic, with b along the needle direction.

Crystal Data
System: monoclinic
Unit-cell dimensions: $a = 18.72$, $b = 5.773$, $c = 12.594$ Å, $\beta = 125.47°$
V_c: 1,108.5 Å^3
D_m: 1.49 g cm^{-3}
M_r: 492.38
Z: 2.02 or 2 to the nearest integer
Unit-cell contents: 4Se, 52C, and 44H atoms
Absent spectra: hkl: $h + k = 2n$
Possible space groups: $C2$, Cm, $C2/m$

Symmetry Analysis
Where the space group is not determined uniquely by the X-ray diffraction pattern, it may be possible to eliminate certain alternatives at the outset of the structure determination by other means.

Space groups $C2$ and Cm each require four general positions:

$$C2 : (0,0,0;\quad \tfrac{1}{2},\tfrac{1}{2},0) + \{x,y,z;\quad \bar{x},y,\bar{z}\} \qquad Cm : (0,0,0;\quad \tfrac{1}{2},\tfrac{1}{2},0) + \{x,y,z;\quad x,\bar{y},z\}$$

Since Z is 2, the molecular symmetry is either 2, in $C2$, or m, in Cm. In both $C2$ and Cm, all atoms could satisfy general position requirements, and neither arrangement would be stereochemically unreasonable.

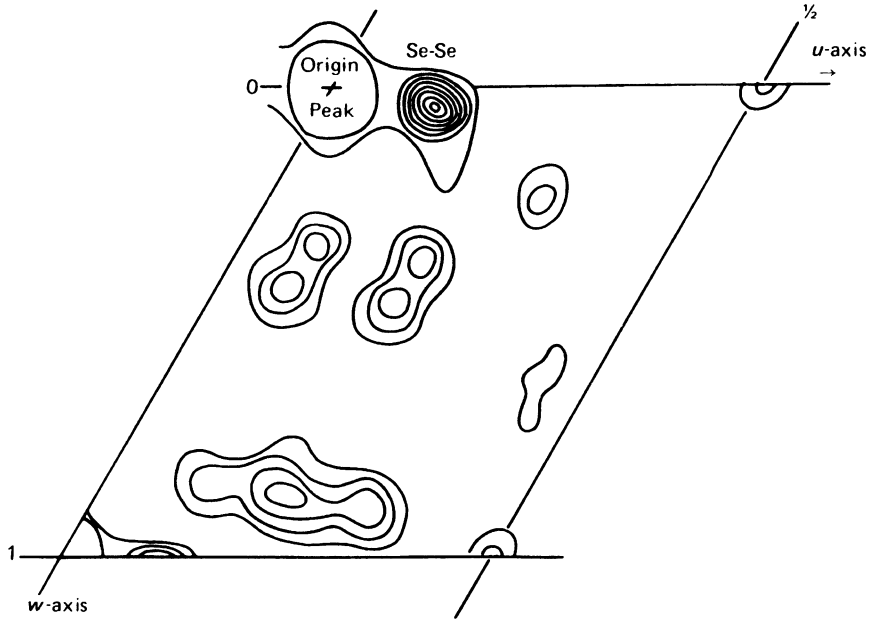


Fig. 7.12 Patterson section, $P(u0w)$; the origin peak (height = 100) has not been contoured. Contours around the Se–Se peaks are at intervals of 4; elsewhere at intervals of 2

Space group $C2/m$ requires eight general equivalent positions per unit cell. Only special position sets, such as 0, 0, 0 and $\frac{1}{2}, \frac{1}{2}, 0$ correspond with $Z = 2$. These positions have symmetry $2/m$, but it is not possible to construct the molecule in this symmetry without contradicting known chemical facts. Consequently, we shall regard this space group as highly improbable for the compound under investigation.

Patterson Studies

Whatever the answer to the questions remaining from this symmetry analysis, we expect, from the covalent radius of selenium (1.22 Å) that the two selenium atoms will be covalently bonded at a distance of about 2.4 Å. This Se–Se interaction will produce a strong peak in the Patterson function at about this distance from the origin.

The atomic numbers of Se, C, and H are 34, 6, and 1, respectively. Hence, the important vectors in the Patterson function would have approximate single-weight peak heights, from (7.18), as follows:

(a) Se–Se: 1156.
(b) Se–C: 204.
(c) C–C: 36.

Because of the presence of identical vectors arising from the C unit cell, all vectors will be double these values.

Figure 7.12 is the Patterson section $P(u0w)$, calculated with 1053 $F_o(hkl)^2$ data, with grid intervals of 50th along u, v, and w. The origin peak $P(000)$ was scaled to 100 and, from (7.17), $\sum_{j=1}^{N} Z_j^2 = 6540$. Hence, the vector interactions (a), (b), and (c) should have peak heights in the approximate ratio 32:5.7:1.

The section is dominated by a large peak of height 39 at a distance of about 2.4 Å from the origin. Making the reasonable assumption that it represents the Se–Se vector, and since there are no

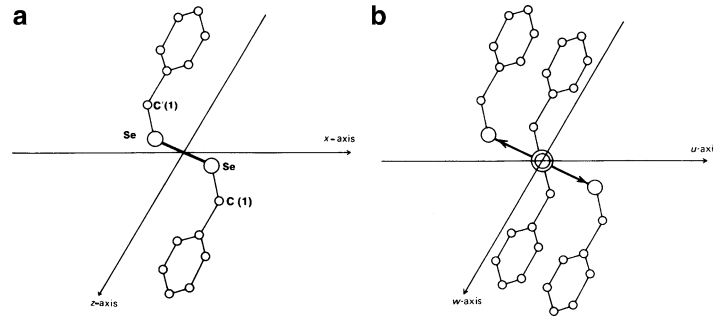


Fig. 7.13 (**a**) Hypothetical structure fragment $C_6H_5CHSe_2CHC_6H_5$ in real space; (**b**) idealized set of Se–Se and Se–C vectors in Patterson space

significant peaks on the v axis, the Harker line in Cm, it follows that the space group cannot be Cm, thus leaving $C2$ as the most logical choice.

In space group $C2$ from the above list, it follows that vectors between atoms in general positions take the form: $(u, 0, w)$, where $u = 2x$, and $w = 2z$, plus the C-centered equivalents. For the sulphur atom vectors $u = 2x_{Se}$, $w = 2z_{Se}$. Hence by measurement on the section, the Patterson coordinates are $u = 6.7/50 = 0.134$ and $w = 2.2/50 = 0.044$, so that $x_{Se} = 0.067$ and $z_{Se} = 0.022$.

In space group $C2$, the unit-cell origin is fixed in the x, z plane by the twofold axis. There is no symmetry element that defines the origin in the y direction, which must be fixed by specifying the y coordinate for a selected atom. For convenience, we may set $y_{Se} = 0$, and our analysis so far may be given as the positions

Se: 0.067, 0, 0.022

Se′: −0.067, 0, −0.022

A space-group ambiguity is not always resolved in this manner. Sometimes it is necessary to proceed further with the structure analysis, even to refinement stages, before confirmation is obtained.

What of the atoms other than selenium? Is it possible to determine the positions of the carbon and hydrogen atoms? We shall find that we can locate the carbon atoms in this structure from the Patterson synthesis. To explain the procedure, we consider first only part of the structure, including one phenyl ring of the asymmetric unit, Fig. 7.13a, and neglect all but the C–Se vectors. The vector set generated by the 2 Se atoms and 14 C atoms in this hypothetical arrangement contains two images of the structure fragment, one per Se atom, which are displaced from each other by the Se–Se vector. The idealized vector set is shown in Fig. 7.13b. By shifting one of these images by a *reverse*[1] Se–Se vector displacement, it is possible to bring the two images into coincidence. Verify this statement by making a transparent copy of Fig. 7.13b and placing its origin over an Se–Se vector position in the original figure, keeping the pairs of u and w axes parallel. Certain peaks overlap, producing a single, displaced image of the structure. Shade the peaks that overlap. This image is displaced with respect to the true space-group origin, which we know to be midway between the two Se atoms. A correctly placed image of the structure can be recovered by inserting the true origin position on to the tracing and neglecting any peaks that are not shaded.

The partial vector set was formed from the image of all atoms of the fragment in each Se atom; each image is weighted by Z_j, the atomic number of the jth atom, (carbon, in this example), imaged in Se. The displacement arises because the Patterson synthesis transfers all vectors to a common origin.

[1] If the forward direction of this vector is used, the structure obtained would, in general, be inverted through the origin. This does not happen with the example under study because the molecule possesses only twofold symmetry.

Patterson Superposition

The technique just described depends upon the recognition of the vector interaction from a given pair of atoms, the two Se atoms in this example. At least a partial unscrambling of the structure images in the Patterson function was effected by correctly displacing two copies of the Patterson map and noting the positions of overlap.

To illustrate the method further and to derive a systematic procedure for its implementation, we return to the Patterson section in Fig. 7.12. The two Se atoms have the same y coordinate, which means that the vector shift takes place in this section. Now, make two copies on tracing paper of the half unit-cell *outline*, $x = 0 - \frac{1}{2}$ and $z = 0$–1, and label them copy 1 and copy 2.

On copy 1 mark in the position S of the point, $-(x_{\mathrm{Se}}, z_{\mathrm{Se}})$, which is at -0.067, -0.022, and on copy 2 mark in the position S' of the point $(x_{\mathrm{Se}'}, z_{\mathrm{Se}'})$, which is at 0.067, 0.022. Think of these two unit cells as existing in crystal space, not Patterson space. Place copy 1 over the Patterson (u, 0, w) section, maintaining a parallel orientation, with S over the origin, and trace out the Patterson map, Fig. 7.14a, excluding the origin peak in each case. Repeat this procedure with copy 2, placing S' over the Patterson section origin, Fig. 7.14b.

Finally, superimpose copy 1 and copy 2. As in the exercise with Fig. 7.13a, b, some peaks overlap and some lie over blank regions in one or the other map. The overlaps correspond to regions of high electron density in the crystal. They are best mapped out by compiling a new diagram which contains the *minimum* value of the vector density between copy 1 and copy 2 for each point, thus eliminating or decreasing in height those regions where one copy has no or only slight overlap. A map prepared in this way is shown in Fig. 7.14c.

Minimum Function

The technique outlined above follows the method of Buerger.[2] An analytical expression for the minimum function $M_n(xyz)$ is given by (7.21); it may be regarded as an approximation to the electron density $\rho(xyz)$.

$$M_n(xyz) = \mathrm{Min}[P(u - x_1, v - y_1, w - z_1),\ P(u - x_2, v - y_2, w - z_2), \ldots, P(u - x_n, v - y_n, w - z_n)] \tag{7.21}$$

where $\mathrm{Min}(P_1, P_2, \ldots, P_n)$ represents the lowest value at the point x, y, z in the set of super positions P_1, P_2, …, P_n; n corresponds with the number of known or trial atomic positions. The following general comments on the application of the minimum function procedure should be noted:

1. The n trial atoms should form within themselves a set or sets of points related by the appropriate space-group symmetry.
2. In a non-centrosymmetric space group, n should be three or more in order to remove the Patterson center of symmetry.
3. If the various n trial atoms have different atomic numbers, the corresponding Patterson copies should be weighted accordingly in order to even out the different image strengths.
4. Incorrectly placed atoms in the trial set tend to confuse the structure image. New atom sites therefore should be added to the model with caution.

Figure 7.15 shows a composite electron density map of the atoms in the asymmetric unit that were revealed by a *three-dimensional* minimum function M_2. This result is quite satisfactory; only C(9), C(10), and C(11) are not yet located. The composite map of the complete structure [5] and the

[2] See Bibliography, Buerger (1959).

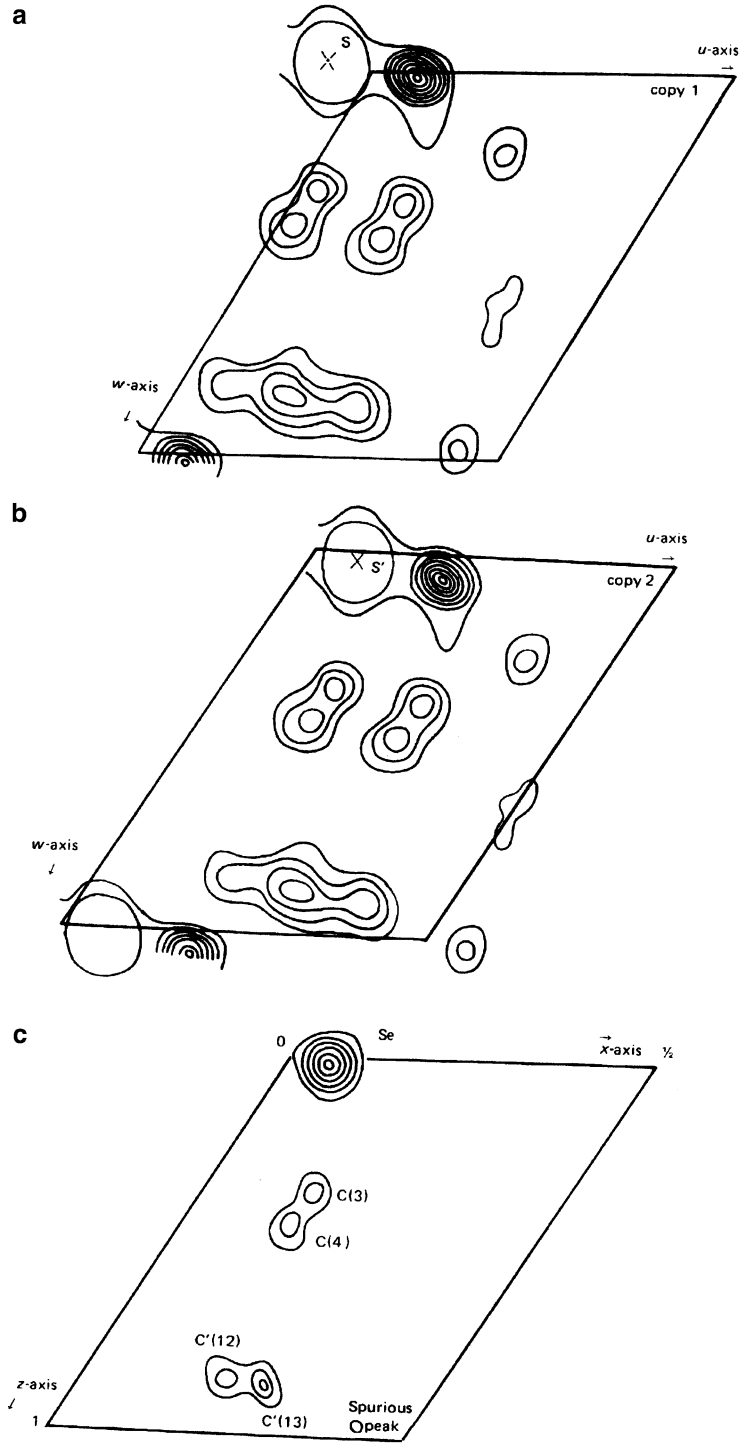


Fig. 7.14 Bisdiphenylmethyldiselenide. (**a**, **b**) Shifted copies 1 and 2 prepared from the ($u0w$) section. (**c**) Minimum-function M_2 section at $y = 0.0$; C′(12) and C′(13) are symmetry-related to C(12) and C(13) in Fig. 7.15

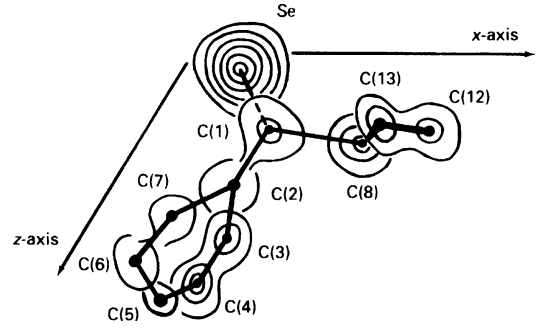


Fig. 7.15 Composite map of the three-dimensional minimum function $M_2(xyz)$ for bisdiphenylmethyldiselenide

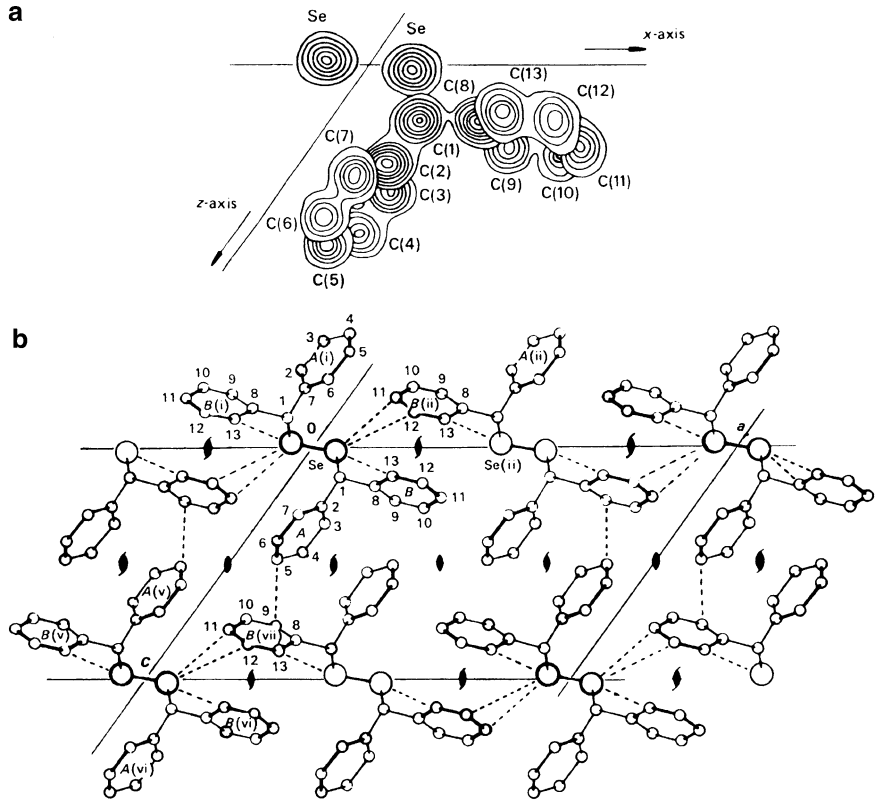


Fig. 7.16 Bisdiphenylmethyldiselenide: (**a**) Composite electron density map as seen along *b*; (**b**) crystal structure as seen along *b*; the *dashed lines* indicate the closest intermolecular contacts

packing of the molecules in the unit cell are shown in Fig. 7.16. In favorable circumstances, the Patterson function can be solved for the majority of the heavier atoms in the crystal structure. The atoms not located by M_2 in this example were obtained from an electron density map phased on those atoms that were found, a standard method for attempting to complete a partial structure, see Sect. 7.5.

7.4.8 Determination of the Chlorine Atom Positions in Papaverine Hydrochloride [1]

The crystal data for this compound have been given in Sect. 7.2.1. The calculated origin peak height is approximately 4700, and a single-weight Cl–Cl vector would have a height of about 6% of that of the origin peak. The Cl–Cl vector may not be located as easily as that of Se–Se in the previous example. The general equivalent positions in $P2_1/c$ give rise to the vectors shown in Table 7.2. The assignment of coordinates to the chlorine atoms follows the recognition of peaks A, B, and C as Cl–Cl vectors on the Patterson maps, Fig. 7.17a–c. Figure 7.17d is the sharpened section, $(u, \frac{1}{2}, w)$. The steps in the solution of the problem are set out in Table 7.3.

The results are completely self-consistent, and we may list the Cl coordinates in the unit cell:

$$\begin{aligned} 4\mathrm{Cl}: \ & 0.025, \quad 0.169, \quad 0.038; \quad 0.025, \quad 0.331, \quad 0.538 \\ & -0.025, -0.169, -0.038; -0.025, -0.331, -0.538 \end{aligned}$$

For simplicity, peak A was assigned as $-(\frac{1}{2}+2y)$, which is crystallographically the same as $\frac{1}{2}-2y$, in order to obtain $y \leq \frac{1}{2}$. For a similar reason, B was retained as $\frac{1}{2}+2z$.

The specification of the peak parameters in this manner is, to some extent, dependent on the observer. A different choice, for example, $\frac{1}{2}+2y$ in A, merely results in a set of atomic positions located with respect to one of the other centers of symmetry as origin. In space groups where the origin might be defined with respect to other symmetry elements, similar arbitrary peak specifications may be possible.

7.4.9 Determination of the Mercury Atom Positions in KHg_2

This example illustrates the application of the Patterson function to the determination of the coordinates of atoms in special positions of space group *Imma*.

Crystal Data[3]
System: orthorhombic
Unit-cell dimensions: $a = 8.10$, $b = 5.16$, $c = 8.77$ Å
V_c: 366.55 Å^3
D_m: 7.95 g cm^{-3}
M_r: 440.28

Table 7.2 Patterson peaks in space group $P2_1/c$

Label	Vector	Peak strength	Harker region
A	$\pm\{0, \frac{1}{2}+2y, \frac{1}{2}\}$	Double weight	Line: $(0, v, \frac{1}{2})$
B	$\pm\{2x, \frac{1}{2}, \frac{1}{2}+2z\}$	Double weight	Section: $(u, \frac{1}{2}, w)$
C	$\pm\{2x, 2y, 2z\}$	Single weight	General region
D	$\pm\{2x, 2\bar{y}, 2x\}$	Single weight	General region

[3] In the original paper, the origin in *Imma* was chosen on a center of symmetry displaced by $\frac{1}{4}, \frac{1}{4}, \frac{1}{4}$ from this origin.

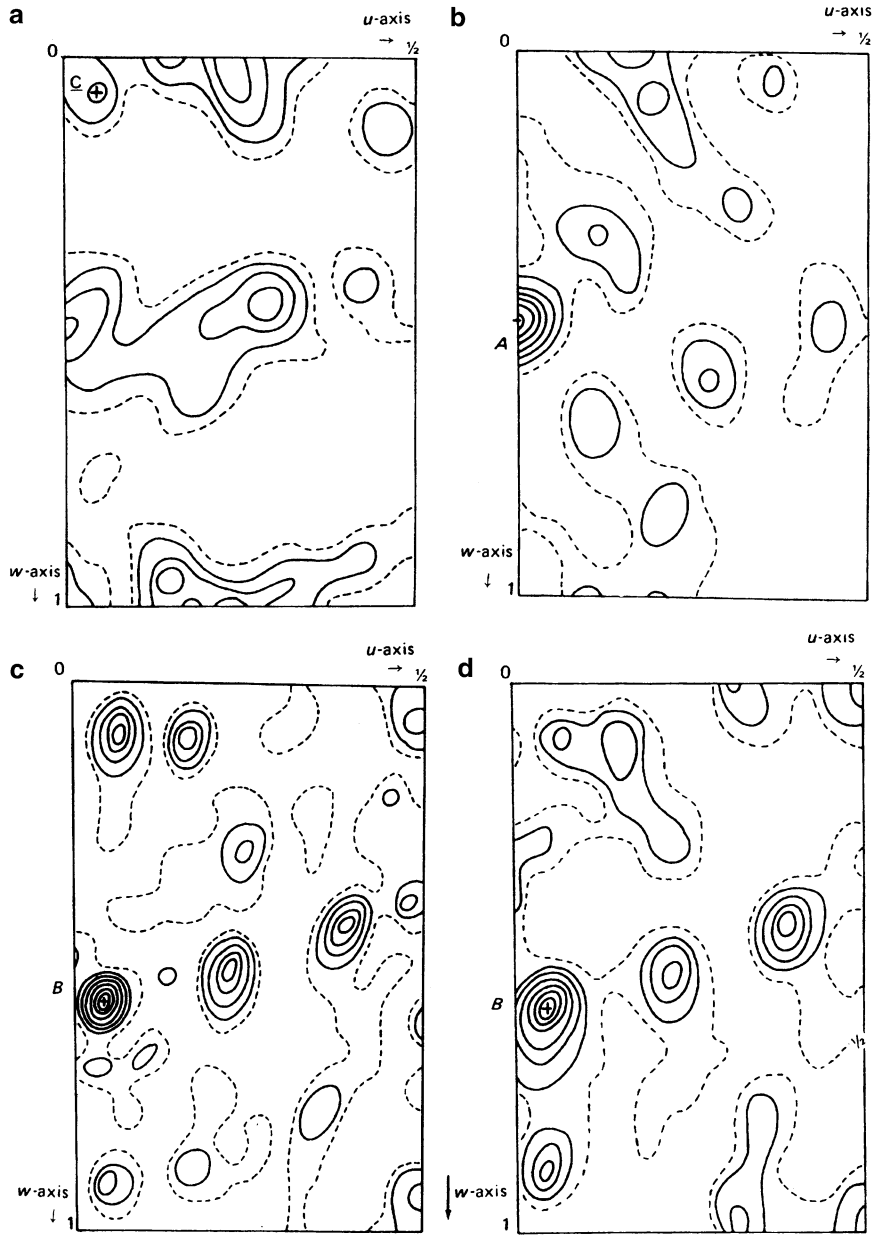


Fig. 7.17 Three-dimensional Patterson sections for papaverine hydrochloride; the Cl–Cl vectors are labeled A, B, and C. (**a**) $v = 8.4/52$, (**b**) $v = \frac{1}{2}$. (**c**) $v = 17.6/52$. (**d**) $v = \frac{1}{2}$ (sharpened section)

Table 7.3 Heavy-atom coordinates for papaverine hydrochloride

Patterson map	Label	Vector coordinates[a]	Cl coordinates
Figure 7.17a, level $v = 8.4/52$	A	$\frac{1}{2} - 2y = 8.4/52$	$y = 0.169$
Figure 7.17b, level $v = \frac{1}{2}$	B	$2x = 2.2/44$ $\frac{1}{2} + 2z = 17.3/30$	$x = 0.025$ $z = 0.038$
Figure 7.17c, level $v = 17.6/52$	C	$2x = 2.2/44$ $2y = 17.6/52$ $2z = 2.3/30$	$x = 0.025$ $y = 0.169$ $z = 0.038$

[a]The Patterson synthesis was computed with the intervals of subdivision 44, 52, and 30 and along u, v, and w, respectively

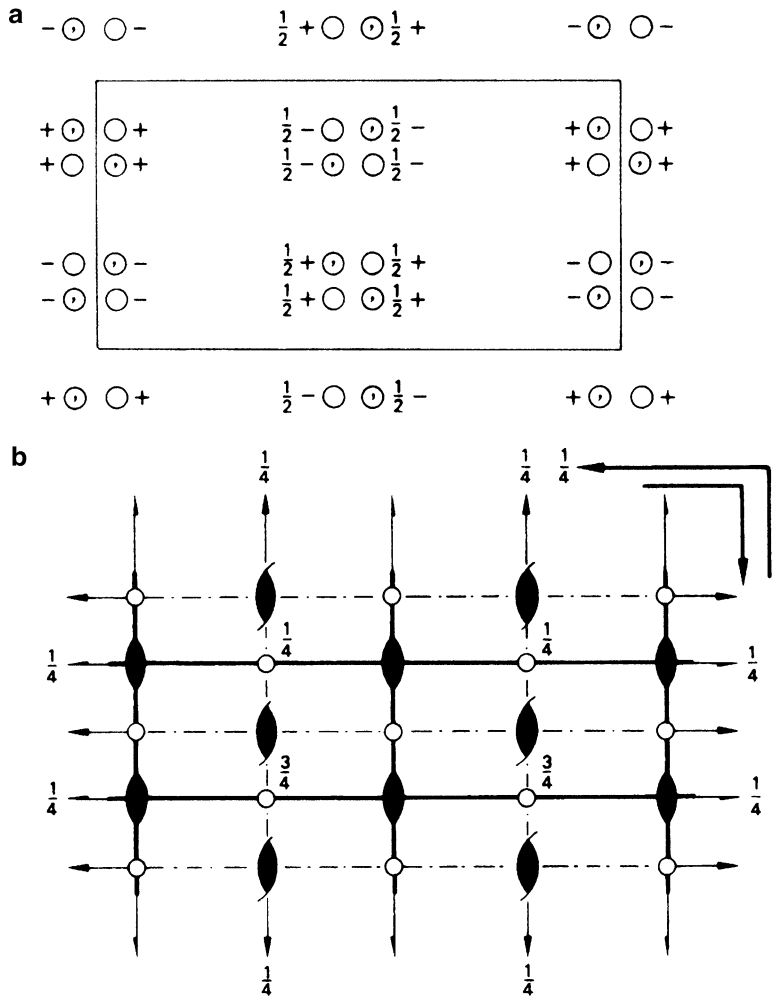


Fig. 7.18 Space group *Imma* (rotated by 90° from the standard setting): (**a**) general equivalent positions, (**b**) symmetry elements

Z: 3.99 or 4 to the nearest integer
Unit-cell contents: 4K and 8Hg atom
Absent spectra: *hkl*: $h + k + l = 2n + 1$; *hk*0: $h = 2n + 1$ ($k = 2n + 1$)

From the diffraction data, possible space groups are *Im*2*a*, *I*2*ma*, or *Imma*. In the absence of further information on the space group, we shall proceed with the analysis in *Imma*, Fig. 7.18a, b; the reader may like to consider how easily this figure may be derived from *Pmma*, origin on $\bar{1}$, + *I*.

Symmetry and Packing Analyses

Since *Z* is 4 and there are 16 general equivalent positions in *Imma*, all atoms must lie in special positions. Table 7.4 lists these positions for this space group, with a center of symmetry (2/*m*) as origin [6].

This list presents a quite formidable number of alternatives for examination. The eight Hg atoms could lie in (f), (g), (h), or (i). However, further consideration of sets (f), (g), and (i) and sets (c) and (d) shows that they would all involve pairs of Hg atoms being separated by distances less than *b*/2 (2.58 Å). This value is much shorter than known Hg–Hg bond distances in other structures, and we shall reject these sets. The positions in these sets may be plotted to scale in order to verify the spatial limitations.

Table 7.4 Special positions in *Imma*

4	(a)	2/*m*	$0,0,0;\quad \frac{1}{2},0,0;\quad \frac{1}{2},\frac{1}{2},\frac{1}{2};\quad 0,\frac{1}{2},\frac{1}{2}$
4	(b)	2/*m*	$0,\frac{1}{2},0;\quad \frac{1}{2},\frac{1}{2},0;\quad \frac{1}{2},0,\frac{1}{2};\quad 0,0,\frac{1}{2}$
4	(c)	2/*m*	$\frac{1}{4},\frac{1}{4},\frac{1}{4};\quad \frac{1}{4},\frac{3}{4},\frac{1}{4};\quad \frac{3}{4},\frac{3}{4},\frac{3}{4};\quad \frac{3}{4},\frac{1}{4},\frac{3}{4}$
4	(d)	2/*m*	$\frac{1}{4},\frac{3}{4},\frac{3}{4};\quad \frac{1}{4},\frac{1}{4},\frac{3}{4};\quad \frac{3}{4},\frac{1}{4},\frac{1}{4};\quad \frac{3}{4},\frac{3}{4},\frac{1}{4}$
4	(e)	*mm*2	$\frac{1}{4},0,z;\quad \frac{3}{4},0,\bar{z};\quad \frac{3}{4},\frac{1}{2},\frac{1}{2}+z;\quad \frac{1}{4},\frac{1}{2},\frac{1}{2}-z$
8	(f)	2	$\pm\left\{0,y,0;\quad \frac{1}{2},y,0;\quad \frac{1}{2},\frac{1}{2}+y,\frac{1}{2};\quad 0,\frac{1}{2}+y,\frac{1}{2}\right\}$
8	(g)	2	$\pm\left\{x,\frac{1}{4},\frac{1}{4};\quad x,\frac{3}{4},\frac{1}{4};\quad \frac{1}{2}+x,\frac{3}{4},\frac{3}{4};\quad \frac{1}{2}+x,\frac{1}{4},\frac{3}{4}\right\}$
8	(h)	*m*	$\pm\left\{x,0,z;\quad \frac{1}{2}-x,0,z;\quad \frac{1}{2}+x,\frac{1}{2},\frac{1}{2}+z;\quad \bar{x},\frac{1}{2},\frac{1}{2}+z\right\}$
8	(i)	*m*	$\pm\left\{\frac{1}{4},y,z;\quad \frac{1}{4},\bar{y},z;\quad \frac{3}{4},\frac{1}{2}+y,\frac{1}{2}+z;\quad \frac{3}{4},\frac{1}{2}-y,\frac{1}{2}+z\right\}$

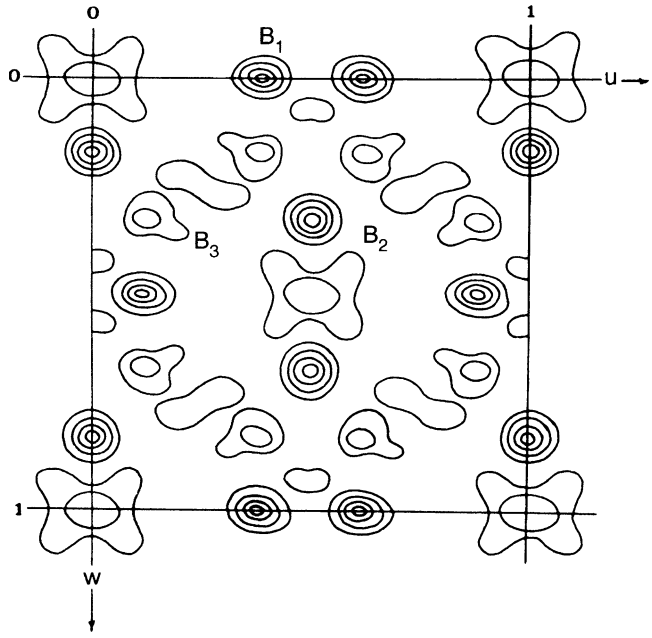


Fig. 7.19 Patterson projection $P(uw)$ for KHg_2; the origin peak has not been contoured and the labeled peaks are Hg–Hg vectors

Of the remaining sets, (a) and (b) together would again place neighboring Hg atoms too close to one another. There are three likely models:
Model I: four Hg in (a) + four Hg in (e).
Model II: four Hg in (b) + four Hg in (e).
Model III: eight Hg in (h).

The Patterson function enables us to differentiate among these alternative models.

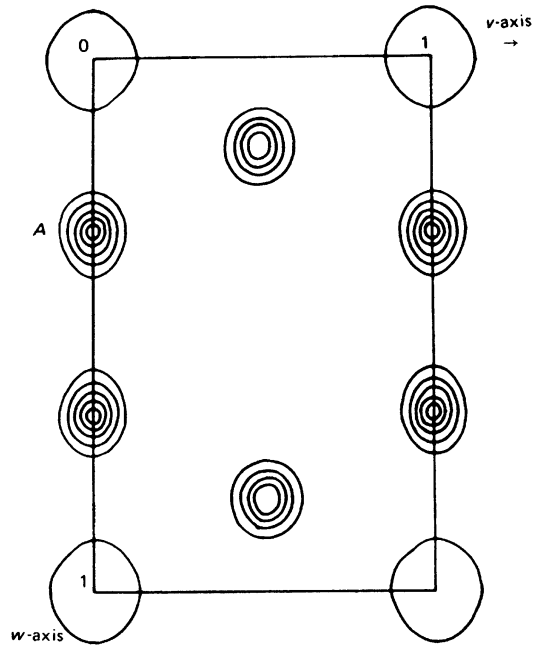


Fig. 7.20 Patterson projection $P(vw)$ for KHg_2; the origin peak has not been contoured

Vector Analysis of the Alternative Hg Positions

Model I would produce, among others, an Hg–Hg vector at $u = \frac{1}{2}$, $w = 0$, from the atoms in set (a). The b axis Patterson projection, Fig. 7.19, shows no peak at that position, and we eliminate model I. For a similar reason, with the atoms of set (b), model II is rejected. It is necessary to show next that model III is consistent with the Patterson function. The a-axis projection is shown in Fig. 7.20.

Interpretation of P(uw)

In this projection, no reference is made to the y coordinates, and we look for vectors of the type $\pm\{\frac{1}{2} + 2x, 0\}$ and $\pm\{\frac{1}{2}, 2z\}$, and four vectors related by $2mm$ symmetry $\pm\{2x, 2z\}$ and $\pm\{2\bar{x}, 2z\}$. The double-weight peak labeled $B(1)$ is on the line $w = 0$, and $B(2)$ is on the line $u = \frac{1}{2}$. Hence, $x_{Hg} = 0.064$ and $z_{Hg} = 0.161$. These values are corroborated by measurements from the single weight peak $B(3)$.

Interpretation of P(vw)

Vectors like A, Fig. 7.20, are of the type $\pm\{0, 2z\}$. We deduce $z_{Hg} = 0.161$, in excellent agreement with the value obtained from the b-axis projection.

Superposition techniques applied to the a-axis projection indicate that the K atoms are in special positions (b), but this result is not supported by the b-axis projection. Evidently, the Patterson results can give only a partial structure, and supplementary methods are needed to carry the analysis to completion. In summary, we have determined the positions of the mercury atoms to be in set (h),[4] Table 7.4, with $x = 0.064$, $z = 0.161$.

[4] In the work of Duwell and Baenziger [6], the positions listed are 8 (i), with $x = 0.186$ and $z = 0.089$, each being $\frac{1}{4}$ *minus* the value given here.

7.5 Heavy-Atom Method and Partial Fourier Synthesis

The heavy-atom method was conceived originally as a method for determining the positions of light atoms in a structure containing a relatively small number of heavier atoms. However, the technique can be applied to most situations where a partial structure analysis has been effected, provided that certain conditions are met.

Imagine a situation where N_k, of the N atoms in a unit cell have been located; N_k may be only one atom, if it is a heavy atom. There will be N_u atoms remaining to be located, and we may express the structure factor in terms of known (k) and unknown (u) atoms:

$$\mathrm{F}(hkl) = \sum_{j=1}^{N_k} g_j \exp[\mathrm{i}2\pi(hx_j + ky_j + lz_j)] + \sum_{u=1}^{N_u} g_u \exp[\mathrm{i}2\pi(hx_u + ky_u + lz_u)] \tag{7.22}$$

or

$$|\mathrm{F}(hkl)| = |\mathrm{F_c}(hkl)| + |\mathrm{F_u}(hkl)| \tag{7.23}$$

In practice, F_o data, appropriately scaled, replace $|\mathrm{F}(hkl)|$, and $\mathrm{F_c}(hkl)$ refers to the known (N_k) atomic positions. As more of the structure becomes known, the values of $|\mathrm{F_c}(hkl)|$ approach $F_\mathrm{o}(hkl)$ and the phase angle ϕ_c approaches the unobservable but required value $\phi(hkl)$. Figure 7.21 illustrates this argument for any given reflection. The values of ϕ_c may provide sufficiently reasonable approximations to $\phi(hkl)$ for an electron density map to be calculated with some confidence. The nearer $|\mathrm{F_c}|$ is to F_o, the better the values of the phase angles, and this is clearly dependent upon the percentage of the scattering power which is known. As a guide to the effective phasing power of a partial structure, the quantity r may be calculated:

$$r = \sum_{j=1}^{N_k} Z_j^2 \Big/ \sum_{u=1}^{N_u} Z_u^2 \tag{7.24}$$

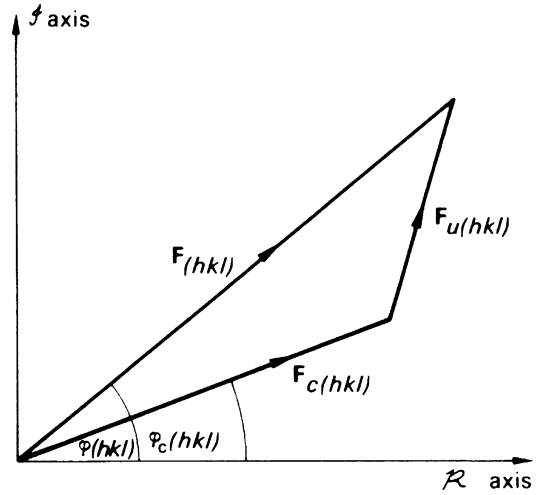


Fig. 7.21 Partial-structure phasing; F(hkl) is the true structure factor of modulus $F_\mathrm{o}(hkl)$ and phase $\phi(hkl)$

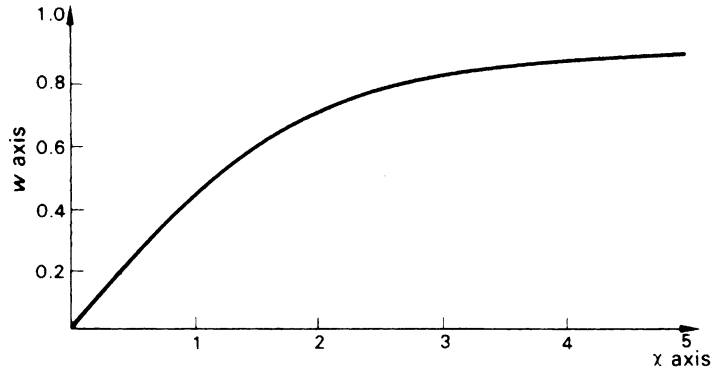


Fig. 7.22 Weighting factor $w(hkl)$ as a function of χ in non-centrosymmetric crystals

where Z refers to the atomic number of a species. A value of r near unity is considered to provide a useful basis for application of the heavy-atom method. However, values of r quite different from unity have produced successful results, because for a given reflection the important quantity is really r', the scattering ratio, given by

$$r' = \sum_{j=1}^{N_k} g_j^2 / \sum_{u=1}^{N_u} g_u^2 \tag{7.25}$$

If r is large, however, the heavy-atom contributions tend to swamp those from the lighter atoms, which may then not be located very precisely from electron density maps. On the other hand, if r is small, the calculated phases may deviate widely from the desired values, and the resulting electron density map could be very difficult to interpret. These extreme situations are found in two of the structures just studied, bisdiphenylmethyldiselenide ($r = 2.4$) and papaverine hydrochloride ($r = 0.28$), based, in each case, on the heavy atoms alone in N_k.

The underlying philosophy of the heavy-atom method depends on the acceptance of calculated phases, even if they contain errors, for the computation of the electron density synthesis. Large phase errors give rise to high background features, which mask the image of the correct structure. The calculated phases ϕ_c contain errors arising from inadequacies in the model, but the F_o data, although subject to experimental errors, hold information on the complete structure. Phase errors may be counteracted to some extent by weighting the Fourier coefficients according to the degree of confidence in a particular phase. For centrosymmetric structures, the weight $w(hkl)$ by which $F_o(hkl)$ is multiplied is given by [7]

$$w(hkl) = \tanh(\chi/2) \tag{7.26}$$

where χ is given by

$$\chi = 2F_o|F_c| / \sum g_u^2 \tag{7.27}$$

The subscripts c and u refer, respectively, to the known and unknown parts of the structure. In non-centrosymmetric structures, $w(hkl)$ can be obtained from the graph [8] in Fig. 7.22. Weighting factors should be applied to F_o values that have been placed on an absolute, or approximately absolute, scale.

We can show, in a simplified manner, how one may reasonably expect the heavy-atom procedure to be successful. In a centrosymmetric structure, the two terms on the right-hand side of (7.22) would

be cosine expressions. The sum over the N_k atoms would have a magnitude $M1$ and be either + or − in sign. Similarly, the sum over the N_u atoms would have a magnitude $M2$ together with a + or a − sign. Over a number of reflections, we may say that there is a 50% chance that the true signs are those given by the heavy atom. For the other 50%, there is a 25% chance that $M1 > M2$, so that again the sign given by the heavy atom is correct. Thus, there is a good chance that a large percentage of the reflections will be given the correct sign in a favorable heavy-atom application.

Bearing all these points in mind, it follows that the best electron density map one can calculate with phases determined from a partial structure is given by

$$\rho(xyz) = \frac{2}{V_c} \sum_{h=0} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty} w(hkl) F_o(hkl) \cos[2\pi(hx + ky + lz) - \phi_c(hkl)] \tag{7.28}$$

where

$$\phi_c(hkl) = \tan^{-1}[B'_c(hkl)/A'_c(hkl)] \tag{7.29}$$

and $A'_c(hkl)$ and $B'_c(hkl)$ are the real and imaginary components, respectively, of the calculated structure factor, which is included in the right-hand side of (7.23).

Electron density maps calculated from partial-structure phasing contain features which characterize both the true structure and the partial, or trial, structure. We have considered this situation in Sect. 6.9.1. Now, we may let each observation in (6.75), be multiplied by the weight w so as to give a better statistical significance to each term in the calculation.

If the model includes atoms in reasonably accurate positions, we can expect two important features in the electron density map: (a) atoms of the trial structure should appear, possibly in corrected positions, and (b) additional atoms should be revealed by the presence of peaks in stereochemically sensible positions.

If neither of these features is observed in the electron density synthesis, it may be concluded that the trial structure contains very serious errors, and we would be on a false trail. Correspondingly, there would be poor agreement in the *pattern* of relationship between F_o and $|F_c|$.

7.5.1 Reliability Factor

Introduction

The use of different R factors is a popular way of quantifying various procedures at different stages of an analysis. While the calculation of each R factor is not necessarily based on rigorous principles, it is a simple and convenient way to record a quantitative assessment of quality as an aid to subsequent decision making. In essence each R factor measures an agreement in terms of the ratio $R = \Delta/M$, where Δ is the mean difference between measurements that are expected to have the same value within experimental error and M is the mean value of the set of measured quantities. These averages are usually derived from a large number of measurements in a typical crystal structure analysis. As an example, we consider intensity data recorded from a crystal having a twofold symmetry axis parallel to the y axis. Ideally, $I(hkl) = I(\bar{h}\,k\,\bar{l})$ so that $\Delta I = |I(hkl) - I(\bar{h}\,k\,\bar{l})|$ may be used to calculate an R factor called R_{sym}.

If the intensity measurements do actually correspond according to twofold symmetry, R_{sym} calculated in this way will be very small, because each ΔI term will tend to be zero within experimental error. An unacceptably high value of R_{sym} could indicate that either there is not

actually a twofold axis parallel to ***y***, or the crystal quality is poor. A solution to this problem would be to repeat data measurements with a fresh crystal with subsequent reassessment of the situation. In the case of a macromolecular crystal, it should be remembered that diffracted intensities tend to much more accurate values at low resolution than at high resolution where the actual intensity values become much smaller owing to the effect of disorder and high thermal factors. For this reason, it is common to calculate R_{sym} as a function of resolution by partitioning the data into resolution "shells" or "bins."

A quantity very similar to R_{sym}, known as R_{merge}, may be calculated as a by-product of data collection: R_{sym} and R_{merge} are based on different sets of data and should not be treated as being equivalent. As a practical warning it should be remembered that neither R_{sym} nor R_{merge}, as with other R indices, can stand up to rigorous analysis, and should be treated only as guides to data quality and aides to decision making. As a rule of thumb for high-resolution small-molecule structures R_{sym} may be as low as 2% and more usually 4–6%. For low resolution protein data R_{sym} would rarely fall below 5%, and may rise to about 20% for high resolution data. A worse result than this would indicate poor data quality, assuming that a sufficient sweep of intensity data has been recorded. In such cases it may be wise either to exercise caution and discard some of the data by using a resolution cut-off or to remeasure the data with a fresh crystal.

R_{sym}

If separate intensity measurements of $I(hkl)$ and its symmetry-related equivalents have been made, each measurement may be expressed as $I_i(hkl)$, $(i = 1, 2, \ldots, n)$ where n is greater than 2 but has possibly different values for different reflections, with a maximum value equal to the total number of symmetry equivalents for the space group in question. The best estimate of a given $I(hkl)$ is $\bar{I}(hkl) = \sum I_i(hkl)/n$. The discrepancy between the ith observation of $I(hkl)$ and the best estimate is $I_i(hkl) - \bar{I}(hkl)$; and the mean discrepancy of all the observations of this intensity is $\sum_n (I_i(hkl) - \bar{I}(hkl))/n$. In a given group of intensity measurements, let there be N for which at least two symmetry equivalents have been measured. To form R_{sym} the mean discrepancy for each of these N reflections is compared to their mean intensity:

$$R_{\text{sym}} = \sum_N \left[\sum_n |I_i(hkl) - \bar{I}(hkl)|/n \right] \Bigg/ \sum_N \bar{I}(hkl)$$

Weights derived from the data collection statistics may be employed in order to make a better estimate of $\bar{I}(hkl)$ as:

$$\bar{I}(hkl) = \sum_{i=1,n} \omega_i I_i(hkl) \Bigg/ \sum_{i=1,n} w_i$$

R_{sym} can be thought of as the mean error of an intensity measurement, compared to the mean intensity, averaged over the chosen group of reflections.

We draw attention to the facts (1) that R_{sym} as defined here is calculated on the basis of intensity measurements, (2) that some programs, such as SHELX-97, use the notation R_{int} instead of R_{sym} and employ $|\text{F}|^2(hkl)$ instead of $I(hkl)$, and (3) that since $I(hkl)$ is related to $|\text{F}|^2(hkl)$, the fractional error in $I(hkl)$ is approximately double the fractional error in $|\text{F}|(hkl)$ so that if R_{sym} is 0.04, the mean fractional error in $|\text{F}|$ would be about 0.02. This sort of value would usually be considered to be "highly acceptable."

$R_{\mathbf{merge}}$

The essential difference between R_{merge} and R_{sym} lies in the data subsets from which they are usually calculated: R_{merge} is calculated, not from symmetry-equivalent intensities, but from *repeats of the same* $I(hkl)$ measured during the course of a data collection session. Such repeat measurements may be the result of employing overlapping angular ranges when the crystal is rotated. The program SHELX-97 uses the notation R_{sigma} instead of R_{merge}.

It is possible to merge the repeat measurements to produce a new basic data set prior to merging the symmetry equivalents. A value of R_{sym} could then be derived from this new basic set and the symmetry equivalents combined to produce the final working data set. A possible scheme is set out as follows.

Intensity Data Collection and Processing Scheme

Data Set 1

Data are collected to the maximum possible θ for which intensities are strong enough to measure satisfactorily. This is the basic data set of non-merged $I(hkl)$ data, and may include repeats:

- For some unique $I(hkl)$ data whose values should be equal within experimental error;
- For a number of space group symmetry-equivalent $I(hkl)$ data whose values should also be equal within experimental error.

Data Set 2

Data processing can be performed first by merging the unique repeats in data set 1. R_{merge} is thus produced and examined for quality; if acceptable, it defines data set 2 which is then further processed.

Data Set 3

Data set 2 is next used to produce the final data set 3 in which the space group symmetry-equivalent $I(hkl)$ data have now been combined, and R_{sym} is produced as a result of this process. If R_{sym} has an acceptable value, data set 3 becomes the current standard data set for the analysis.

Data Collection from Macromolecules

If R_{merge} from data set 2 is not acceptable, poor crystal quality may be indicated. The data collection should be repeated with a new crystal until a good specimen is found.

If R_{sym} from data set 3 is not acceptable, it could indicate an incorrect assignment of the space group. Then, other possible space groups should be investigated by further analysis of the data and recalculation of R_{sym} until a satisfactory result is obtained. If it is still unsuccessful, it may be necessary to recollect data from fresh crystals.

Anomalous Scattering

Small Molecules: Calculation of R_{sym} for Non-centrosymmetric Structures

In small-molecule analysis for structures in non-centrosymmetric space groups, it is possible to determine the absolute configuration of the structure by virtue of anomalous differences between Friedel pairs $I(hkl)$ and $I(\overline{hkl})$. The success of this operation depends on the types of atom present and the radiation used for data collection. Thus, it is usual in non-centrosymmetric small-molecule structure analysis not to merge Friedel opposites or any opposites generated by the space group symmetry (Bijvoet differences, q.v.) unless it can be shown from calculation of the Flack parameter (Sect. 7.6.1) that the absolute configuration cannot be determined with certainty. If and when this stage is reached, the

refinement should be repeated and completed with a data set for which the Friedel opposites and Bijvoet pairs have been merged.

R_{anom} in Macromolecular Analysis

With macromolecular structures that are also non-centrosymmetric, the application of anomalous scattering for phasing in structure analysis also requires similar anomalous measurable differences to exist. Again, this would preclude the inclusion of such reflections from the calculation of R_{sym}. In order to assess the possible strength of anomalous scattering, an R index is initially calculated as $R_{\text{anom}} = \Delta/M$ where Δ is now the mean Friedel or Bijvoet difference and M is the mean of all amplitudes involved.

We know that Bijvoet amplitude differences are equivalent by space group symmetry to Friedel opposite differences. In a given data collection experiment, one or both quantities may be measured. In practice, it is possible to derive useful phase indications from anomalous differences for a macromolecular crystal which includes one or more strong anomalous scatters in relation to the radiation employed provided that $2R_{\text{anom}} \gg R_{\text{merge}}$. The factor 2 derives from the fact that R_{merge} is based on intensity whereas R_{anom} is based on amplitude.

R_{deriv}

This R factor is used to assess the phasing power of a heavy-atom derivative in macromolecular structure analysis by the isomorphous replacement method:

$$R_{\text{deriv}} = \sum ||\mathrm{F}_{PH}| - |\mathrm{F}_P|| \Big/ \sum |\mathrm{F}_P|$$

The summations here are over the N-independent reflections for which both $|\mathrm{F}_P|$ and $|\mathrm{F}_{PH}|$ have been measured. It is assumed that $|\mathrm{F}_P|$ and $|\mathrm{F}_{PH}|$ have both been scaled, for example, by Wilson statistics (Sect. 4.2.1). For a moderately sized protein with one fully occupied heavy atom site, the value of R_{deriv} calculated for the data between 10 and 3.5 Å resolution would be expected to have a value between 0.1 and 0.2. Larger changes than this may indicate several sites of heavy atom substitution or may be due to lack of isomorphism. If the latter is the case, the heavy atom derivative will not provide any useful phasing and would be discarded; R_{deriv} is also known as R_{diff} and R_{iso}.

Reliability Factor in Structure Refinement

The conventional R factor that is used in structure refinement assessment reflects the difference between the *scaled*-observed and the calculated structure-factor amplitudes, and provides a measure of the quality of the trial structure. Large differences correspond to poor reliability, and vice versa. The R is defined as

$$R = \sum_{hkl} |KF_{\text{o}} - |\mathrm{F}_{\text{c}}|| \Big/ \sum_{hkl} KF_{\text{o}} \tag{7.30}$$

R is sometimes also known as R_1, where R_2 is used for the corresponding weighted value. For a well-refined structure model, the value of R approaches a small value, about 1% at the very best, corresponding to the residual errors in both the experimental data and the model. In the early stages of the analysis, however, R may lie between 0.4 and 0.5. It expresses the first criterion of correctness, namely, good agreement between F_{o} and $|\mathrm{F}_{\text{c}}|$. It should be noted that trial structures with an R factor of more than 50% have been known to be capable of refinement; R is only a rough guide at that stage of the analysis. A better basis for judgment is a comparison of the *pattern* of F_{o} and $|\mathrm{F}_{\text{c}}|$, which requires care and experience.

Refinement *R* Indices and Weights with SHELX-97

Structure refinement with the SHELX-97 system is carried out against $|\mathrm{F}|^2$ instead of $|\mathrm{F}|$ and the values of the R indices tend to be more than twice as high. For comparison with older refinements, SHELX-97 produces a conventional index R_1 based on F_o values $>4\sigma(F_\mathrm{o})$ and a weighted R factor wR_2 based on $|\mathrm{F}|^2$ where R_1 is governed by (7.30) and the weighted R factor by

$$wR_2 = \left\{ \sum \left[w({F_\mathrm{o}}^2 - |\mathrm{F_c}|^2)^2 \right] \Big/ \sum \left[w({F_\mathrm{o}}^2)^2 \right] \right\}^{1/2}$$

where F_o is a scaled value, as in (7.30), and $w = 1/[\sigma^2(F_\mathrm{o}^2) + (aP)^2 + bP]$ where $P = [2|F_\mathrm{c}|^2 + \max({F_\mathrm{o}}^2, 0)]/3$. The use of this combination of F_o^2 and $|\mathrm{F_c}|^2$ was shown [9] to reduce statistical bias. As a rough guide, for high resolution small-molecule structures R_1 may be as low as 2% but more usually 4–6%. For low resolution protein data R_1 would rarely fall below 5%, and may rise to 15–20% at high resolution. A result less than this would indicate poor data quality and require careful investigation.

R_{free} and Protein Refinement

The R_{free} index [10] is currently used in macromolecular structure analysis, and this is a requirement in most journals as a condition of publication. Use of this technique requires that the measured $F_\mathrm{o}(hkl)$ data available for refinement of the protein structure is partitioned into two separate sets. The smaller portion of the data is usually selected at random and comprises 5–10% of the whole data set. The small subset is used in the calculation of R_{free} and the larger subset is known as the "working" data set. The R_{free} subset of data is used only for calculation of R_{free}. These reflections are subsequently excluded from the summation in the equations for R (R_1) given above, which is calculated only from the working data set. Consequently there is a permanent loss of this selected data from the refinement procedure. This test subset is never allowed to be included in the refinement process from start to finish. The purpose of excluding this subset of data is to enable the calculation of

$$R_{\mathrm{free}} = \sum_{hkl} |KF_\mathrm{o} - |\mathrm{F_c}|| \Big/ \sum_{hkl} KF_\mathrm{o}$$

where the summations are carried out over the randomly selected 5–10% subset of test data, and K is the scale factor for the data set.

Since these randomly selected data are not available at any stage in the refinement procedure, any improvements such as a decrease in R_{free} during the course of the refinement will reflect genuine improvements in the model, rather than the fitting of an incorrect model to the F_o data which could, within limits, improve the conventional R factor. After each cycle of refinement calculations, it is essential to review the refined model for errors by studying in detail the $F_\mathrm{o} - |\mathrm{F_c}|$ difference map. If this map has any significant positive features, they are a clear indication that a part of the map needs to be reinterpreted. In practice, during the course of structure refinement R_{free} lags behind R because it is calculated from reflections that the refinement procedure does not "know about." A lag of about 3% is usually considered acceptable. Further guidelines can be found in the documentation for the Phenix program [11].

R_{free} as a Simple Quantitative Guide to Correct Coordinate in a Refined Protein Structure

A method [12] that was originally intended to aid evaluation of the accuracy of well-resolved atomic structures in small-molecule crystallography has been modified [13, 14] for application to refined protein structures. As with the original method, the modification is based largely on the R factor.

It enables the positional error for an atom having an average isotropic temperature factor to be estimated. From (7.30), we see that R will be small if the F_o values have been measured accurately and the $|F_c|$ values are derived from a well behaved and highly reliable structure. A modified approximation [14] provides a rapid method for estimating the precision $\sigma(r_{x,y,z}, \bar{B})$ of the positions of atoms in a refined protein crystal structure:

$$\sigma(r_{x,y,z}, \bar{B}) = 2.2\sqrt{N_{\text{atoms}}} V_a^{1/3} n_o^{-5/6} R_{\text{free}}$$

where N_{atoms} is the number of fully occupied protein sites in the structure, V_a is the volume of the crystal asymmetric unit in Å^3, n_o is the number of $I(hkl)$ measurements in the refinement data set, and R_{free} is its final value after refinement is complete. This expression applies only to protein structures, because the average atomic weight of atoms present in the structure is assumed to be about 14.1. The measure $\sigma(r_{x,y,z}, \bar{B})$ is actually the standard error of position for an atom that has an average B value; it does not include any information about $\sigma(B)$, and the atom positions involved in N are those that are fully occupied. Some of the atoms in a protein structure may be either partially occupied or have larger than average B values. According to this analysis, $\sigma(r_{x,y,z}, \bar{B})$ has the following very approximate expectation values as a function of resolution d_{min} and R_{free}:

d_{min} (Å)	R_{free}	$\sigma(r_{x,y,z}, \bar{B})$ (Å)
1.0	0.25	0.06
2.0	0.25	0.30

The values of R_{free} at this level would be considered worryingly high.

d_{min} (Å)	R_{free}	$\sigma(r_{x,y,z}, \bar{B})$ (Å)
1.0	0.20	0.045
2.0	0.20	0.25
1.0	0.15	0.035
2.0	0.15	0.18

The values of R_{free} within these ranges would generally be considered to be acceptable.

7.5.2 Pseudosymmetry in Electron Density Maps

The electron density map calculated with phases derived from the heavy-atom positions may not exhibit the true space-group symmetry. Suppose space group $P2_1$, for example, has one heavy atom per asymmetric unit. The origin is defined with respect to the x and z axes by the 2_1 axis along $[0, y, 0]$, but the y coordinate of the origin is determined with respect to an arbitrarily assigned y coordinate for *one* of the atoms. Consider the heavy atoms at x, y, z and symmetry-related at $\bar{x}$, $\frac{1}{2} + y$, $\bar{z}$. This arrangement of heavy atoms has the symmetry of $P2_1/m$, with the m planes cutting the y axis at whatever y coordinate is chosen for the heavy atom, say y_H, and at $\frac{1}{2} + y_H$. If y_H is selected as $\frac{1}{4}$, a center of symmetry is at the origin, and the calculated phases will be 0 or π. This situation is illustrated in Fig. 7.23, which indicates that an unscrambling of the images must be carried out. If the heavy atom is given any general value for y_H, $B'(hkl)$ will not be zero and the phase angles will not be 0 or π, but the pseudosymmetry will still exist.

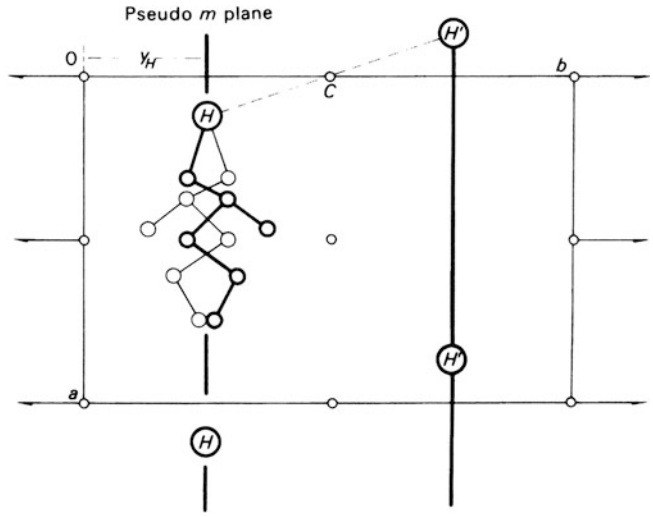


Fig. 7.23 Introduction of pseudosymmetry into space group $P2_1$ by single heavy-atom phasing. H is the heavy atom and C is a center of symmetry introduced between H and its $P2_1$ equivalent H'. The space group (for the heavy atoms alone) thus appears as $P2_1/m$ with mirror (and $\bar{1}$) pseudosymmetry. The electron density map, phased on the H and H' species, will contain two mirror-related images in the asymmetric unit, with a certain degree of confusion between them

7.5.3 Successive Fourier Refinement

A single application of the Fourier method described above does not usually produce a complete set of atomic coordinates. It should lead to the inclusion of more atoms into subsequent structure factor calculations and so to a better electron density map, and so on. This iterative process of Fourier refinement should, after several cycles, result in the identification of all non-hydrogen atoms in the structure to within about 0.1 Å of their true positions. Further improvement of the structure would normally be carried out by the method of least squares, which is described in Sect. 8.4.

7.5.4 Difference-Fourier Synthesis

Some errors present in the trial structure may not be revealed by Fourier synthesis. In particular, the following situations are important.

1. Atoms in completely wrong positions tend to be returned by the Fourier process with similar fractional coordinates, but sometimes with a comparatively low electron density.
2. Correctly placed atoms, may have been assigned either the wrong atomic number, for example, C for N, or an incorrectly estimated temperature factor.
3. Small corrections to the fractional coordinates may be difficult to assess from the Fourier map.

In these circumstances, a difference-Fourier synthesis is valuable. We shall symbolize the Fourier series with F_o coefficients as $\rho_o(xyz)$ and the corresponding synthesis with $|F_c|$ instead as $\rho_c(xyz)$; the difference-Fourier synthesis $\Delta\rho(xyz)$ may be obtained in a single-stage calculation from the equation, using the difference of observed and calculated structure amplitudes:

$$\Delta\rho(xyz) = \frac{2}{V_c} \sum_h \sum_k \sum_l (F_o - |F_c|) \cos[2\pi(hx + ky + lz) - \phi_c] \tag{7.31}$$

Since the phases are substantially correct at this stage, it is in effect, a subtraction, point by point, of the "calculated," or trial, Fourier synthesis from that of the "observed," or experimentally based, synthesis. The difference synthesis has the following useful properties.

1. Incorrectly placed atoms correspond to regions of high electron density in $\rho_c(xyz)$ and low density in $\rho_o(xyz)$; $\Delta\rho(xyz)$ is therefore negative in these regions.
2. A correctly placed atom with either too small an atomic number or too high a temperature factor shows up as a small positive area in $\Delta\rho$. The converse situations produce negative peaks in $\Delta\rho$.
3. An atom requiring a small positional correction tends to lie in a negative area at the side of a small positive peak. The correction is applied by moving the atom into the positive area.
4. Very light atoms, such as hydrogen, may be revealed by a $\Delta\rho$ synthesis when the phases are essentially correct, after least-squares refinement has been carried out.
5. As one final test of the validity of a refined structure, the $\Delta\rho$ synthesis should be effectively featureless within two to three times the standard deviation of the electron density, Sect. 8.7.

7.5.5 Limitations of the Heavy-Atom Method

The Patterson and heavy-atom techniques are effective over a wide range of structures. It is sometimes necessary to introduce heavy atoms artificially into structures. This process may not be desirable because a possible structural interference may arise, and there will be a loss in the accuracy of the light-atom positions. An introduction to direct methods, capable of solving the phase problem for such structures, is given in the next chapter.

7.5.6 Patterson Selection

It is possible that the Patterson function for the crystal of a heavy-atom compound may not reveal the heavy-atom vector unambiguously. Figure 7.24a is the Patterson projection on to the x, z plane for euphenyl iodoacetate, $C_{32}H_{53}O_2I$, and Fig. 7.24b is the Harker section for the same material. There are two high peaks, A and B, in the asymmetric unit where only one was expected. For the Harker section, the coefficients were sharpened and further modified by the multiplicative function $\exp(-9\sin^3\theta)$ so as to smooth out any undesirable fluctuations caused by the sharpening which enhances the high-order reflections relative to those of low order.

In the *selection* process, the F_o^2 data, averaged in zones of $2\sin\theta$, are plotted as a function of $2\sin\theta$ in Fig. 7.24c(i). Between the values for $2\sin\theta$ of 0.5 and 0.7, the average values of F_o^2 are enhanced, owing to the multiplicity of similar distances in the structure, unconnected with the heavy atom, Fig. 1.7, compared with the corresponding smoothed curve, Fig. 7.24c(ii). The many equal, or nearly equal, vectors between atoms in the molecule are superimposed in the Patterson function and lead to additional large peaks.

The F_o^2 data were "selected" by excluding from the next Patterson synthesis all those data lying within the range $0.5<2\sin\theta<0.7$, and the resulting sharpened Harker section is shown in Fig. 7.25; clearly, the heavy-atom vector is at B. The selection process has effectively removed the "structure" that was giving rise to the additional peak A, so that the heavy-atom vector was then sought among the vectors from a more random array of atoms [15].

Figure 7.24c(iii) is the average F_o^2 curve for a random arrangement corresponding to $C_{32}O_2I$, with the same unit-cell geometry as that of euphenyl iodoacetate, and with no two atoms closer than 1.6 Å. The R factor for this hypothetical structure was 0.58; the value for a completely random

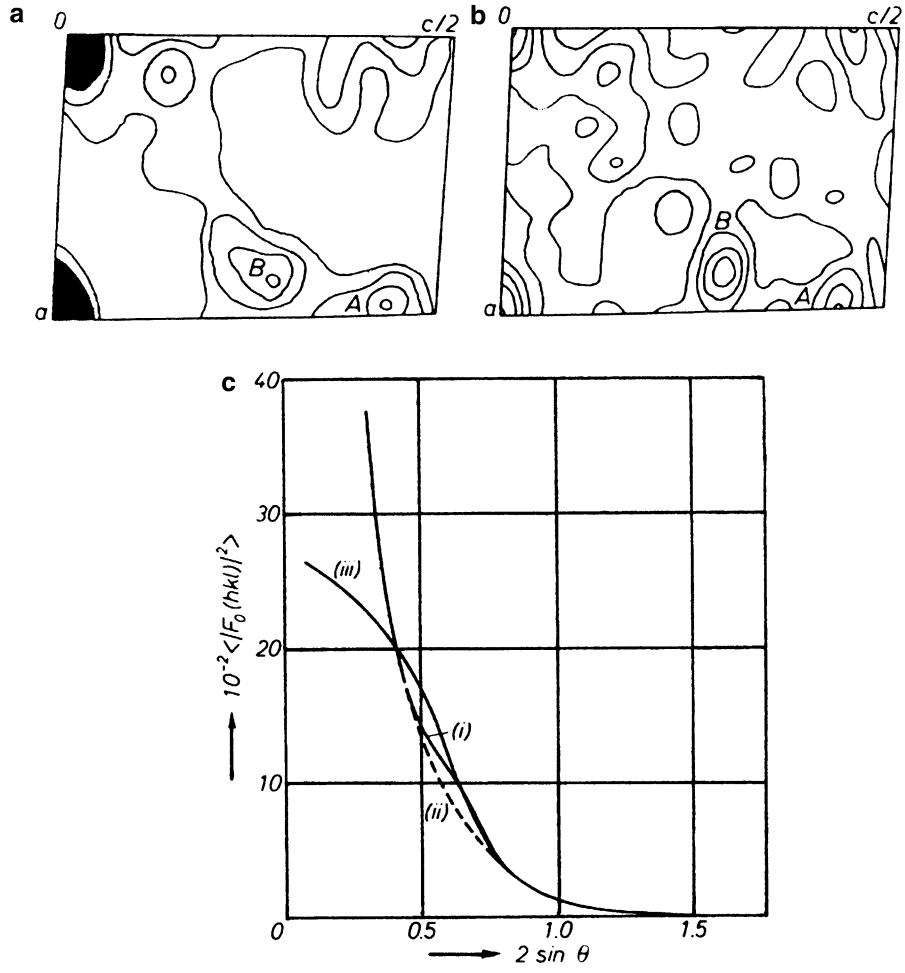


Fig. 7.24 Patterson studies on the heavy-atom compound $C_{32}H_{53}O_2I$ ($P2_1$; $Z = 2$). (**a**) Projection $P(uw)$. (**b**) Sharpened Patterson–Harker section $P(u\frac{1}{2}w)$. (**c**) Plots of average F_o^2 against 2 sin θ: (i) experimental data, (ii) "smooth" curve through the experimental points, (iii) hypothetical (random) structure $C_{32}O_2I$, with the same crystal geometry as $C_{32}H_{53}O_2I$

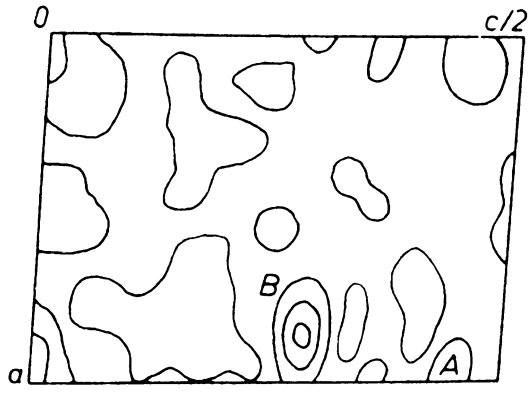


Fig. 7.25 Sharpened Patterson–Harker section $P(u\frac{1}{2}w)$, with the F_o^2 data selected as described in the text. The I–I vector is clearly at B, whereas in Fig. 7.24a, b, there were two peaks of equal height in the asymmetric unit

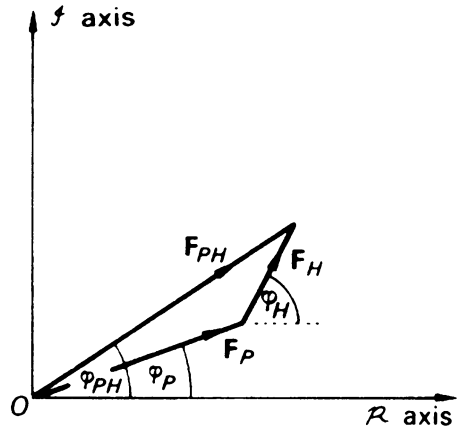


Fig. 7.26 Graphical interpretation of the isomorphous replacement equation. In practice, the phases ϕ_P and ϕ_{PH} are unknown initially. F_H may be known with a fair degree of accuracy if the heavy-atom position in each isomorphous derivative is known. This enables a solution, as illustrated in Fig. 7.27, to be obtained

non-centrosymmetric structure is 0.586 [16]. The curve evolves in a manner similar to an f^2 curve and, as would be expected, shows no structural effects of the nature of Fig. 7.24c(i).

7.5.7 Isomorphous Replacement

A common feature of biologically important substances is their high molecular weight. Proteins and enzymes, for example, are polymers built up from various amino acid residues and form very large atomic assemblies with molecular weights greater than about 5000. The study of the conformations of these giant molecules is necessary for an understanding of their biological functions, and the principal method of obtaining structural detail is by X-ray analysis.

Because of their high molecular weight, protein structures do not yield to analysis by the straightforward heavy-atom method. The value of r, from (7.24), is typically 0.03 for a protein molecule of molecular weight 5000 containing one incorporated mercury atom. This value of r is too small to be useful. Another difficulty is that most proteins and enzymes contain neither very heavy atoms nor easily replaceable groups to facilitate the introduction of heavy atoms. In spite of these difficulties, if a heavy-atom derivative of a large molecule can be prepared, it may be possible to induce it to crystallize in a similar size of unit cell and with the same space group as the native compound. Such pairs of crystals are said to be isomorphous.

The structure factor of the heavy-atom derivative F_{PH} may be handled, as before, in the manner of a vector as

$$F_{PH} = F_P + F_H \tag{7.32}$$

where F_P and F_H are the structure factors for the parent protein and the heavy atoms alone, respectively, for the same reflection. This relationship is shown in Fig. 7.26.

Assuming that the positions of the N_H heavy atoms in the unit cell can be determined, their contribution can be calculated:

$$\mathrm{F}_H = \sum_{j=1}^{N_H} g_j' \exp[\mathrm{i}2\pi(hx_j + ky_j + lz_j)] \tag{7.33}$$

where $g_j' = f_j' \exp(-B_j \sin^2\theta/\lambda^2)$, and $f_j' = K_j f_j$: K_j is a site occupation factor, less than or equal to unity, provided that $|\mathrm{F}_P|$ and $|\mathrm{F}_{PH}|$ are on an absolute scale, and depending on the degree of substitution at the heavy-atom site j; some heavy-atom binding sites of the protein molecules in the crystal may not be fully substituted.

In order to obtain an idea of the effect of a heavy atom on the intensities of X-ray reflections from a protein, we shall carry out a simple calculation for a crystal containing one protein molecule per unit cell in space group $P1$. Assuming that it has a molecular weight of about 13000, about 1000 non-hydrogen atoms would also be present in the molecule; we shall assume that the non-hydrogen atoms are all carbon ($Z_{\mathrm{C}} = 6$). Accepting Wilson's approximation (4.35), and replacing g_j by f_{C}, we have

$$\overline{|\mathrm{F}_P|^2} \approx \sum_{j=1}^{1000} f_{\mathrm{C}}^2 \tag{7.34}$$

At $\sin\theta = 0$, $\overline{|\mathrm{F}_P|^2}$ is 36000. If the derivative contains one mercury atom ($Z_{\mathrm{Hg}} = 80$), then

$$\overline{|\mathrm{F}_{PH}|^2} \approx \sum_{j=1}^{1000} f_C^2 + \sum_{j=1}^{N_H} f_j^2 \tag{7.35}$$

which has the value 42400 at $\sin\theta = 0 (f_j = 80)$. Hence, the maximum change in intensity is about 18%, which is a surprisingly high value.

Experimentally, two sets of data $|\mathrm{F}_P(hkl)|$ and $|\mathrm{F}_{PH}(hkl)|$ are measured and, because of the comparative nature of the phase-determining procedure with isomorphous compounds, they must be placed on the same relative scale, which can be achieved by Wilson's method. Rewriting (7.32), we have

$$|\mathrm{F}_P| \exp(\mathrm{i}\phi_P) = |\mathrm{F}_{PH}| \exp(\mathrm{i}\phi_{PH}) - \mathrm{F}_H \tag{7.36}$$

Assuming that F_H can be determined, this equation involves two unknown quantities, ϕ_P and ϕ_{PH}, and cannot yield a unique solution. However, Fig. 7.27 shows that ideally only two solutions for ϕ_P are real, corresponding to the vectors OP_1 and OP_2, one of which is the true F_P vector. A second isomorphous derivative with a *different set* of heavy-atom positions will also have two solutions for F_P. The derivatives are denoted 1 and 2, and the solutions for F_P are OP_{11} and OP_{12} (derivative 1) and OP_{21} ad OP_{22} (derivative 2), as shown in Fig. 7.28. Two of the solutions should agree with either OP_1 or OP_2 within experimental error, thus resolving the ambiguity, Fig. 7.28. With a more extensive series of isomorphous derivatives, it is possible to obtain phases capable of yielding interpretable electron density maps. Many protein structures have been investigated successfully by this technique. Further details are given in Sect. 7.4.8.

Centrosymmetric Projections

Proteins always crystallize in non-centrosymmetric space groups because the amino acid residues in the polypeptide have left-handed configurations about the α-carbon atoms. Amino acid residues with

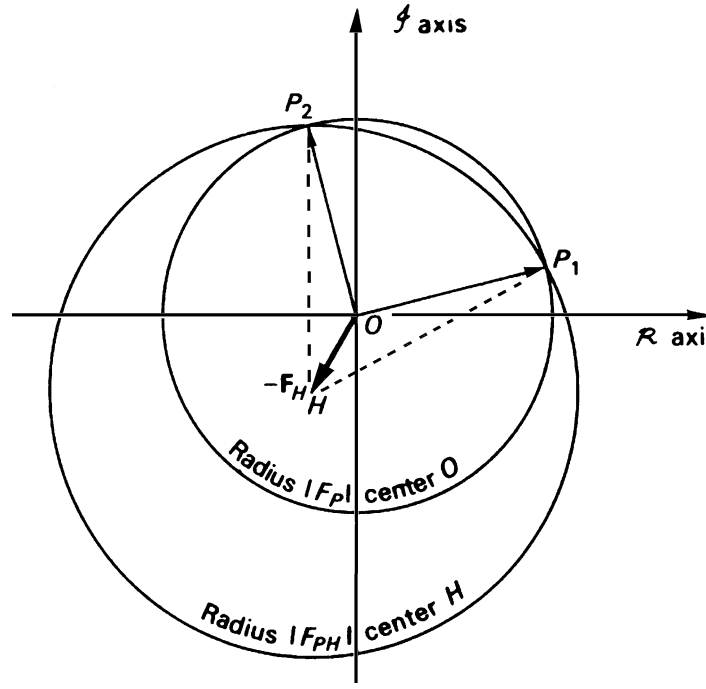


Fig. 7.27 Single isomorphous replacement (SIR) phase–amplitude diagram in the SIR method. $OH(-F_H)$ is the known reversed heavy-atom vector. The triangles OHP_1 and OHP_2 both satisfy (7.32), giving a twofold ambiguity with vectors either OP_1 or OP_2 as the solution for F_P

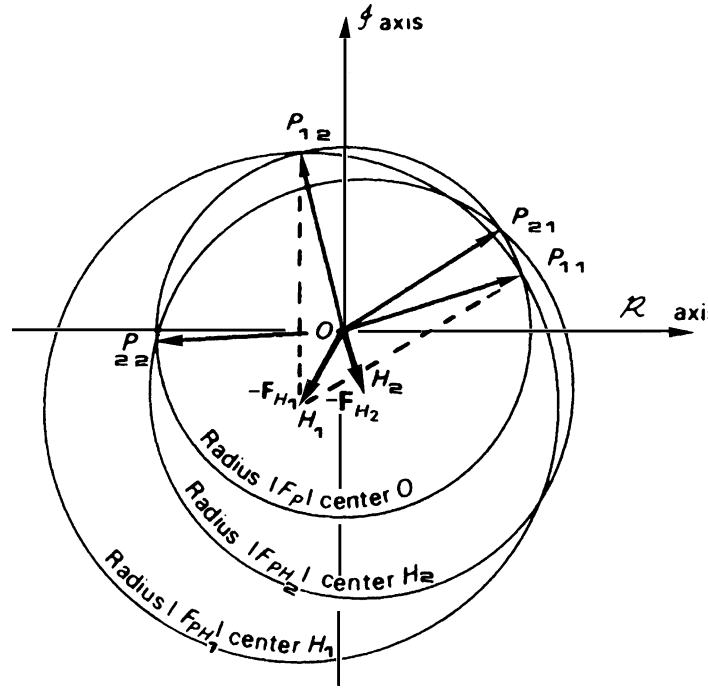


Fig. 7.28 Multiple isomorphous replacement (MIR) resolution of the phase ambiguity with a second isomorphous heavy-atom derivative PH_2. The determined direction of F_P is near vectors OP_{11} and OP_{21}. In practice, P_{11} and P_{21} rarely coincide, due to inaccuracies in the heavy-atom parameters and lack of complete isomorphism

right-handed configurations are very rare in nature. Although non-centrosymmetric structures usually present more difficulties than centrosymmetric structures, there is a compensation in the relative ease of determination of the space group; ambiguities such as $P2_1$ and $P2_1/m$ do not exist for the protein crystallographer. Most non-centrosymmetric space groups have at least one centric zone. In such a case, (7.32) becomes

$$s_{PH}F_{PH} = s_P F_P + s_H F_H \tag{7.37}$$

where s, the sign of the structure factor is ± 1.

Unless both F_{PH} and F_P are very small compared with F_H, it is unlikely that s_{PH} will differ from s_P. Generally F_P and F_{PH} are pointing in the same direction. Accepting this statement, we may substitute, s_P for s_{PH} in (7.37):

$$s_P(F_{PH} - F_P) = s_H F_H \tag{7.38}$$

or

$$s_P = s_H F_H / \Delta F \tag{7.39}$$

where $\Delta F = F_{PH} - F_P$. Since we are interested only in the signs, (7.39) may be rewritten as

$$s_P = s_H s_\Delta \tag{7.40}$$

where s_Δ is +1 if $F_{PH} > F_P$ and -1 if $F_{PH} < F_P$. In this way, signs can often be determined for centric reflections in a protein crystal with only a single isomorphous derivative, and we shall illustrate the method by the following example.

Sign Determination for Centric Reflections in Protein Structures

We shall consider data for both the enzyme ribonuclease and a heavy-atom derivative prepared by soaking pre-grown crystals of the enzyme in $K_2[PtCl_6]$ solution.

Crystal Data for Ribonuclease

System: monoclinic

Unit-cell dimensions: $a = 30.31$ Å, $b = 38.26$ Å, $c = 52.91$ Å, $\beta = 105.9°$

M_r: 13500 (ribonuclease)

Z_P: two molecules of ribonuclease plus an unknown number of water molecules

Z_{PH}: as for $Z_P + N_H$ $[PtCl_6]^{-2}$ groups per unit cell (N_H is the number of heavy atoms groups)

Absent spectra: $0k0$: $k = 2n + 1$

Space group: $P2_1$. The $h0l$ zone is centrosymmetric

Given the heavy-atom positions, Table 7.5 shows how the signs for some $h0l$ reflections have been determined. Notice that experimental errors in F_P and F_{PH}, together with errors in the calculated F_H arising from inaccuracies in the heavy-atom model, are reflected in the inequality of ΔF and F_H. The validity of (7.40) is upheld by these data.

Location of Heavy-Atom Positions in Proteins

In a centrosymmetric zone, it follows from (7.39), since s_P and s_H are ± 1, that

$$F_H = |\Delta F| \tag{7.41}$$

where $|\Delta F| = |F_{PH}| - |F_P|$. A Patterson function calculated with $|F_H|^2$ as coefficients would give the vector set of the substituted heavy atoms in the protein molecule. Since $|F_H|$ cannot be observed

Table 7.5 *h0l* Data for ribonuclease

Observed data					Calculated data		Deduced sign
hkl	F_P	F_{PH}	$\lvert\Delta F\rvert$	s_Δ	$\lvert F_H\rvert$	s_H	$s_p = s_H s_\Delta$
003	437	326	111	−1	50	+1	−1
006	59	48	11	−1	27	−1	+1
007	182	109	73	−1	90	−1	+1
$10, \overline{17}$	144	196	52	+1	31	−1	−1
10,13	146	82	64	−1	52	+1	−1
$10\bar{9}$	97	165	68	+1	55	−1	−1
106	183	242	59	+1	45	+1	+1
$30\bar{4}$	746	861	115	+1	72	+1	+1
405	103	57	46	−1	56	+1	−1

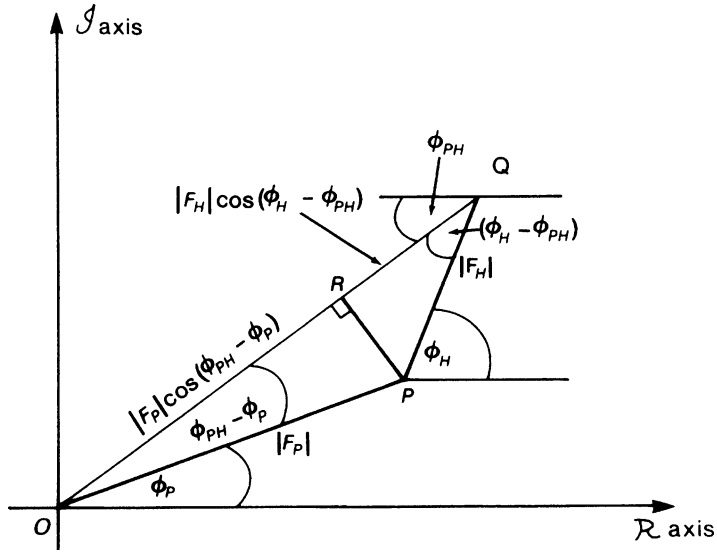


Fig. 7.29 Location of heavy-atom positions in proteins

directly, one calculates a difference Patterson map with $(\Delta F)^2$ as coefficients. If the experimental errors in $|F_P|$ and $|F_{PH}|$ are not significant, and not too many sign "cross-overs" with s_P and s_{PH} occur, then the $(\Delta F)^2$ Patterson projection would be expected to reveal the heavy-atom vectors. In the case of general non-centrosymmetric reflections, we note in Fig. 7.29 that since $OQ = |F_{PH}|$, $OP = |F_P|$, and $OR = |F_P| \cos(\phi_{PH} - \phi_P)$, we have

$$\begin{aligned} RQ &= |F_H| \cos(\phi_H - \phi_{PH}) \quad \text{and} \\ OQ &= OR + RQ \end{aligned} \tag{7.42}$$

Hence,

$$|F_{PH}| - |F_P| \cos(\phi_{PH} - \phi_P) = |F_H| \cos(\phi_H - \phi_{PH}) \tag{7.43}$$

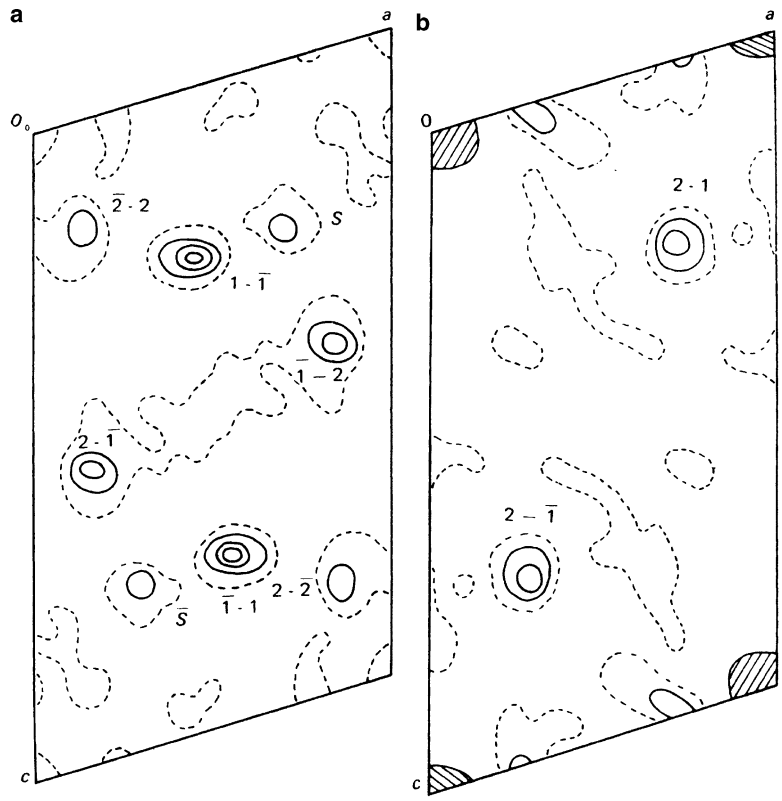


Fig. 7.30 $(\Delta F)^2$ Patterson sections for the Pt derivative of ribonuclease: (**a**) $P(u\frac{1}{2}w)$. (**b**) $P(u0w)$

If $\phi_{PH} - \phi_P$ is small, it follows that $\cos(\phi_{PH} - \phi_P) \approx 1.0$ and

$$|\mathrm{F}_{PH}| - |\mathrm{F}_P| = \Delta F \approx |\mathrm{F}_H| \cos(\phi_H - \phi_{PH}) \tag{7.44}$$

In practice, since the angle $\phi_H - \phi_{PH}$ is undeterminable at this stage, one can calculate a Patterson function with $(\Delta F)^2$ coefficients as for centrosymmetric reflections, but as an added precaution to ensure that $(\phi_{PH} - \phi_P)$ is small, use only those terms for which both $|\mathrm{F}_P|$ and $|\mathrm{F}_{PH}|$ are large. Although the non-centrosymmetric $(\Delta F)^2$ synthesis is not a true Patterson function, it has been used successfully to determine the heavy-atom distribution in proteins.

The most useful derivatives contain a small number of highly substituted sites. Unlike the structure analysis of smaller molecules, it is not known initially how many heavy-atom sites have been incorporated into the molecule.

As an example, we shall consider the $(\Delta F)^2$ Patterson map for the Pt derivative of ribonuclease, space group $P2_1$. The vectors between symmetry-related atoms occur on the Harker section $(u, \frac{1}{2}, w)$. Eight peaks occur on the Harker section and four at $v = 0$, Fig. 7.30a, b. This result suggests that there is more than one heavy-atom site per protein molecule. The most obvious choice is two, since four heavy atoms per unit cell would give rise to 12 non-origin peaks. If the two sites are labeled 1 and 2, their Harker peaks will be of the form $\pm(2x_1, \frac{1}{2}, 2z_1)$ and $\pm(2x_2, \frac{1}{2}, 2z_2)$.

Interpretation of the Patterson function is best undertaken in terms of the Harker section, assuming that the peaks represent nonoverlapping vectors and ignoring the possibility that some peaks could be

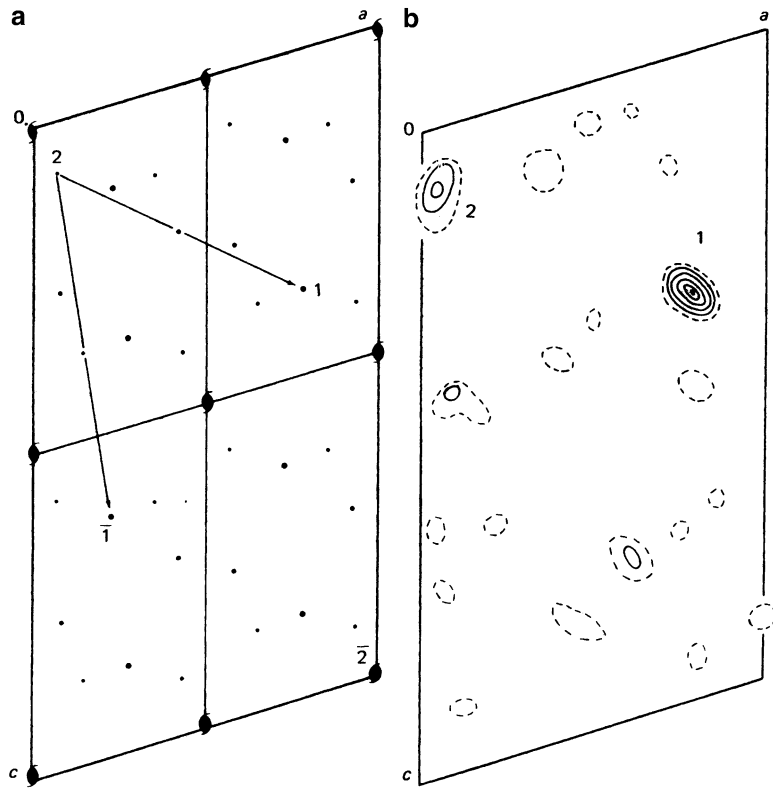


Fig. 7.31 Interpretation of the $(\Delta F)^2$ Patterson sections for ribonuclease. (**a**) Implication diagram. (**b**) Electron density map showing the Pt atom sites

non-Harker peaks. Since the true Harker peaks are of the form $2x$, $2z$, values of x and z can be obtained from the fractional coordinates on the Harker section.

This analysis may be carried out graphically. The peak positions from the Harker section are replotted, on tracing paper, on a unit-cell projection in which the a and c dimensions are each reduced by a factor of $\frac{1}{2}$. This procedure results in one quadrant of Fig. 7.31a. The diagram is completed by operating on the first quadrant with the translation of $a/2$, and then on both quadrants by $c/2$, thus completing an area the size of the true unit-cell projection.

All points marked on this map locate potential (x, z) coordinates for the heavy atoms. In fact, it contains four equivalent solutions with respect to the four unique 2_1 axes in the unit cell. Cross-vector peaks are found by moving this *implication diagram*[5] to other sections of the Patterson function, using pairs of potential sites to generate potential vectors. To see how this mechanism operates, place the site marked 2 on the tracing paper over the origin of the section $v = 0$ and note the coincidence of site 1 with the peak 2-1. Similarly, the peak $2-\bar{1}$ and others on the section $u = \dfrac{1}{2}$ can be generated from the sites 1, 2, $\bar{1}$, and $\bar{2}$ on the implication diagram. Peaks S and $\bar{S}$ are not explained in this way; they may be assumed to be spurious: remember $(\Delta F)^2$ is not a true representation of $|\mathrm{F}_H|^2$.

Figure 7.31b shows a composite electron density map of the Pt atom sites which were prepared by an independent method, and confirms the Patterson analysis. The y coordinates of the two heavy-atom sites are almost equal, which accounts for the presence of the non-Harker peaks $\pm$(2-1) on the Harker

[5] See Bibliography, Buerger (1959).

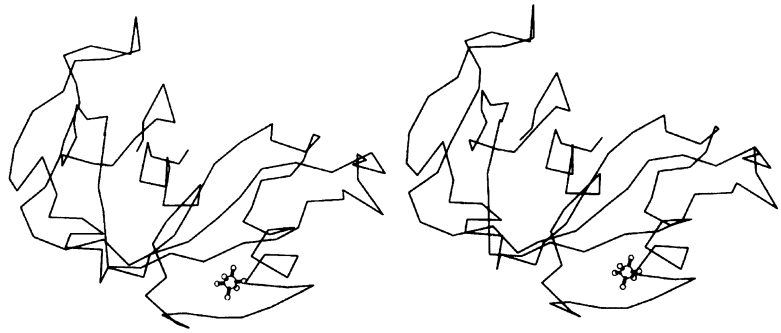


Fig. 7.32 Stereoviews of the polypeptide chain in ribonuclease; the main site in the Pt derivative is shown as a simulated octahedrally coordinated group

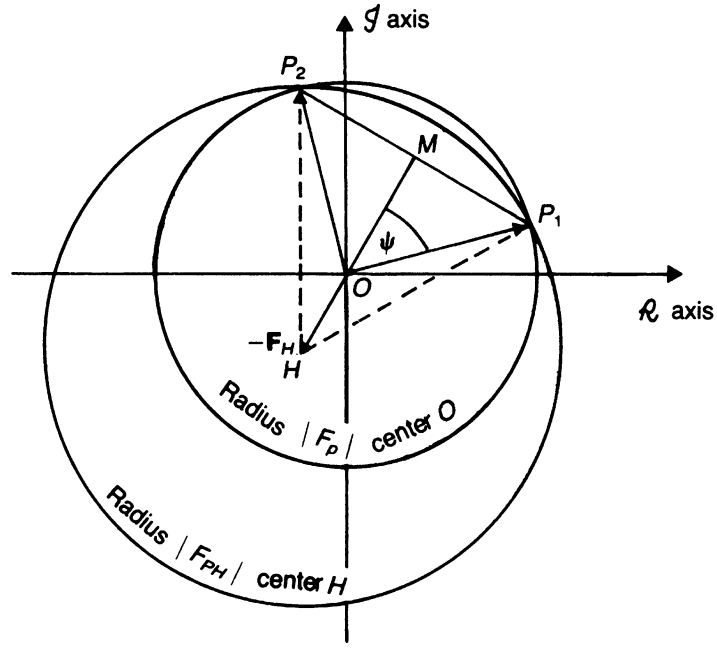


Fig. 7.33 Phase-amplitude diagram in the SIR method; $OH(-F_H)$ is the known reversed heavy-atom vector. The triangles OHP_1 and OHP_2 both satisfy (7.32), giving twofold ambiguity with the vectors either OP_1 or OP_2 as the solution for F_P; OM defines the weighted SIR solution for F_P

section. Figure 7.32 is a stereo-pair showing the course of the polypeptide chain in ribonuclease and the position of the main site in the Pt derivative.

7.5.8 Further Details of the Isomorphous Replacement Phasing Procedure

In single isomorphous replacement (SIR), the ambiguity in ϕ_P is best resolved, Fig. 7.33, by taking $\phi_P = \phi_H$ with F_P along the median OM between P_1 and P_2; $|F_P|$ should be weighted [17] by $m = \cos \Psi$ where Ψ is the semi-angle between P_1 and P_2. By the cosine rule

$$m = \cos \Psi = ||F_{PH}|^2 - |F_P|^2 - |F_H|^2|/(2|F_P||F_H|) \tag{7.45}$$

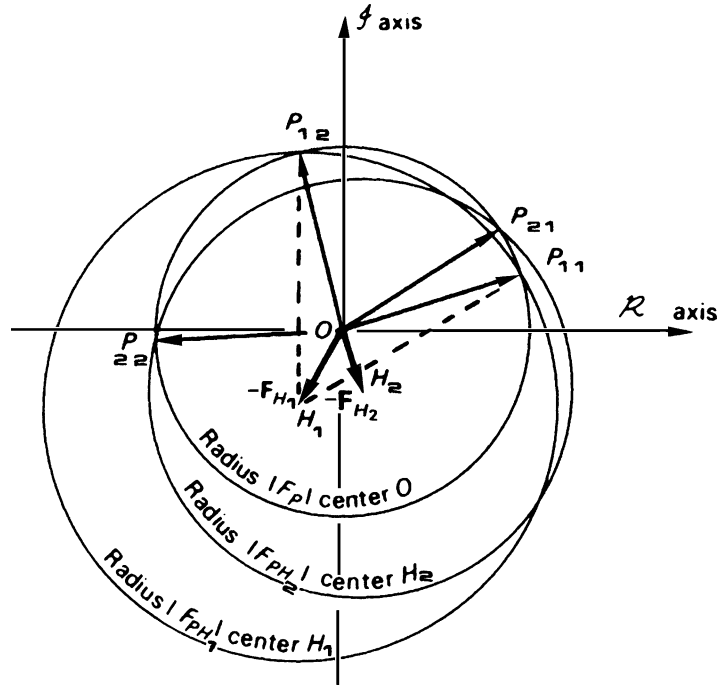


Fig. 7.34 MIR resolution of the phase ambiguity with a second isomorphous derivative PH_2. The determined direction is near the vectors OP_{11} and OP_{21}. In practice P_{11} and P_{21} rarely coincide, owing to inaccuracies in the heavy-atom parameters and a lack of true isomorphism

we see that m would have a maximum value of unity in the special case for which $|\mathrm{F}_H| = ||\mathrm{F}_{PH}| - |\mathrm{F}_P||$, where the two circles in Fig. 7.33 are tangential. The coefficients in the SIR electron density map would be composed of $m|\mathrm{F}_P|$ and ϕ_H. The electron density of such a map would be subject to the pseudo-symmetry effects discussed in Sect. 7.4.4. Taking $\phi_P = \phi_H$ is thus the SIR equivalent of the initial stage of the heavy-atom method, in which we take $\phi = \phi_H$.

Analytical Calculation of Phases in SIR and MIR

The geometrical determination of phases by the isomorphous replacement method using Harker's construction is impractical for several reasons:

1. In MIR (multiple isomorphous replacement), phase-circle intersections, arising from accumulated errors, do not usually give absolutely clear indications of ϕ_P, as shown in Fig. 7.34 by P_{11} and P_{21}. Actual phase determination in MIR, exemplified by Fig. 7.35, contains a complexity of multiple-derivative phase indications, the phase circles intersecting in rather ill-defined regions.
2. The size of the task of estimating thousands of $\phi_P(hkl)$ values in a typical protein analysis necessitated the development of an analytical formula suitable for computer programming, as outlined below.

The basis for a computational algorithm, alternative to the Harker construction for SIR, is shown in Fig. 7.36, in which the inner circle represents $|\mathrm{F}_P|$ and the spokes represent ϕ_T, a series of trial values of ϕ_P for $\phi_T = 0$ to 360° in steps of 30°. The amplitude $|\mathrm{F}_H|$, which would be calculated from (7.33), using the known heavy-atom parameters, is plotted at the end of each spoke. In order to simplify the drawing, the third side of the isomorphous replacement triangle, representing $|\mathrm{F}_{PH}|$, has not been joined up. The SIR solutions, corresponding to P_1 and P_2 in Fig. 7.33, would occur when $|\mathrm{F}_H|$ just touches the $|\mathrm{F}_{PH}|$ circle, which is plotted concentrically with the $|\mathrm{F}_P|$ circle and is the outer circle in Fig. 7.36. These two positions are indicated in the diagram, which should be compared with Fig. 7.33. Because of the method of selecting ϕ_T, neither is generated exactly in this method.

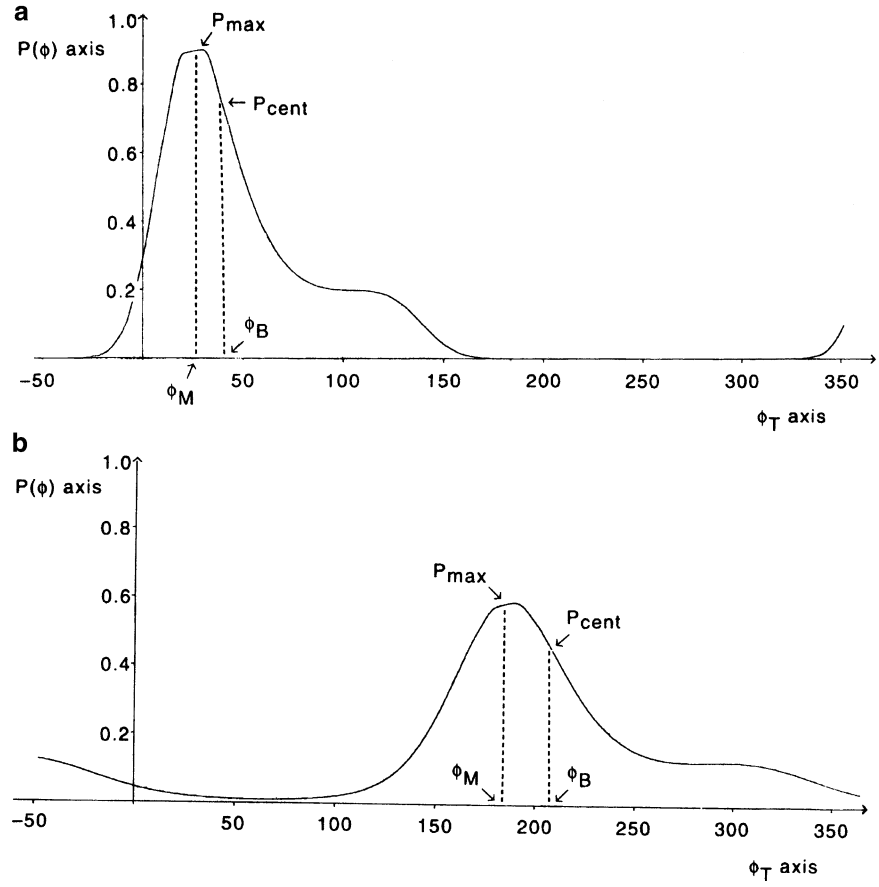


Fig. 7.35 Examples of MIR where the phase circles do not intersect at a point. The most probable value of the phase ϕ (ϕ_M at $P_{\max}$) and the centroid phase (ϕ_B at P_{cent}) are indicated. (**a**) $P(\phi)$ calculated for the two-derivative case in Fig. 7.34, with $\phi_B = 43°$, $\phi_M = 31°$, and $m = 0.80$. (**b**) An example of three-derivative phasing (see Problem 7.14) with $\phi_B = 204°$, $\phi_M = 185°$, and $m = 0.59$. Although based on three derivatives, the probability distribution in (**b**) is not as sharp as that in (**a**), resulting in a lower figure of merit, m, and larger $\phi_B \rightarrow \phi_M$ difference

Now consider Fig. 7.37, which shows a more detailed representation of the case where $\phi_T = 30°$, Fig. 7.36. This is one of the general cases, neither P_1 nor P_2, where the $|\mathrm{F}_{PH}|$ trial value on the Argand diagram would not close the third side of the phase triangle properly. In the case shown $|\mathrm{F}_{PH}|$ is too short; other situations evident in Fig. 7.36 would correspond to $|\mathrm{F}_{PH}|$ being too long. In these situations there is a *lack-of-closure error* denoted by $\varepsilon(\phi_T)$, Fig. 7.37. For the SIR solutions, $\varepsilon(\phi_T) = 0$. In general, $\varepsilon(\phi_T)$ may be calculated as follows:

$$D^2_{\phi_T} = |\mathrm{F}_P|^2 + |\mathrm{F}_H|^2 + 2|\mathrm{F}_P||\mathrm{F}_H|\cos(\phi_T - \phi_H) \tag{7.46}$$

and

$$\varepsilon^2(\phi_T) = [|\mathrm{F}_{PH}| - D(\phi_T)]^2 \tag{7.47}$$

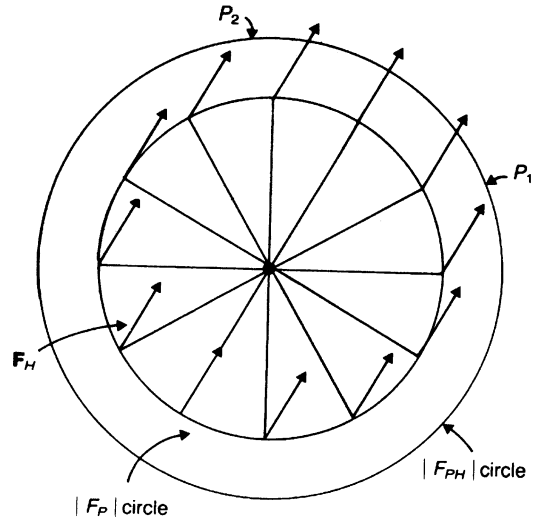


Fig. 7.36 The concept of lack of closure; the inner circle represents $|F_P|$ and the outer circle $|F_{PH}|$. Trial values of ϕ_P are plotted at 30° intervals, each carrying the known F_H. At P_1 and P_2, F_H ends exactly on the F_{PH} circle; otherwise it fails to close, being too long for the smaller region spanning P_1–P_2 and too short for the rest; see Fig. 7.27 for the Harker construction of this SIR case

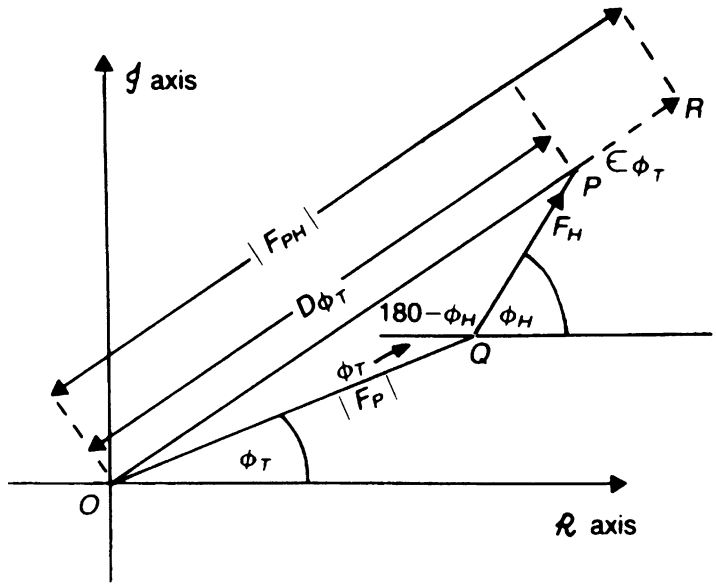


Fig. 7.37 Calculation of the lack-of-closure error $\varepsilon(\phi_T)$

The SIR solutions could be determined to a satisfactory degree of precision by plotting $\varepsilon(\phi_T)$ against ϕ_T and locating the two ϕ_T values for which $\varepsilon(\phi_T) = 0$. This is shown for the example in Figs. 7.33 and 7.36 by the graph of Fig. 7.38. Both solutions P_1 and P_2 are of course equally probable in the SIR method. In the theory of phase analysis by the MIR method, errors may be assumed to reside in $|F_{PH}|$ [18], which simplifies the calculations. For a given trial value of ϕ_T, the probability that ϕ_T is the correct value is

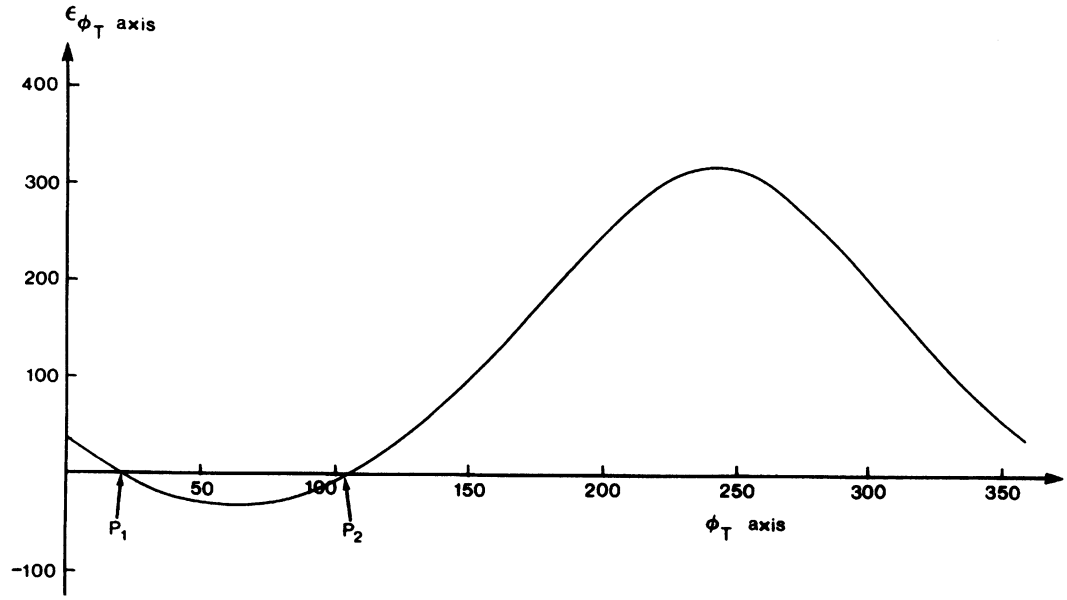


Fig. 7.38 Values of $\varepsilon(\phi_T)$ plotted against ϕ_T; P_1 and P_2 are the two positions for which $\varepsilon_{\phi_T} = 0$, corresponding to the SIR solutions; see Figs. 7.33 and 7.36

$$P(\phi_T) = \exp(-\varepsilon^2(\phi_T)/2E^2) \tag{7.48}$$

where E here is the root-mean-square error in $|F_{PH}|$ arising from data errors.

In MIR there would be one value of $\varepsilon(\phi_T)$ per derivative. Let $\varepsilon_i(\phi_T)$ be the value for derivative i, where $i = 1, 2, \ldots$ to the total number of derivatives. Then the probability for the ith derivative is

$$P_i(\phi_T) = \exp[-\varepsilon_i^2(\phi_T)/2E_i^2] \tag{7.49}$$

and the joint probability over all derivatives is

$$P(\phi_T) = P_1(\phi_T)P_2(\phi_T)P(\phi_T)\ldots$$

or

$$P(\phi_T) = \exp\left[-\sum \varepsilon_i^2(\phi_T)/2E_i^2\right] \tag{7.50}$$

Typical examples of probability distributions met in practice are given in Fig. 7.35. Generally the distributions are bimodal, indicating a stronger preference for one maximum over the other. The most probable electron density map uses coefficients ($|F_P|$, ϕ_M), where ϕ_M is the phase angle corresponding to the maximum probability in the range 0–360°. However, the electron density map with the least overall root-mean-square error uses coefficients ($|F_P|$, ϕ_B), where ϕ_B is the "best" phase angle, corresponding to the centroid of the probability distribution, and m is a weighting function (or figure of merit), given by

$$m\cos\phi_B = \sum_{\phi_T} P(\phi_T)\cos\phi_T \Big/ \sum P(\phi_T) \tag{7.51}$$

$$m \sin \phi_B = \sum_{\phi_T} P(\phi_T) \sin \phi_T \Big/ \sum P(\phi_T) \tag{7.52}$$

It is convenient in practice to evaluate these expressions by stepping from 0 to 360° in regular intervals of 5 or 10°. The probability distributions and corresponding phases may be readily evaluated by suitable programming [19].

For each derivative, the root-mean-square estimate of error may be taken initially as the average

$$E_j^2 = \langle (|\Delta F_i| - |\mathrm{F}_{H_i}|)^2 \rangle_{hkl} \tag{7.53}$$

where the quantity

$$\Delta F_i = |\mathrm{F}_{PH_i}| - |\mathrm{F}_P| \tag{7.54}$$

is evaluated for *centric* reflections only.

The error in a phase angle may be defined as $\Delta\phi = \phi_B - \phi_M$ and $m = \cos \Delta\phi$. A value of $m = 0.7$ corresponds to ≈45°. The average value of m is a measure of the average of cos $\Delta\phi$. In a typical protein analysis at resolution 2 Å[6] (sin $\theta_{\max} = \lambda/4$), an average m of 0.6–0.7 would be acceptable. Further practical details of MIR are to be found in Sect. 10.2.12ff.

Electron Density Maps Used in Large-Molecule Analysis

The correlation of heavy-atom sites between derivatives requires one to establish the coordinates of heavy atoms in a derivative i with respect to those of another derivative or combination of derivatives for which phases $\phi_{P \neq i}$ have been determined. A difference electron density map may be calculated as

$$\rho\Delta_i(xyz) = \frac{1}{V_\mathrm{c}} \sum_h \sum_k \sum_l (|\mathrm{F}_{PH_i}| - |\mathrm{F}_P|) \cos[2\pi(hx + ky + lz) - \phi_{P \neq i}] \tag{7.55}$$

This should reveal the heavy atoms in derivative i with respect to the same origin as in the other heavy-atom derivatives. Derivative i can then be added into the MIR procedure.

For a trial structure in which phases have been calculated, as in small-molecule analysis, a difference electron density map may be used in order to effect corrections to the structure:

$$\Delta\rho(xyz) = \frac{1}{V_\mathrm{c}} \sum_j \sum_k \sum_l (|\mathrm{F}_P| - |\mathrm{F}_\mathrm{c}|) \cos[2\pi(hx + ky + lz) - \phi_\mathrm{c}] \tag{7.56}$$

Alternatively a double-difference map

$$\rho'(xyz) = \frac{1}{V_\mathrm{c}} \sum_h \sum_k \sum_l (2|\mathrm{F}_P| - |\mathrm{F}_\mathrm{c}|) \cos[2\pi(hx + ky + lz) - \phi_\mathrm{c}] \tag{7.57}$$

[6] The resolution of a protein X-ray analysis is loosely defined as $d_{\min}$, where $d_{\min} = \lambda/2 \sin \theta_{\max}$, $\theta_{\max}$ being the maximum Bragg angle associated with the analysis: initially $\theta_{\max}$ may be temporarily restricted in order to limit the work required, but at the expense of the quality of the electron density image.

where $\rho'(xyz)$, equal to $\rho_c(xyz) + \Delta\rho(xyz)$, may be used since new features may be more easily recognized in $\Delta\rho(xyz)$ against the background of the known $\rho_c(xyz)$ structure. This map is very useful in computer graphics analysis, Sect. 10.4.4.

In MIR the most error-free electron density is calculated as

$$\rho_P(xyz) = \frac{1}{V_c} \sum_h \sum_k \sum_l m|\mathrm{F}_P| \times \cos[2\pi(hx + ky + lz) - \phi_B] \tag{7.58}$$

where ϕ_B is the MIR phase corresponding to the centroid of the phase probability distribution (the best phase) and m is the figure of merit; see (7.51) and (7.52).

7.6 Anomalous Scattering

Friedel's law is not an exact relationship, and becomes less so as the atomic numbers of the constituent atoms in a crystal increase. The law breaks down severely if X-rays are used that have a wavelength just less than that of an absorption edge of an atom in the crystal, Sect. 3.1.3. However, this criterion is not essential for anomalous scattering to be used in two important aspects of crystal structure analysis, namely, the determination of absolute stereochemical configurations and the phasing of reflections.

Anomalous scattering introduces a phase change into a given atomic scattering factor, which becomes complex:

$$f = f_o + \Delta f' + \mathrm{i}\Delta f'' = f' + \mathrm{i}\Delta f'' \tag{7.59}$$

$\Delta f'$ is a real correction, usually negative, and $\Delta f''$ is an imaginary component which is rotated anticlockwise through 90° in the complex plane with respect to f_o and $\Delta f'$, that is, to f'.

A possible situation is illustrated in Fig. 7.39. In Fig. 7.39a, atom A is assumed to be scattering in accordance with Friedel's law, and it is clear that $|\mathrm{F}(\mathbf{h})| = |\mathrm{F}(\bar{\mathbf{h}})|$, where $\mathbf{h}$ stands for hkl. In Fig. 7.39b, atom A is represented as an anomalous scattering species, with its three components according to (7.59). In this situation, $|\mathrm{F}(\mathbf{h})| \neq |\mathrm{F}(\bar{\mathbf{h}})|$, and intensity measurements of Friedel pairs of reflections produce different values.

We can assume safely that procedures for measuring $I(hkl)$ differentiate correctly between hkl and $\bar{h}\,\bar{k}\,\bar{l}$. In any non-centrosymmetric space group a structure model can be inverted, as if through a center of symmetry, and used to recalculate the structure factors. Because structure factor formulae involve $(hx + ky + lz)$, the two models will produce different values for $|\mathrm{F}_c(hkl)|$ and $|\mathrm{F}_c(\bar{h}\,\bar{k}\,\bar{l})|$. The correct enantiomorph is expected to produce better agreement between F_o and $|\mathrm{F}_c|$ and, thus, a lower R-factor. Some typical results are listed in Table 7.6, from which it may be deduced that the structure giving $|\mathrm{F}_c|_{\bar{x},\bar{y},\bar{z}}$ corresponds to the absolute configuration. An equivalent procedure would be to measure the values of both $|\mathrm{F}(hkl)|$ and $|\mathrm{F}(\bar{h}\,\bar{k}\,\bar{l})|$ and compare them with $|\mathrm{F}_c|_{x,y,z}$. The technique can be used only with crystals which are non-centrosymmetric, because $|\mathrm{F}(\mathbf{h})| = |\mathrm{F}(\bar{\mathbf{h}})|$ in centrosymmetric crystals, but this limitation is not important because molecules which crystallize with a single enantiomorph cannot do so in a space group containing any form of inversion symmetry.

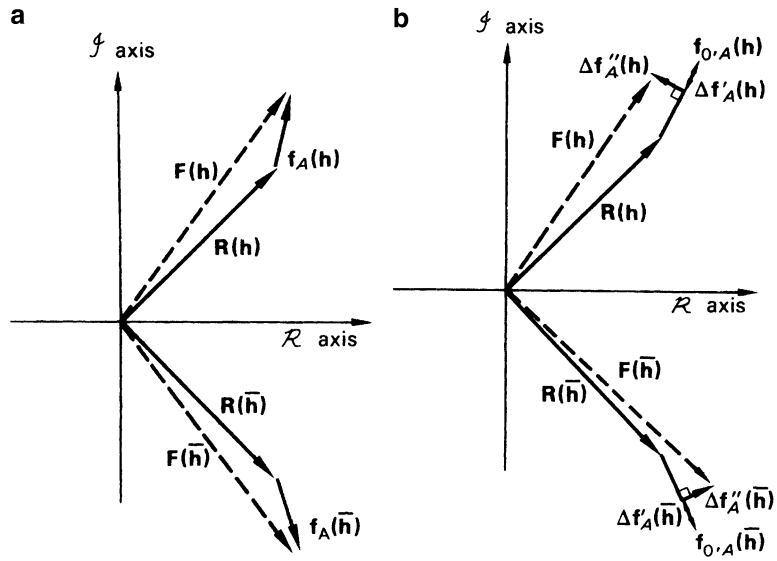


Fig. 7.39 Anomalous scattering of atom A with respect to the rest of the structure R. (**a**) Normal case: $|\mathrm{F}(\mathbf{h})| = |\mathrm{F}(\bar{\mathbf{h}})|$. (**b**) Anomalous case: $|\mathrm{F}(\mathbf{h})| \neq |\mathrm{F}(\bar{\mathbf{h}})|$. The general non-centrosymmetric case is illustrated. For centrosymmetric crystals, $|\mathrm{F}_\mathbf{h}|$ always equals $|\mathrm{F}_\mathbf{h}|$, but $\phi_\mathbf{h}$ differs to a small degree from 0 or π if anomalous scattering is significant

Table 7.6 Example of some Friedel pairs and the corresponding |F| values

hkl	F_o	$\lvert F_c\rvert_{x,\,y,\,z}$	$\lvert F_c\rvert_{\bar{x}\bar{y}\bar{z}}$
121	17.0	19.1	18.3
122	21.2	22.9	21.9
123	41.4	44.4	42.8
341	36.7	38.7	35.5
342	7.8	9.5	8.2
413	14.2	15.3	13.5

7.6.1 The Flack *x* Parameter

It is clear from the above treatment that a more complete, rigorous, and easily applied method for the determination of absolute configuration would be highly desirable. To this end several proposals have been made: Hamilton's R-ratio test [20] requires the number of observations to be known; Rogers' η parameter [21] attempted to overcome operational difficulties associated with Hamilton's test, but assumed that reflections for which $|\mathrm{F}(hkl)| = |\mathrm{F}(\bar{h}\,\bar{k}\,\bar{l})|$, and there are always some except in space group $P1$, had no influence on the refinement of η and its standard deviation. Flack's x parameter method [22] showed that the assumption of Rogers was not always valid, and it has proved to be the most useful and reliable method; it is currently employed in program packages such as SHELX-97. The Flack x parameter is refined by least squares, together with the other structural parameters, Sect. 8.4.3, to a final value and corresponding standard deviation. The Flack parameter x is defined in terms of $|\mathrm{F_c}|$ by the equation

$$|\mathrm{F_c}(hkl)|^2 = (1-x)|\mathrm{F_c}(hkl)|^2 + x|\mathrm{F_c}(\bar{h}\,\bar{k}\,\bar{l})|^2 \tag{7.60}$$

When the atomic coordinate set and the crystal have the same chirality, x takes the value zero; if they are different, x is equal to unity. A result is considered acceptable if x lies within 3 standard deviations, that is, $\pm 3\sigma(x)$ as calculated by the program, of either zero or unity. If a value of unity is returned, the coordinates of the model should be inverted in the origin; the data collection routines ensure that the *hkl* data have the correct polarity. After further refinement a value of $0.0 \pm 3\sigma(x)$ for x should result.

This method has withstood innumerable tests and can be considered reliable in most circumstances. Some examples are given below. A further advantage of this method is that although, as with all analyses, the measured intensity data set should cover as large a volume of reciprocal space as possible, it is not absolutely necessary to measure Friedel or Bijvoet pairs, Sect. 7.6.2. In practice, for the test to be reliable, $\sigma(x)$ in x should be 0.05 or less. Because the refinement routine in SHELX refines against $|F|^2$ as opposed to $|F|$ Sect. 8.4.2, the program determines x within a small number of iterations. It should also be noted that since the value of x may be related to twinning, Sect. 5.10.2, it is advisable to carry out further tests for this effect as prescribed, for example, in the SHELX system manual. Quite rightly the SHELX manual advises that it is important to refine every non-centrosymmetric structure as the correct absolute structure in order to avoid introducing systematic errors into the molecular geometry. In fact, it cannot be overemphasized that it is often imperative to determine the absolute configuration, especially in the case of new drug molecules for which a reversal of chirality is almost certain to lead to a dramatic change in potency and other binding characteristics. The classic example [23] of this is that of thalidomide in the early 1960s. In some cases the absolute structure will be known with certainty, for example, proteins which are predominantly L-amino acid polypeptides, but in others it has to be deduced from the X-ray structure.

If the structure includes a reasonably heavy atom, such as phosphorus, sulphur or chlorine, the anomalous scattering effect from these atoms using Cu $K\alpha$ or even accurate high-resolution low-temperature Mo $K\alpha$ radiation may be sufficient for determining the absolute configuration, particularly if Friedel opposites have been measured. Sometimes the presence of as few as two oxygen atoms per molecule has been shown to produce a reliable result, Sect. 9.4.4. In SHELX the refinement program estimates [22] the parameters x and $\sigma(x)$ for the absolute structure.

The parameter x is effectively the fractional contribution of the inverted component of a "racemic twin" and, as we have said, should be zero if the absolute structure is correct, or 1 if it needs to be inverted in the origin; but it could be somewhere between 0 and 1 if racemic twinning is really present. If this is the case, x can be refined along with all the other parameters, and the instructions for TWIN refinement in the SHELX manual must be followed faithfully; fortunately such cases are fairly uncommon.

Examples to Illustrate Results of Flack Parameter Refinement

The following six crystal structures were all refined using SHELX-97 and have different degrees of heavy atom content. The Flack parameter produced in the final least-squares cycle was within three standard deviations of zero in each of the first five structures. All six structures are, of course, in non-centrosymmetric space groups and the first four have single molecular species of one specific hand, determined absolutely by the X-ray analysis. The fifth structure, atropine, was crystallized from a commercial racemic mixture in the unusual space group *Fdd*2 and contains both enantiomorphic forms of the atropine molecule. The space group itself is enantiomorphous, Tables 2.7, and the crystal structure has therefore established, via the Flack parameter, the correct hand of the three-dimensional arrangement of molecules in the crystal. For the sixth example the value of x was not within 3σ of 0 (or 1).

1. Form I of Cholesteryl Iodide [24], $C_{27}H_{45}I$

   X-ray data were collected at low temperature (100 K) with Mo $K\alpha$ radiation ($\lambda = 0.71073$ Å). The crystals are monoclinic, space group $P2_1$ with $a = 12.577(1)$ Å, $b = 9.009(1)$ Å, $c = 12.862$

(1) Å, $\beta = 119.00(1)^\circ$, and $Z = 2$ molecules per unit cell. The crystal size was 0.35, 0.56, 0.06 mm, and the θ-range for data collection 1.81–29.15°. The number of reflections collected was 14457, with 5729 unique reflections ($R_{\text{int}} = 0.0632$). Refinement by full-matrix least-squares on $|F|^2$ gave final R indices ($I > 2\sigma_I$) $R_1 = 0.0499$, weighted $R_2 = 0.1119$; R indices on all data were $R_1 = 0.0751$ and weighted $R_2 = 0.1345$. The absolute structure parameter x was $-0.04(3)$ which is zero within 3σ, σ being less than 0.05.

2. Form II of Cholesteryl Iodide [24], $C_{27}H_{45}I$
X-ray data were collected at room temperature with Cu $K\alpha$ radiation ($\lambda = 1.54178$ Å). Crystals are monoclinic, space group $P2_1$ with $a = 11.005(1)$ Å, $b = 10.469(1)$ Å, $c = 11.6840(10)$ Å, $\beta = 106.827(6)^\circ$, and $Z = 2$. The crystal size was 0.24, 0.42, 0.24 mm, and the θ-range for data collection was 3.95–69.91°. The number of reflections collected was 3543, with 2470 unique reflections ($R_{\text{int}} = 0.0191$). Refinement was by full-matrix least-squares on $|F|^2$ which gave final R indices ($I > 2\sigma_I$): $R_1 = 0.0480$, weighted $R_2 = 0.1299$; R indices on all data were $R_1 = 0.0537$, weighted $R_2 = 0.1447$. The absolute structure parameter $x = 0.00(1)$ which is zero within 3σ, σ being less than 0.05.

The structures (1) and (2) are polymorphs, and have exactly the same chemical composition. The iodine atom is a strong anomalous scattering species for either Mo radiation in the case of (1) or Cu radiation in the case of (2). Both analyses confirm the absolute configuration of cholesterol, an important biological substance, and of steroids in general.

3. Compound BW202W92: $C_{12}H_{12}Cl_3FN_4O_3$ ($C_{11}H_9Cl_3FN_4{\cdot}CH_3SO_3$) (*R*-form) [25]
X-ray data were collected at room temperature (293 K) with Cu $K\alpha$ radiation ($\lambda = 1.54178$ Å). The crystals are monoclinic, space group $P2_1$ with $a = 8.384(2)$ Å, $b = 16.984(3)$ Å, $c = 12.480(3)$ Å, $\beta = 104.14(6)^\circ$, and $Z = 2$. The crystal size was 0.30, 0.16, 0.16 mm, and the θ-range for data collection 4.49–74.42°. The number of reflections collected was 3990 reflections, with 3192 unique reflections ($R_{\text{int}} = 0.0496$). Refinement was by full-matrix least-squares on $|F|^2$ which gave final R *i*ndices ($I > 2\sigma_I$) $R_1 = 0.0490$, weighted $R_2 = 0.1254$; R indices on all data were $R_1 = 0.0644$, weighted $R_2 = 0.1324$. The absolute structure parameter $x = 0.05(3)$ which is zero within 3σ, σ being less than 0.05.

4. Compound BW203W92: $C_{12}H_{12}Cl_3FN_4O_3$ ($C_{11}H_9Cl_3FN_4{\cdot}CH_3SO_3$) (*S*-form) [25]
X-ray data were collected at low temperature (123 K) with Mo $K\alpha$ radiation ($\lambda = 0.71073$ Å). The crystals are triclinic, space group $P1$ with $a = 7.716(2)$ Å, $b = 8.120(2)$ Å, $c = 13.719(3)$ Å, $\alpha = 74.91(3)^\circ$, $\beta = 87.69(3)^\circ$, $\gamma = 89.83(3)^\circ$, and $Z = 1$. The crystal size was 0.30, 0.20, 0.20 mm, and the θ-range for data collection 2.60–27.47°. The number of reflections collected was 6456, with 5302 unique reflections ($R_{\text{int}} = 0.0733$). Refinement was by full-matrix least-squares on $|F|^2$, which gave final R *i*ndices ($I > 2\sigma_I$) $R_1 = 0.0614$, weighted $R_2 = 0.1204$; R indices on all data were $R_1 = 0.1248$, weighted $R_2 = 0.1468$. The absolute structure parameter $x = -0.10(9)$ which is zero within 3σ, but σ here is greater than 0.05.

Structures (3) and (4) are atropisomers [25] and have exactly the same chemical structure. Chemical and spectroscopic studies have identified (3) as the *R*-isomer and (4) as the *S*-isomer, but with insufficient details to enable their three-dimensional structures to be finalized; this was achieved through X-ray structure analyses. The structures are good candidates for anomalous X-ray scattering by virtue of their three chlorine, one sulphur, and three oxygen atoms. It is clear, however, that the result for structure (3) is more convincing, being based on copper X-radiation data and resulting in $x = 0.05(3)$, which is zero within 3σ ($\sigma < 0.05$), while structure (4), based on molybdenum X-radiation data, has $x = -0.10(9)$ which is also zero within 3σ, but with σ greater than 0.05, slightly larger than the recommended limit. Nevertheless, it is safe to say that the X-ray analyses have established the absolute configurations of these important drugs for which (3) is a potent/selective sodium-ion channel blocker and (4) is about 30 times less potent in this capacity.

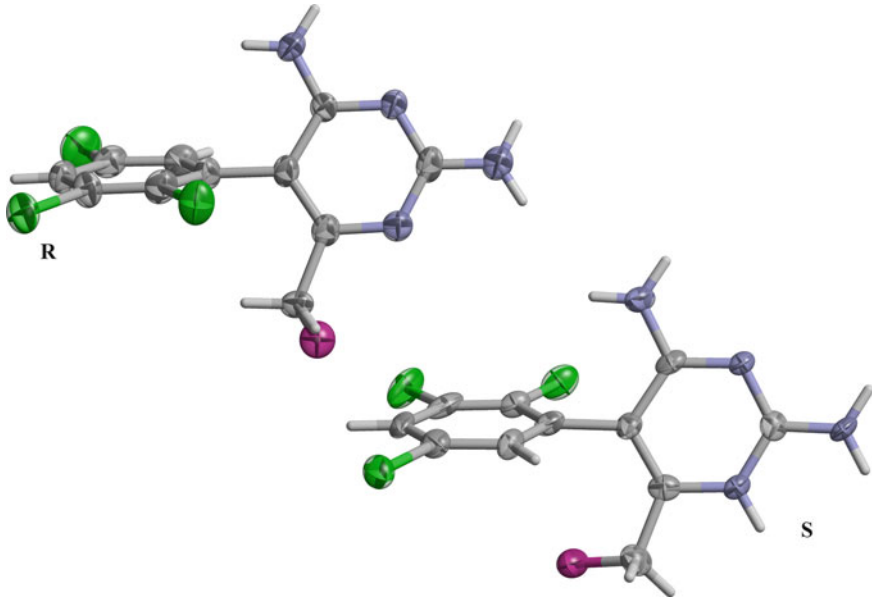


Fig. 7.40 Molecular conformations of BW202W92: $C_{11}H_9Cl_3FN_4{\cdot}CH_3SO_3$ (*R*-form) and BW203W92: $C_{12}H_{12}Cl_3FN_4O_3$ (*S*-form) [25] the *R* and *S* enantiomers respectively of 2,4-diamino-6-fluoromethyl-5-(2,3,5-trichlorophenyl)-pyrimidine (Drawn with Ortep/Raster [26, 27]). Thermal ellipsoids are shown at 50% probability; for clarity the CH_3SO_3 moieties have been omitted. The views are shown perpendicular to the pyrimidine ring (RHS) in order to emphasize the difference between the two molecules with respect to the central linking bond

The *R* and *S* enantiomers derived from the X-ray analyses are shown in Fig. 7.40; see also Sect. 8.2.16.

5. Atropine [28], $C_{17}H_{23}NO_3$
   X-ray data were collected at low temperature (100 K) with Mo $K\alpha$ radiation ($\lambda = 0.71073$ Å). The crystals are orthorhombic, space group $Fdd2$ with $a = 24.291(5)$ Å, $b = 39.538(8)$ Å, $c = 6.473(1)$ Å, and $Z = 16$. The crystal size was 0.20, 0.15, 0.35 mm, and the θ-range for data collection 3.30–25.02°. The number of independent reflections collected was 14353, with 2710 unique reflections ($R_{int} = 0.0820$). Refinement was by full-matrix least-squares on $|F|^2$, which gave final R *i*ndices ($I > 2\sigma_I$) $R_1 = 0.0452$, weighted $R_2 = 0.1206$; R indices on all data $R_1 = 0.0558$, weighted $R_2 = 0.1294$. The absolute structure parameter $x = -0.1(2)$ which is zero within 3σ, but with σ greater than 0.05. Considering that there are no heavy anomalous scattering species in this structure, only three oxygen atoms, this is an excellent result and fully acceptable; see also Sect. 8.3.7.
6. 1-Benzyl-1*H*-tetrazole, $C_8H_8N_4$
   X-ray data were collected at low temperature (120 K) with Mo $K\alpha$ radiation ($\lambda = 0.71073$Å). The crystals are monoclinic, space group $P2_1$ with $a = 7.6843(5)$ Å, $b = 5.5794(7)$ Å, $c = 9.4459(7)$ Å, and $Z = 2$. The crystal size was 0.09, 0.03, 0.02 mm, and the θ-range for data collection 3.14–27.45°. The number of independent reflections collected was 5624, with 1718 unique reflections ($R_{int} = 0.0588$). The refinement was by full-matrix least-squares on $|F|^2$, which gave final R *i*ndices ($I > 2\sigma_I$) $R_1 = 0.0447$, weighted $R_2 = 0.0786$; R indices on all data were $R_1 = 0.0695$, weighted $R_2 = 0.0870$. The absolute structure parameter $x = 0(3)$. Inversion of the structure produced a result with an absolute structure parameter $x = 1(3)$.

The molecular structure is shown in Fig. 7.41. What does the Flack parameter mean in this case? In fact, the values of x for the structure and its inverse are exactly correct, 0 and 1, but the estimated standard deviation of 3 in each case deprives this result of a statistical significance. We can relate the

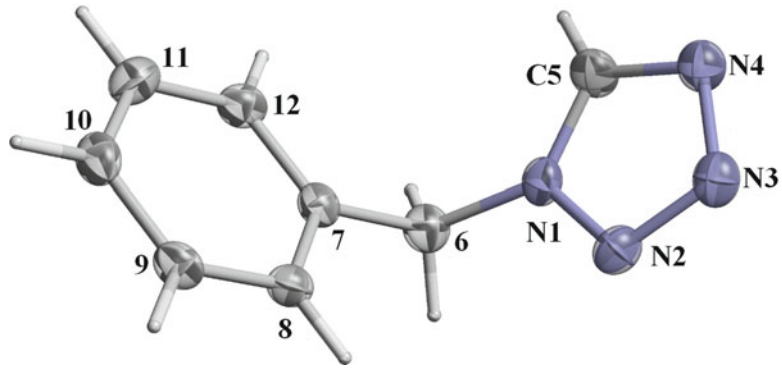


Fig. 7.41 X-ray molecular structure of 1-benzyl-*H*-tetrazole [29] (Drawn with Ortep/Raster [26, 27]). Thermal ellipsoids are shown at 50 % probability

failure of the Flack refinement to the lack of significant anomalous scattering species in the molecule as well as to the choice of molybdenum radiation. Inspection of Fig. 7.41 reveals that there are no chiral centers in this molecule but the space group $P2_1$ is enantiomorphic, Table 10.1, and therefore could be more associated with molecules which are chiral. In this case, there is no need for concern over the absolute configuration of the molecule as it does not have one. It should be noted that in fact molecules chemically similar to this species usually crystallize in *centrosymmetric* space groups, avoiding space groups such as $P2_1$.

Conclusions
These examples illustrate the successful use of Flack parameter refinement to determine absolute configuration. It is easy to apply and has a high chance of success particularly where there are a good number of anomalous scattering species in the structure and the diffraction data has been measured carefully and completely. The cause of the failed result lay in the given experimental conditions. It should be said, however, that it is always sensible to refine the Flack *x* parameter in all cases and run the final refinement stages again if the test fails.

7.6.2 Effect of Anomalous Scattering on the Symmetry of Diffraction Patterns

We have seen that when Friedel's law holds, the X-ray diffraction pattern, considered as a three-dimensional weighted reciprocal lattice, exhibits a symmetry equivalent to that of the point group of the crystal with an additional center of symmetry, if not already present, that is, to one of the eleven diffraction symmetry groups (Laue groups) listed in Table 1.6. However, for a structure in which some of the atoms scatter anomalously Friedel's law breaks down and the symmetry of the diffraction pattern then reverts to that of the point group of the crystal. For a centrosymmetric crystal $|F(hkl)|$ still equals $|F(\bar{h}\,\bar{k}\,\bar{l})|$, although the phase angle is no longer 0 or π. As an example to illustrate the effects of anomalous scattering let us consider a crystal in space group $P2_1$.

Diffraction Symmetry for a Crystal in Space Group $P2_1$ with No Anomalous Scattering
A given crystal is monoclinic, belonging to point group 2, and the diffraction symmetry if Friedel's law holds is 2/*m*. The $|F(hkl)|$ equivalents are thus $|F(hkl)| = |F(\bar{h}\,k\,\bar{l})| = |F(h\,\bar{k}\,l)| = |F(\bar{h}\,\bar{k}\,\bar{l})|$. On X-ray photographs, for example, zero-level and upper-level *a*-axis precession photographs, this symmetry will be manifested as indicated in Fig. 7.42. The zero level, Fig. 7.42a(i), clearly

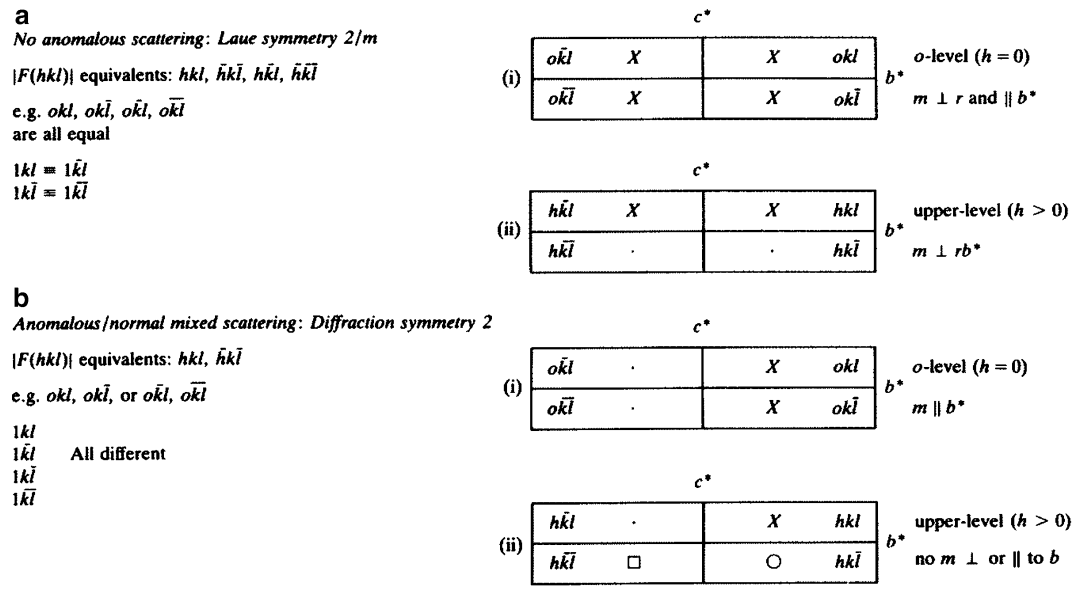


Fig. 7.42 Schematic representation of symmetry exhibited in b^*c^* sections of the reciprocal lattice for a crystal with space group $P2_1$ (or any other monoclinic space group in crystal class 2). (**a**) No anomalous scattering; the diffraction symmetry is $2/m$. (i) Symmetry of $0kl$ section is $mm2$. (ii) Symmetry of $1kl$, $2kl$, ... sections is $m \perp b^*$. Equivalent hkl are represented by the same symbol X, etc. (**b**) Anomalous scattering case; the diffraction symmetry is 2. (i) Symmetry of $0kl$ section is $m \perp b$.* (ii) Symmetry of $1kl$, $2kl$, ... sections is reduced to 1 (no m present)

demonstrates symmetry $mm2$ with $|\mathrm{F}(0kl)| = |\mathrm{F}(0\bar{k}\,l)| = |\mathrm{F}(0k\bar{l})| = |\mathrm{F}(0\bar{k}\,\bar{l})|$, while upper-level photographs exhibit symmetry m (perpendicular to b^*) with, for example $|\mathrm{F}(1kl)| = |\mathrm{F}(1\bar{k}l)|$ and $|\mathrm{F}(1k\bar{l})| = |\mathrm{F}(1\bar{k}\,\bar{l})|$ respectively, but $|\mathrm{F}(1kl)| \neq |\mathrm{F}(1k\bar{l})|$, Fig. 7.42a(ii).

Diffraction Symmetry for a Crystal in Space Group $P2_1$ for a Structure Containing Some Anomalous Scattering Species

The diffraction symmetry is now that of point group 2, for which $|\mathrm{F}(hkl)| = |\mathrm{F}(\bar{h}k\bar{l})|$ and $|\mathrm{F}(\bar{h}\,\bar{k}\,\bar{l})| = |\mathrm{F}(h\bar{k}l)|$. On the a-axis precession photographs shown schematically in Fig. 7.42b(i) we observe, on the zero level, $|\mathrm{F}(0kl)| = |\mathrm{F}(0k\bar{l})|$ and $|\mathrm{F}(0\bar{k}l)| = |\mathrm{F}(0\bar{k}\,\bar{l})|$, but $|\mathrm{F}(0kl)| \neq |\mathrm{F}(0\bar{k}l)|$ and $|\mathrm{F}(0k\bar{l})| \neq |\mathrm{F}(0\bar{k}\,\bar{l})|$ because there is now no m plane perpendicular to b^*. On the upper-level photographs, no symmetry is observable and all four $|\mathrm{F}|$ values are different, Fig. 7.40b(ii).

We may define the anomalous difference as

$$\begin{aligned} \Delta F_{ANO} &= (|\mathrm{F}(hkl)| - |\mathrm{F}(\bar{h}\,\bar{k}\,\bar{l})|) \\ &\text{or as} \\ \Delta F_{\mathrm{ANO}} &= (|\mathrm{F}(+)| - |\mathrm{F}(-)|) \end{aligned} \tag{7.61}$$

From the above discussion, for the case of space group $P2_1$ with anomalous scattering species present, ΔF_{ANO} is also given by

$$\Delta F_{\mathrm{ANO}} = (|\mathrm{F}(hkl)| - |\mathrm{F}(h\bar{k}l)|) \tag{7.62}$$

since $|\mathrm{F}(\bar{h}\,\bar{k}\,\bar{l})| = |\mathrm{F}(h\bar{k}l)|$. Differences such as (7.62), equivalent by point-group symmetry to the difference between $|\mathrm{F}|$ for a Friedel pair, are known as *Bijvoet differences*, the two reflections involved being denoted as a *Bijvoet pair* [30].

Bijvoet differences can be observed on X-ray photographs represented in Fig. 7.42a, b as follows:
(a) Zero level: $h = 0$

$$\begin{aligned} &(|\mathrm{F}(0kl)| - |\mathrm{F}(0\bar{k}\bar{l})|) \quad \text{Bijvoet pair} \\ &(|\mathrm{F}(0k\bar{l})| - |\mathrm{F}(0\bar{k}\,\bar{l})|) \quad \text{Bijvoet pair} \end{aligned}$$

These differences will be equivalent by symmetry.
(b) Upper level

$$\begin{aligned} &(|\mathrm{F}(hkl)| - |\mathrm{F}(h\bar{k}\bar{l})|) \quad \text{Bijvoet pair} \\ &(|\mathrm{F}(hk\bar{l})| - |\mathrm{F}(h\bar{k}\,\bar{l})|) \quad \text{Bijvoet pair} \end{aligned}$$

These differences are not equivalent.

It is, thus, possible to monitor anomalous differences indirectly on *the same photograph* through the use of Bijvoet pairs, whereas Friedel pairs, in the true sense of the definition, would necessarily always occur on different photographs (except for axial reflections, such as $00l$ and $00\bar{l}$).

7.6.3 Form of the Structure Factor for a Structure Composed of Heavy-Atom Anomalous Scattering Species

For a structure composed of N_H heavy-atom anomalous scattering species, the structure factor becomes

$$F(hkl) = \sum_{j=1}^{N_H} (f'_j + \mathrm{i}\Delta f''_j) \exp 2\pi\mathrm{i}(hx_j + ky_j + lz_j) \tag{7.63}$$

which can be written as

$$\mathrm{F}_H(+) = \mathrm{F}'_H(+) + \mathrm{iF}''_H(+)$$

where

$$\mathrm{F}'_H(+) = \sum_{j=1}^{N_H} f'_j \exp 2\pi\mathrm{i}(hx_j + ky_j + lz_j) \tag{7.64}$$

and

$$\mathrm{F}''_H(+) = \sum_{j=1}^{N_H} \Delta f''_j \exp 2\pi\mathrm{i}(hx_j + ky_j + lz_j) \tag{7.65}$$

Similarly the structure factor $F(\bar{h}\,\bar{k}\,\bar{l})$ can be written as

$$\mathrm{F}_H(-) = \mathrm{F}'_H(-) + \mathrm{i}F''_H(-)$$

where

$$\mathrm{F}'_H(-) = \sum_{j=1}^{N_H} f'_j \exp\{-2\pi \mathrm{i}(hx_j + ky_j + lz_j)\} \tag{7.66}$$

and

$$\mathrm{F}''_H(-) = \sum_{j=1}^{N_H} \Delta f''_j \exp\{-2\pi \mathrm{i}(hx_j + ky_j + lz_j)\} \tag{7.67}$$

The form of the structure factor for a protein heavy-atom derivative crystal composed of protein atoms P, light atom, negligible-anomalous scattering species, mainly C, N, O atoms, and heavy atoms H, anomalous scattering species, becomes

$$\mathrm{F}_{PH}(+) = \mathrm{F}_P(+) + \mathrm{F}'_H(+) + \mathrm{iF}''_H(+) \quad \text{for} \quad hkl \tag{7.68}$$

$$\mathrm{F}_{PH}(-) = \mathrm{F}_P(-) + \mathrm{F}'_H(-) + \mathrm{iF}''_H(-) \quad \text{for} \quad \bar{h}\bar{k}\bar{l} \tag{7.69}$$

F_P is the normal protein structure factor and $\mathrm{F}_P(+)$ is clearly equal to $\mathrm{F}_P(-)$, Fig. 7.39a.

The structure factors $\mathrm{F}_{PH}(+)$ and $\mathrm{F}_{PH}(-)$ are represented in Fig. 7.43, where $OC = \mathrm{F}_{PH}(+)$ and $OD = \mathrm{F}_{PH}(-)$. Clearly $|\mathrm{F}_{PH}(+)| \neq |\mathrm{F}_{PH}(-)|$. Figure 7.43 also shows the effect of reflecting $\mathrm{F}_{PH}(-)$ across the real axis OR. This device simplifies the following calculation. Using the cosine rule in triangles OBC and OBD', we obtain

$$\begin{aligned} |\mathrm{F}_{PH}(+)|^2 = {} & |\mathrm{F}_{PH}|^2 + |\mathrm{F}''_H|^2 - 2|\mathrm{F}_{PH}||\mathrm{F}''_H| \\ & \times \cos(\phi_{PH} - \phi_H + 90) \end{aligned} \tag{7.70}$$

$$\begin{aligned} |\mathrm{F}_{PH}(-)|^2 = {} & |\mathrm{F}_{PH}|^2 + |\mathrm{F}''_H|^2 - 2|\mathrm{F}_{PH}||\mathrm{F}''_H| \\ & \times \cos(90 - \phi_{PH} - \phi_H) \end{aligned} \tag{7.71}$$

Subtracting (7.71) from (7.70) gives

$$\begin{aligned} & |\mathrm{F}_{PH}(+)|^2 - |\mathrm{F}_{PH}(-)|^2 \\ & = [|\mathrm{F}_{PH}(+)| + |\mathrm{F}_{PH}(-)|][|\mathrm{F}_{PH}(+)| - |\mathrm{F}_{PH}(-)|] \\ & = -2|\mathrm{F}_{PH}||\mathrm{F}''_H|[\cos(\phi_{PH} - \phi_H + 90] - \cos(90 - \phi_{PH} - \phi_H)] \\ & = 4|\mathrm{F}_{PH}||\mathrm{F}''_H| \sin(\phi_{PH} - \phi_H) \end{aligned} \tag{7.72}$$

But, intuitively,

$$|\mathrm{F}_{PH}(+)| + |\mathrm{F}_{PH}(-)| = OC + OD' \approx 2OB (= 2|\mathrm{F}_{PH}|) \tag{7.73}$$

Thus, it follows that the anomalous difference

$$\Delta F_{ANO} = |\mathrm{F}_{PH}(+)| - |\mathrm{F}_{PH}(-)| \approx 2|\mathrm{F}''_H| \sin(\phi_{PH} - \phi_H) \tag{7.74}$$

For a given heavy-atom type, it is known that the ratio $|\mathrm{F}'_H|/|\mathrm{F}''_H| = f'/\Delta f''$, and is approximately constant, κ, say. Thus,

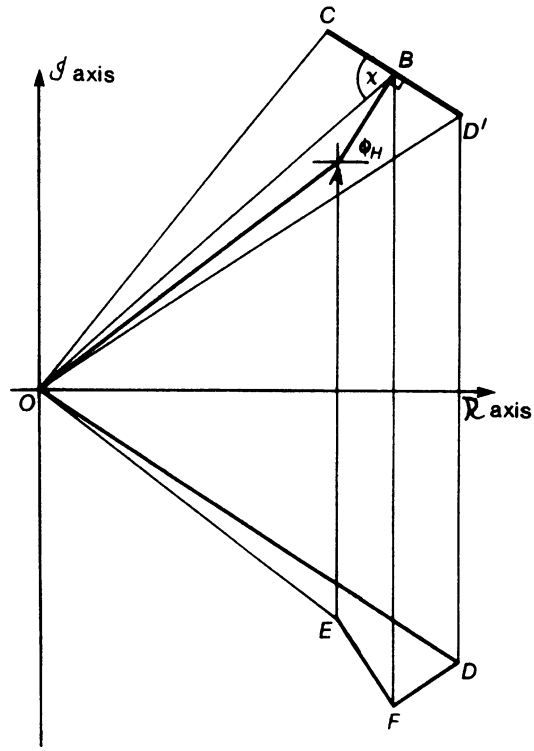


Fig. 7.43 Structure factors $F_{PH}(+)$ and $F_{PH}(-)$ for a protein crystal containing heavy-atom anomalous scattering species: $\widehat{ROC} = \phi_{PH}(+)$; $\widehat{ROD} = \widehat{ROD}$; $\widehat{ROB} = \phi_{PH}$; $\widehat{ROA} = \phi_P$; $\widehat{OAB} = \phi_P + (180 - \phi_H)$; $OA = |F_P(+)|$; $AB = |F_H(+)|$; $BC = |F''_H(+)|$; $OB = |F_{PH}|$; $OC = |F_{PH}(+)|$; $OD = |F_{PH}(-)|$; OD' = mirror reflection of OD; $OE = |F_P(-)|$; $EF = |F_H(-)|$; $FD = |F''_H(-)|$; $\widehat{OBC} = \chi = \phi_{PH} - \phi_H + 90$; $AC = AD' = ED = |F_H|$

$$\Delta F_{ANO} \approx (2|F'_H|/\kappa)\sin(\phi_{PH} - \phi_H)$$

or

$$|F'_H|\sin(\phi_{PH} - \phi_H) \approx (\kappa/2)\Delta F_{ANO} \tag{7.75}$$

From (7.44), we see that $\Delta F_{\mathrm{ISO}} \approx |F_H|\cos(\phi_H - \phi_{PH})$, and in the case of anomalous scattering $|F_H|$ becomes $|F'_H|$. It follows that $|F'_H| \approx [(\Delta F_{\mathrm{ISO}})^2 + (k/2)\Delta F_{\mathrm{ANO}})^2]^{1/2}$.

This provides a possible method for estimating $|F'_H|$ for calculation of a difference Patterson map when anomalous scattering measurements are available.

7.6.4 Phasing by Use of Anomalous Scattering

Anomalous scattering can be used in phasing reflections. We saw in the previous section that the isomorphous replacement technique in non-centrosymmetric crystals leads to an ambiguity in phase determination, Fig. 7.27. The ambiguity cannot be resolved unless the replaceable site is changed. Merely using a third derivative with the same replaceable site would lead to a situation comparable with that in Fig. 7.27. The heavy-atom vector would still be directed along OH, and its different length would be just balanced by the change in F_o, so that *three* circles would intersect at P_1 and P_2.

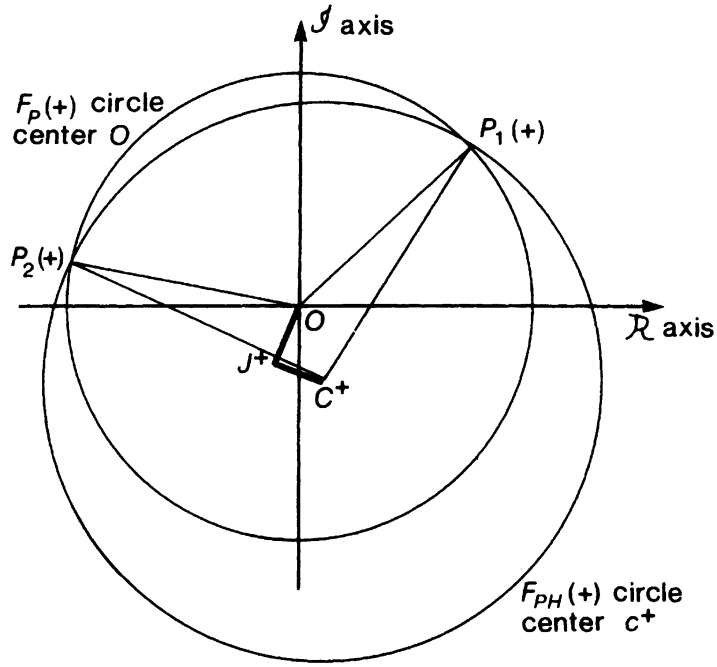


Fig. 7.44 Harker diagram for $F_P(+)(hkl)$ for a protein crystal with heavy atoms scattering anomalously. OJ^+ is the vector $-F'_H(+)$, and J^+C^+ is the imaginary component of OC^+, the vector for $-F_H(+)$. The $F_P(+)$ and $F_{PH}(+)$ *circles* intersect at $P_1(+)$ and $P_2(+)$. $OP_1(+)$ and $OP_2(+)$ are the ambiguous solutions for $F_P(+)$

A vector change in the heavy-atom contribution can be brought about through anomalous scattering in a given derivative, instead of invoking a different replaceable site. Following Fig. 7.39b we see that two different F_o values can arise for $\mathbf{h}$ and $\bar{\mathbf{h}}$. Consequently, $\phi(\mathbf{h}) \neq \phi(\bar{\mathbf{h}})$, and the ambiguity can be resolved by the experimental data. This technique is particularly important with synchrotron radiation, where the wavelength can be tuned to the absorption edge of a relatively heavy atom in the structure so as to obtain the maximum difference between $|F_o(\mathbf{h})|$ and $F_o(\bar{\mathbf{h}})$.

7.6.5 Resolution of the Phase Problem for Proteins Using Anomalous Scattering Measurements (SIRAS Method)[7]

A Harker diagram for any *hkl* reflection $F_P(+)$ may be constructed as discussed earlier by first drawing a circle of center O and radius $|F_P(+)|$. From C^+, the end of the vector $-F_H(+)$ as center, a second circle of radius $|F_{PH}(+)|$ is drawn, as shown in Fig. 7.44. It intersects the $F_P(+)$ circle in points $P_1(+)$ and $P_2(+)$; $OP_1(+)$ and $OP_2(+)$ represent the SIR phase ambiguity noted previously, Sect. 7.4.8.

The Harker diagram for the corresponding $\bar{h}\,\bar{k}\,\bar{l}$ reflection $F_P(-)$ is constructed by drawing a circle of radius $|F_P(-)|$, center O, and finally a circle of radius $|F_{PH}(-)|$, centered at C^-, the end of the value of $-F_H(-)$, as shown in Fig. 7.45. The two circles this time intersect in points $P_1(-)$ and $P_2(-)$, representing the ambiguous SIR solution for $F_P(-)$.

Since $F_P(+)$ and $F_P(-)$ are related by reflection across the real axis OR, Figs. 7.44 and 7.45, the *correct* solutions for $F_P(+)$ and $F_P(-)$ will be mirror-related in this way, while the incorrect pair will not. In the given case the corresponding correct solutions are $P_1(+)$ and $P_1(-)$ and $P_2(+)$ and $P_2(-)$ are the unwanted, erroneous solutions. In order to rationalize this process a combined Harker diagram can be conceived as in Fig. 7.46. This involves plotting the $F_{PH}(-)$ circle at the end of the *mirrored* $F_H(-)$ vector, thus enabling

[7] Single Isomorphous Replacement with Anomalous Scattering.

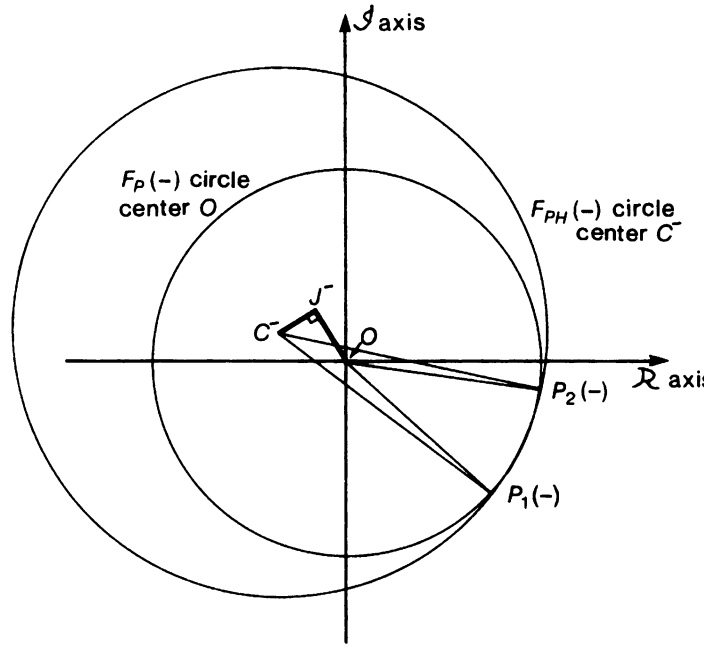


Fig. 7.45 Harker diagram for $F_P(-)(hkl)$ for a protein crystal with heavy atoms scattering anomalously. OJ^- is the vector for $-F_H(-)$, and J^-C^- is the imaginary component of OC^-, the vector for $-F_H(-)$. The $F_P(-)$ and $F_{PH}(-)$ *circles* intersect at $P_1(-)$ and $P_2(-)$. $OP_1(-)$ and $OP_2(-)$ are the ambiguous solutions for $F_P(-)$

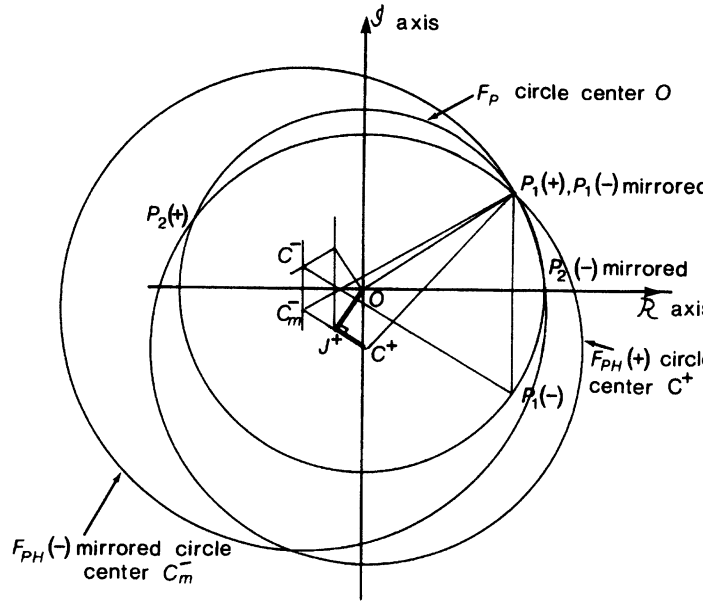


Fig. 7.46 Combined Harker diagram for $F_P(+)$ and $F_P(-)$, with $F_P(-)$ mirrored across the real axis of the Argand diagram: C_m^- is the mirrored point of C^-, where OC^- is the vector for $-F_H(-)$, as in Fig. 7.43. The correct solution for $F_P(+)$ is $OP_1(+)$, coinciding with $OP_1(-)$ mirrored

solutions $P_1(+)$ and $P_1(-)$ mirrored to coalesce. In practice, the phasing process is carried out by calculating a probability distribution in a similar manner to that used in the MIR technique. Algorithms for carrying out these computations have been developed.

Table 7.7 Values of the real (dispersion) $\Delta f'$ and imaginary (absorption) $\Delta f''$ components of anomalous scattering for Cu $K\alpha$ X-rays

		Cu $K\alpha$ radiation, $\lambda = 1.5418$ Å				
		$\Delta f'$		$\Delta f''$		
	Atomic number	$(\sin\theta)/\lambda = 0$	=0.6	$(\sin\theta)/\lambda = 0$	=0.4	μ[a]
C	6	0	0	0	0	4.6
N	7	0	0	0	0	7.52
O	8	0	0	0.1	0.1	11.5
S	16	0.3	0.3	0.6	0.6	89.1
Fe	26	−1.1	−1.1	3.4	3.3	308
Zn	30	−1.7	−1.7	0.8	0.7	60.3
Pd	46	−0.5	−0.6	4.3	4.1	206
Ag	47	−0.5	−0.6	4.7	4.5	218
I	53	−1.1	−1.3	7.2	6.9	294
Sm	62	−6.6	−6.7	13.3	12.8	397
Gd	64	−12	−12	12.0	11.6	439
Lu	71	−7	−7	5	5	153
Pt	78	−5	−5	8	7	200
Au	79	−5	−5	8	8	208
Hg	80	−5	−5	9	8	216
Pb	82	−4	−5	10	9	232
U	92	−4	−5	16	16	306

The atoms selected are those found in proteins and those used frequently for heavy-atom derivatives
[a] μ is the absorption coefficient (see Sect. 3.1.3)

7.6.6 Protein Phasing Using the Multiple-Wavelength Anomalous Dispersion Technique (MAD) with Synchrotron Radiation (SR)

The above treatment shows that, in principle, the phase ambiguity associated with the SIR technique can be resolved by incorporating anomalous dispersion measurements. Conventional laboratories are usually equipped with either sealed-tube or rotating-anode X-ray sources generating X-radiation from a copper target. Anomalous scattering effects for Cu $K\alpha$ radiation are quite small, Table 7.7, the actual differences between $|F(hkl)|$ and $|F(\bar{h}\,\bar{k}\,\bar{l})|$ being difficult to detect without extremely careful measurements. This practical limitation to the method may be overcome by the use of synchrotron radiation. Optimization of anomalous scattering information can be achieved by selection of a wavelength close to an absorption edge of the heavy atom, where $\Delta f''$ is a maximum. It should be remembered, however, that anomalous differences are still small, even for measurements made for wavelengths tuned in this way.

There is another advantage to be gained by the availability of SR radiation, namely that measurements can be made at different wavelengths, possibly even on the same crystal specimen. The second wavelength should be selected such that f' is large and f'' is small for the anomalous scattering species. Measurements for this wavelength would be made only for *hkl* reflections. This technique is known as the *multiple-wavelength anomalous dispersion* (MAD) method for phasing. The measurements required are $|F_P(hkl)|$, $|F_{PH}(+)|_{\lambda_1}$, $|F_{PH}(-)|_{\lambda_2}$, and $|F_{PH}(+)|_{\lambda_2}$; Fig. 7.47 shows an idealized phase diagram for a typical *hkl* reflection using the MAD technique.

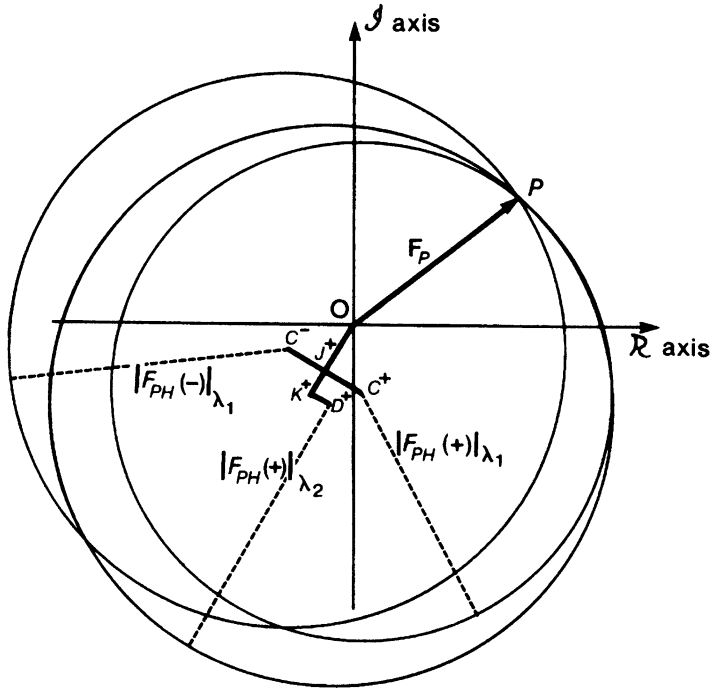


Fig. 7.47 Multiple-wavelength anomalous dispersion (MAD) phasing for wavelengths λ_1 and λ_2. The circles based on λ_1 for $|F_{PH}(+)|$ and $|F_{PH}(-)|$ are drawn as before. Adding the measurement at λ_2 of $|F_{PH}(+)|_{\lambda_2}$ is comparable to the information from a second heavy-atom derivative in the MIR technique. OK^+ is the real component and K^+D^+ the imaginary component of OD^+, the direction $F_H(+)_{\lambda_2}$. For clarity the F_P circle is not shown. F_P gives the MAD solution for the protein structure factor. The measurements are sometimes made only for the derivative crystal, using first λ_1 and then λ_2, during the same experimental session at the synchrotron radiation station. In such a case, the $|F_P|$ data may not be available, whereupon the phasing would be carried out for $|F_{PH}|$ instead. The ensuing electron density map would then apply to the derivative crystal

7.7 Charge Flipping

The determination of good phases may sometimes be achieved by a process known as *charge flipping*, which is based on an algorithm by Oszlányi and Sütü [31]. The process requires a set of experimental $|F_o|$ values, but no other prior information about the structure at this stage. It can be applied to both classical crystals and the aperiodic structures that were discussed briefly in Sect. 1.4.3, and with either single crystal or powder specimens.

It begins by assigning random phases to all values of $|F_o|$ in the data set. A Fourier transform of this phased data produces an electron density map $\rho_1(xyz)$ which will contain both positive and negative regions of density, and will most probably have little if any resemblance to a chemical structure. Then, a new density function $\rho_2(xyz)$ is derived, by inverting the signs of all density less than a chosen positive value, Δ; this is the charge flipping stage. The new function $\rho_2(xyz)$ is transformed to give $|F_c|$ values, which are discarded, and a set of phases. This set of phases is combined with the experimentally determined $|F_o|$ values so as to create $\rho_3(xyz)$. Again, the signs are reversed, and the iterative process is continued until a sensible chemical structure is obtained. The process is shown diagrammatically in Fig. 7.48.

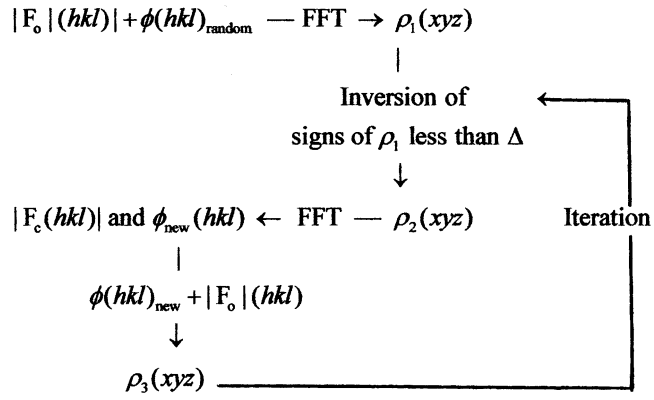


Fig. 7.48 Iterative process for phasing by charge flipping (FFT = fast Fourier transform)

The convergence of the process may be judged in the usual way, by both the chemical sense of the electron density map and the R factor. The procedure has a single free parameter, namely the charge flipping threshold, Δ, and no structural or symmetry information is needed. A number of structures has been solved by this technique [32–36], and there is an interesting applet available [33] which allows an interactive appreciation of the process.

In order to complete the structure determination, it is, of course, necessary to know the unit cell data and space group, since the structure will need to be refined by least squares in the usual way. It may be required also to have recourse to normal Fourier methods, if the phases produced by charge flipping do not lead to all atoms in the molecule. The charge flipping procedure is another useful tool in the crystallographer's armoury, and is also useful in studying incommensurate structures, such as those based on interpenetrating periodic lattices, and aperiodic structures [37]. The algorithm is incorporated into the SHELX and WinGX (Superflip) program systems.

7.8 Location of Hydrogen Atoms

In other sections of this book we have commented on the location of hydrogen atoms: by difference Fourier synthesis, Sect. 8.4.5, from molecular geometry, Sect. 9.2.4, and by neutron diffraction, Sect. 11.7.2. Here we consider a method by which hydrogen atoms can be located in hydrated ionic crystal structures from calculations of electrostatic energy.

The location of the positions of hydrogen atoms in crystalline hydrates can sometimes be carried out by a calculation based on electrostatic energy [38]. The coordinates at the positive end of the dipole in a water molecule may be located by a vector sum of the weighted electrical field strengths around the oxygen atom in the water molecule. All atoms less than 4 Å from the oxygen atom are included in calculating x' from (7.76); the O–H bond distance chosen was 0.99 Å and the H–O–H angle 104.5°.

$$x' = \sum_i (Z_i x_i / d_i^2) \tag{7.76}$$

where x_i is x-coordinate of the ith atom in the structure, distant d_i from the oxygen atom, and Z_i is its charge including the sign. For the ions in the structure, their formal charges were used; for the oxygen

Table 7.8 Hydrogen atom positions in $BaCl_2{\cdot}2H_2O$

	Experimental			By (7.76)		
	x	y	z	x	y	z
H_{11}	0.3534	0.0616	0.8824	0.381	0.064	0.874
H_{12}	0.4756	0.1255	0.0402	0.440	0.119	0.061
H_{21}	0.0977	0.1368	0.5967	0.082	0.154	0.586
H_{22}	0.2466	0.0654	0.4877	0.294	0.091	0.573

Table 7.9 Hydrogen atom positions in $NaBr{\cdot}2H_2O$

	Experimental			By (7.76)		
	x	y	z	x	y	z
H_{11}	0.269	0.029	0.635	0.298	0.052	0.657
H_{12}	0.129	−0.017	0.765	0.124	−0.001	0.744
H_{21}	0.861	0.221	0.725	0.804	0.222	0.753
H_{22}	0.741	0.098	0.792	0.772	0.101	0.882

and hydrogen atoms, theoretical calculations on a water molecule gave charges of −0.310 and +0.155 for Z_O and Z_H, respectively. Results for y' and z' follow similarly.

A number of possible positions for each hydrogen atom in the water molecule was generated in steps ϕ around the point x', y', z', such that all O–H distances were 0.99 Å. Madelung constants were calculated [39] for the complete structure with all possible hydrogen atoms locations and plotted against the step angle ϕ. The positions chosen for the hydrogen atoms were those corresponding to minimum electrostatic energy; about 90% of the energy is electrostatic in ionic hydrate structures.

Results were obtained for barium chloride dihydrate, Table 7.8, and sodium bromide dihydrate, Table 7.9.

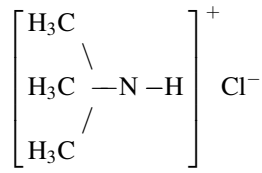
This procedure can be useful for structures in which heavy atoms are present, because the electrostatic contribution of a hydrogen atom in relation to the heavy atoms is significant. Subsequent to the location of the hydrogen atoms, their positions can be refined by least squares.

7.9 Problems

7.1. A structure with the apparent space group $P2_1/c$ consists of atoms at 0.2, $\frac{1}{4}$, 0.1 and the symmetry-related positions; the center of symmetry is at the origin. Evaluate the geometric structure factor for the four general equivalent positions in the unit cell, and, hence, determine the systematic absences among the *hkl* reflections. What are the consequences of these absences as far as the true structure is concerned? Sketch the structure in projection along *b*. What is the true space group?

7.2. Rh_2B crystallizes in space group *Pnma* with $a = 5.42$ Å, $b = 3.98$ Å, $c = 7.44$ Å, and $Z = 4$. Consider Fig. 2.36: show that if no two Rh atoms may approach within 2.5 Å of each other, they cannot lie in general positions. Where could the Rh atoms be placed? Illustrate your answer with a sketch showing possible positions for these atoms in projection on (010).

7.3. Trimethylammonium chloride,



crystallizes in a monoclinic, centrosymmetric space group, with $a = 6.09$ Å, $b = 7.03$ Å, $c = 7.03$ Å, $\beta = 95.73°$, and $Z = 2$. The only limiting condition is $0k0$: $k = 2n$. What is the space group? Comment on the probable positions of (a) Cl, (b) C, (c) N, and (d) H atoms.

7.4. Potassium hexachloroplatinate(IV), $K_2[PtCl_6]$, is cubic, with $a = 9.755$ Å. The atomic positions are as follows ($Z = 4$):

$$(0,0,0; \quad 0,\tfrac{1}{2},\tfrac{1}{2}; \quad \tfrac{1}{2},0,\tfrac{1}{2}; \quad \tfrac{1}{2},\tfrac{1}{2},0) +$$

4	Pt	0, 0, 0
8	K	$\frac{1}{4},\frac{1}{4},\frac{1}{4};\frac{3}{4},\frac{3}{4},\frac{3}{4}$
24	Cl	$\pm\{x, 0, 0; 0, x, 0; 0, 0, x\}$

Show that $|F_c(hhh)| = A'(hhh)$, where

$$A'(hhh) = 4g_{Pt} + 8g_K \cos(3\pi h/2) + 24g_{Cl} \cos 2\pi h x_{Cl}$$

Calculate $|F_c(hhh)|$ for the values of h tabulated below, with $x_{Cl} = 0.23$ and 0.24. Obtain R factors for the scaled F_o data for the two values of x_{Cl}, and indicate which value of x_{Cl} is the more acceptable. Calculate the Pt–Cl distance, and sketch the $[PtCl_6]^{-2}$ ion. What is the point group of this species?

hkl	111	222	333
F_o	491	223	281
g_{Pt}	73.5	66.5	59.5
g_K	17.5	14.5	12.0
g_{Cl}	15.5	13.0	10.5

Atomic scattering factors g_j may be taken to be temperature-corrected values.

7.5. USi crystallizes in space group $Pbnm$, with $a = 5.65$ Å, $b = 7.65$ Å, $c = 3.90$ Å, and $Z = 4$. The U atoms lie at the positions

$$\pm\{x, y, \tfrac{1}{4}; \quad \tfrac{1}{2} - x, \tfrac{1}{2} + y, \tfrac{1}{4}\}$$

Obtain a simplified expression for the geometric structure factor ($\bar{1}$ at 0, 0, 0) for the U atoms. From the data below, determine approximate values for x_U and y_U; the Si contributions may be neglected.

hkl	200	111	210	231	040	101	021	310
$I_o(hkl)$	0	236	251	200	0	170	177	0

Proceed by using 200 to find a probable value for x_U. Then find y_U from 111, 231, and 040.

7.6. Methylamine forms a complex with boron trifluoride of composition $CH_3NH_2BF_3$.
Crystal Data
System: monoclinic
Unit-cell dimensions: $a = 5.06$ Å, $b = 7.28$ Å, $c = 5.81$ Å, $\beta = 101.5°$
V_c: 209.7 Å^3
D_m: 1.54 g cm^{-3}
M_r: 98.86
Z: 1.97 or 2 to the nearest integer
Unit-cell contents: 2C, 10H, 2N, 2B, and 6F atoms
Absent spectra: $0k0 : k = 2n + 1$
Possible space groups: $P2_1$ or $P2_1/m$ ($P2_1/m$ may be assumed)
Determine what you can about the crystal structure.

7.7. Write the symmetry-equivalent amplitudes of |F(hkl)|, |F($0kl$)|, and |F($h0l$)| in (a) the triclinic, (b) the monoclinic, and (c) the orthorhombic crystal systems; Friedel's law may be assumed.

7.8. (a) Determine the orientations of the Harker lines and sections in Pa, $P2/a$, and $P222_1$.
(b) A monoclinic, non-centrosymmetric crystal with a primitive space group shows concentrations of peaks on (u, 0, w) and [0, v, 0]. How might this situation arise?

7.9. Diphenyl sulfoxide, $(C_6H_5)_2SO$, is monoclinic, with $a = 8.90$ Å, $b = 14.08$ Å, $c = 8.32$ Å, $\beta = 101.12°$, and $Z = 4$. The conditions limiting possible X-ray reflections are as follows.

$$hkl : none;\ h0l : h + l = 2n;$$
$$0k0 : k = 2n$$

(a) What is the space group?
(b) Figure P7.1a–c are Patterson sections at $v = \frac{1}{2}$, 0.092, and 0.408, respectively, and contain S–S vector peaks. Write the coordinates of the non-origin S–S vectors in terms of x, y, and z, and from the sections provided determine the best values for the S atoms in the unit cell. Plot these atomic positions as seen along the b axis, with an indication of the heights of the atoms with respect to the plane of the diagram.

7.10. Figure P7.2 shows an idealized u, w vector set for a hypothetical structure C_6H_5S in space group $P2$ with $Z = 2$, projected down the b axis. Only the S–S and S–C vector interactions are considered.
(a) Determine the x and z coordinates for the S atoms and plot them to the scale of this projection.
(b) Use the Patterson superposition method to locate the carbon atom positions on a map of the same projection.

7.11. Hafnium disilicide, $HfSi_2$, is orthorhombic, with $a = 3.677$ Å, $b = 14.55$ Å, $c = 3.649$ Å, and $Z = 4$. The space group is $Cmcm$, and the Hf and Si atoms occupy three sets of special positions of the type

$$\pm\{0, y, \tfrac{1}{4};\quad \tfrac{1}{2}, \tfrac{1}{2} + y, \tfrac{1}{4}\}$$

The contributions from the Hf atoms dominate the structure factors. By combining the terms $\cos 2\pi ky$ and $\cos 2\pi(ky + k/2)$, show that the geometric structure factor $A(0k0)$ is approximately proportional to $\cos 2\pi y_{Hf}$. The $F_o(0k0)$ data are listed below, from which the values of $F_o(0k0)^2$, divided by 10 and rounded to the nearest integer, have been derived.

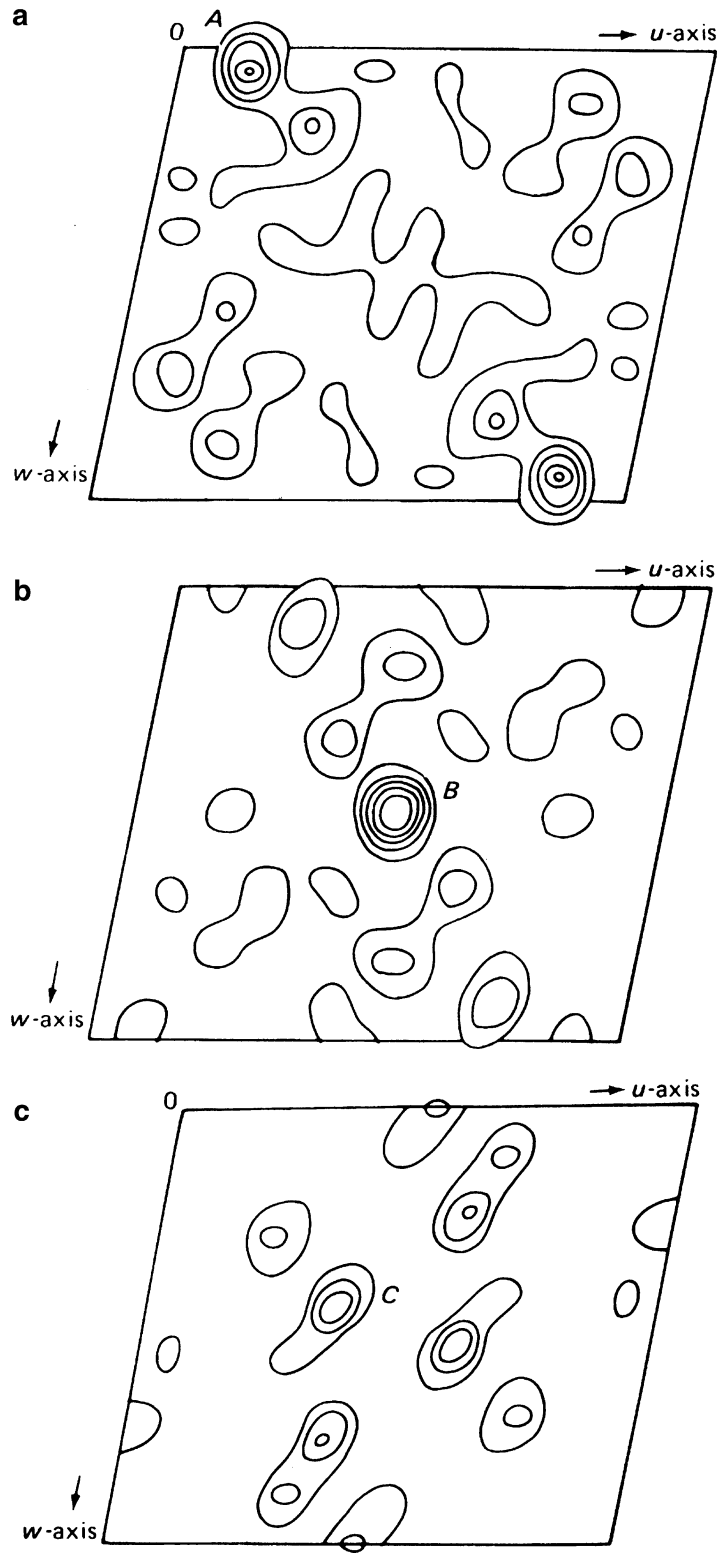


Fig. P7.1 Patterson sections at (**a**) $v = \frac{1}{2}$, (**b**) $v = 0.092$, (**c**) $v = 0.408$

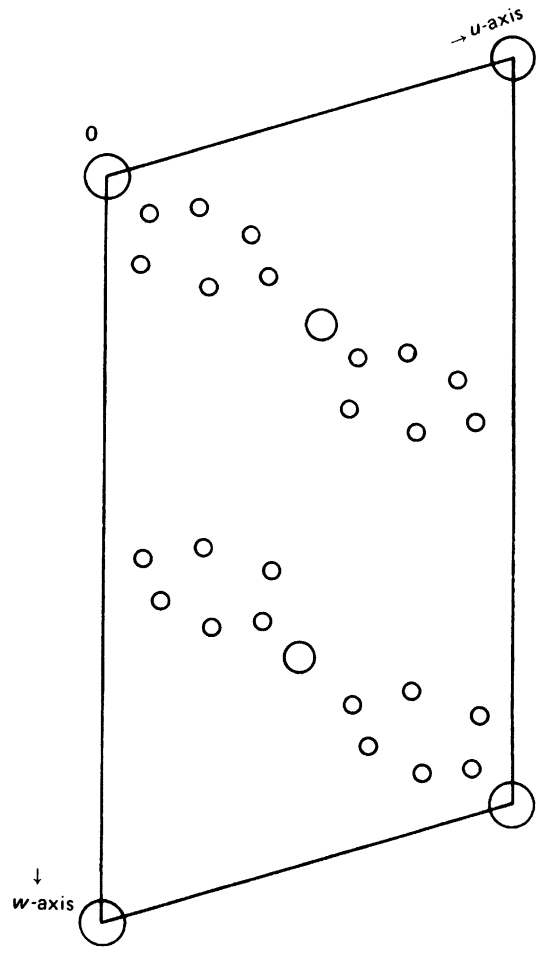


Fig. P7.2 Idealized Vector map for C_6H_4S

$0k0$	020	040	060	080	010,0	012,0	014,0	016,0
$F_o(0k0)$	7	14	18	13	12	<1	20	<1
$F_o(0k0)^2$	5	20	32	17	14	0	40	0

(a) Calculate the one-dimensional Patterson function $P(v)$, using the equation

$$P(v) \propto \sum_k F_o(0k0)^2 \cos 2\pi kv$$

The multiplying factor $2/b$ and the F(0) term have been omitted[8] to simplify the calculation; they can never change the form of the synthesized function, although the neglect of the term involving F(0) gives rise to negative values in the calculated $P(v)$. The Fourier summation here can be carried out readily by means of the program FOUR1D, which also gives a plot of the function at the online printer.

[8] These omissions give rise to the proportionality sign.

Table P7.1 Simulated Beevers-Lipson strips appropriate to problem 7.11

	Amplitude	Index	$\frac{0}{60}$	$\frac{1}{60}$	$\frac{2}{60}$	$\frac{3}{60}$	$\frac{4}{60}$	$\frac{5}{60}$	$\frac{6}{60}$	$\frac{7}{60}$	$\frac{8}{60}$	$\frac{9}{60}$	$\frac{10}{60}$	$\frac{11}{60}$	$\frac{12}{60}$	$\frac{13}{60}$	$\frac{14}{60}$	$\frac{15}{60}$
Add columns for $\sum_k$	5	2	5	5	5	4	3	2	2	1	$\bar{1}$	$\bar{2}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{5}$	$\bar{5}$
	20	4	20	18	13	6	$\bar{2}$	$\overline{10}$	$\overline{16}$	$\overline{20}$	$\overline{20}$	$\overline{16}$	$\overline{10}$	$\bar{2}$	6	13	18	20
	32	6	32	26	10	$\overline{10}$	$\overline{26}$	$\overline{32}$	$\overline{26}$	$\overline{10}$	10	26	32	26	10	$\overline{10}$	$\overline{26}$	$\overline{32}$
	17	8	17	11	$\bar{2}$	$\overline{14}$	$\overline{17}$	8	5	16	16	5	$\bar{8}$	$\overline{17}$	$\overline{14}$	$\bar{2}$	11	17
	14	10	14	7	$\bar{7}$	$\overline{14}$	$\bar{7}$	7	14	7	$\bar{7}$	$\overline{14}$	$\bar{7}$	7	14	7	$\bar{7}$	$\overline{14}$
	40	14	40	4	$\overline{39}$	$\overline{12}$	37	20	$\overline{32}$	$\overline{27}$	27	32	$\overline{20}$	$\overline{32}$	12	39	$\bar{4}$	$\overline{40}$
			128	71	$\overline{20}$...													

Table P7.2 Simulated Beevers–Lipson strips

Amplitude	Index	$\frac{0}{60}$	$\frac{1}{60}$	$\frac{2}{60}$	$\frac{3}{60}$	$\frac{4}{60}$	$\frac{5}{60}$	$\frac{6}{60}$	$\frac{7}{60}$	$\frac{8}{60}$	$\frac{9}{60}$	$\frac{10}{60}$	$\frac{11}{60}$	$\frac{12}{60}$	$\frac{13}{60}$	$\frac{14}{60}$	$\frac{15}{60}$
7	2	7	7	6	6	5	3	2	1	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{5}$	$\bar{6}$	$\bar{6}$	$\bar{7}$	$\bar{7}$
14	4	14	13	9	4	$\bar{1}$	$\bar{7}$	$\overline{11}$	$\overline{14}$	$\overline{14}$	$\overline{11}$	$\bar{7}$	$\bar{1}$	4	9	13	14
18	6	18	15	6	$\bar{6}$	$\overline{15}$	$\overline{18}$	$\overline{15}$	$\bar{6}$	6	15	18	15	6	$\bar{6}$	$\overline{15}$	$\overline{18}$
13	8	13	9	$\bar{1}$	$\overline{11}$	$\overline{13}$	$\bar{6}$	4	12	12	4	$\bar{6}$	$\overline{13}$	$\overline{11}$	$\bar{1}$	9	13
12	10	12	6	$\bar{6}$	$\overline{12}$	$\bar{6}$	6	12	6	$\bar{6}$	$\overline{12}$	$\bar{6}$	6	12	6	$\bar{6}$	$\overline{12}$
20	14	20	2	$\overline{20}$	$\bar{6}$	18	10	$\overline{16}$	$\overline{13}$	13	16	$\overline{10}$	$\overline{18}$	6	20	$\bar{2}$	$\overline{20}$

However, for many years following their introduction in 1936, the summation aid known as Beevers–Lipson[9] strips was used for Fourier summations in one and two dimensions, and even Harker sections in three dimensions. For historical interest, Table P7.1 shows the strips that would be used for this summation. Each line contains the value of the positive coefficients $(F_o^2/10)\cos 2\pi h(n/60)$ for $n = 0$ to 15/60; for negative |F| the sign of all terms for $n = 0$–15 are changed. For reflection symmetry at $\frac{1}{4}$, only the values of n from 0 to 15 are needed. The columns are added vertically to give the sum over h at each value of n. The range of n can be increased by making use of the properties of the cosine function.

Table P7.2 shows clearly how an error is of greater consequence in a phase (sign) than in an amplitude. Consider changing the sign of 080 in the calculation of $\rho(y)$ and then the amplitude by, say, 20%.

Plot the function, extend it to one repeat unit, interpret the four highest non-origin peaks, and determine y_{Hf}.

(b) Use the value of y_{Hf} and the form of the geometric structure factor $A(0k0)$ to determine the signs for the $0k0$ reflections. Hence, compute the electron density:

$$\rho(y) \propto \sum_k \pm F_o(0k0)| \cos 2\pi ky$$

Again the $2/b$ factor and F(0) have been omitted. Plot the function and determine y_{Hf}. What can be deduced about the positions of the Si atoms? In the light of your results, study $P(v)$ again. Table P7.2 contains the simulated, relevant Beevers–Lipson strips, with *positive* values of |F|.

[9] Lipson H, Beevers CA (1936) Proc Phys Soc 48:772.

Table P7.3 $F_o(hhh)$ for isomorphous alums

hkl	NH_4^+ (10 electrons)	K^+ (18 electrons)	Rb^+ (36 electrons)	Tl^+ (80 electrons)
111	86	38	19	113
222	0	19	79	195
333	111	125	158	236
444	25	6	55	125
555	24	49	64	131
666	86	86	122	164
777	53	34	0	18
888	0	16	22	56

7.12. The alums, $M\mathrm{Al}(N\mathrm{O}_4)_2{\cdot}12\mathrm{H}_2\mathrm{O}$, where $M = \mathrm{NH}_4$, K, Rb, Tl, and $N = $ S, Se, are isomorphous. They crystallize in the cubic centrosymmetric space group $Pa3$, with the unit-cell side a in the range 12.2–12.4 Å and $Z = 4$.

A symmetry analysis leads to the following atomic positions:

4*M*	$0,0,0;\ 0,\frac{1}{2},\frac{1}{2};\ \frac{1}{2},0,\frac{1}{2};\ \frac{1}{2},\frac{1}{2},0$
4Al	$\frac{1}{2},\frac{1}{2},\frac{1}{2};\ \frac{1}{2},0,0;\ 0,\frac{1}{2},0;\ 0,0,\frac{1}{2}$
8*N*	$\pm\left\{x,x,x;\ \frac{1}{2}+x,\frac{1}{2}-x,\bar{x};\ \frac{1}{2}+x,\frac{1}{2}-x;\ \frac{1}{2}-x,\bar{x},\frac{1}{2}+x\right\}$

The N atoms lie on cube diagonals, and x_N may be obtained by a one-dimensional Fourier synthesis along the line [111], using F(*hhh*) data. Table P7.3 lists these data for four alums ($N = S$). Tl may be assumed to be sufficiently heavy to make all F_o values positive in this derivative. The same sites in each crystal are occupied by the replaceable atoms.

(a) Use the isomorphous replacement technique to determine the signs of the reflections in Table P7.3.

(b) Compute $\rho[111]$ for K alum, using the following equation:

$$\rho(D) \propto \sum_h \pm F_o(hhh)\cos 2\pi hD$$

where D is the sampling interval along [111]. Plot the function and determine a probable value for d_s.

(c) The corresponding *hhh* data for the isomorphous K/Se alum are listed below. The signs have been allocated by a similar isomorphous replacement procedure. Calculate and plot $\rho(D)$ for these data. Compare the two electron density plots and comment upon the results.

hkl	111	222	333	444	555	666	777	888
F_o	−48	−52	64	0	116	100	−16	0

7.13. A crystal contains five atoms per unit cell. Four of them contribute together $100e^{i\phi}$ to F(010). The fifth atom has fractional coordinates 0.00, 0.10, 0.00, and its atomic scattering factor components f_o, $\Delta f'$, and $\Delta f''$ are 52.2, −2.7, and 8.0, respectively. If $\phi = 60°$, determine, graphically or otherwise, $|F(010)|$, $|F(0\bar{1}0)|$, $\phi(010)$, and $\phi(0\bar{1}0)$.

7.14. A protein crystal structure is to be solved using MIR. Three isomorphous derivatives are prepared using platinum, uranium, and iodine compounds. For the reflection 060, the following measurements were recorded:

For protein: $|F_P| = 858$
For Pt derivative: $|F_{PH_1}| = 756$, $|F_{H_1}| = 141$, $\phi_{H_1} = 78°$
U derivative: $|F_{PH_2}| = 856$, $|F_{H_2}| = 154$, $\phi_{H_2} = 63°$
I derivative: $|F_{PH_3}| = 940$, $|F_{H_3}| = 100$, $\phi_{H_3} = 146°$
Use a Harker construction to obtain an estimate for ϕ_P for this reflection from the native protein crystal.

7.15. Consider (7.19), reproduced here:

$$\Psi(x) = \frac{\pi}{2} + 2\sum_{h=1}^{\infty} \frac{1}{h} \sin hx$$

Show that sin hx can be replaced by cos $(hx - \phi)$, where $\phi = \pi/2$.

7.16. A protein with molecular weight 18000 Da crystallizes in space group $C2$ with unit-cell dimensions $a = 40$ Å, $b = 50$ Å, $c = 60$ Å, $\beta = 100°$.
(a) Estimate the number of protein molecules per unit cell if there are equal masses of protein and solvent in the unit cell (take M_H as 1.66×10^{-24} g).
(b) What symmetry would the protein molecule need to adopt in the crystalline state?

7.17. A non-centrosymmetric structure is composed entirely of N_H identical heavy atom anomalous scattering species per unit cell; it may be assumed here that no normal scattering species is present. Show, graphically or otherwise, that $|F(hkl)| = |F(\bar{h}\,\bar{k}\,\bar{l})|$ but that $\phi(hkl) \neq -\phi(\bar{h}\,\bar{k}\,\bar{l})$.

7.18. A centrosymmetric structure contains a mixture of anomalous and normal scattering atoms. Show that $|F(hkl)|^2 = |F(\bar{h}\,\bar{k}\,\bar{l})|^2$ and $\phi(hkl) = \phi(\bar{h}\,\bar{k}\,\bar{l}) \neq 0$ or π.

7.19. Given the $|F(hkl)|$ equivalents for each of the following space groups, list corresponding Bijvoet pairs:
(a) $C2$: $|F(hkl)| = |F(\bar{h}k\bar{l})| = |F(hk\bar{l})| = |F(\bar{h}\,\bar{k}\,\bar{l})|$
(b) Pm: $|F(hkl)| = |F(\bar{h}k\bar{l})| = |F(hk\bar{l})| = |F(\bar{h}\,\bar{k}\,\bar{l})|$
(c) $P2_12_12_1$:
$|F(hkl)| = |F(\bar{h}kl)| = |F(h\bar{k}l)| = |F(hk\bar{l})| = |F(\bar{h}\bar{k}l)| = |F(\bar{h}k\bar{l})| = |F(h\bar{k}\bar{l})| = |F(\bar{h}\,\bar{k}\,\bar{l})|$
(d) $P4$:
$|F(hkl)| = |F(\bar{k}hl)| = |F(\bar{h}\,\bar{k}\,l)| = |F(k\bar{h}l)| = |F(kh\bar{l})| = |F(hk\bar{l})| = |F(\bar{k}h\bar{l})| = |F(\bar{h}\,\bar{k}\,\bar{l})|$

7.20. X-ray intensity data are to be measured for an orthorhombic crystalline protein in three stages:
(a) For $0 < \theta < 10°$, both $I(hkl)$ and $I(\bar{h}\,\bar{k}\,\bar{l})$ symmetry-equivalent reflections are measured
(b) For $10 < \theta < 20°$, $I(hkl)$ alone are measured
(c) For $20 < \theta < 25°$, $I(hkl)$ alone are measured
If the unit cell is primitive, with $a = 30$ Å, $b = 50$ Å and $c = 40$ Å, and the X-ray wavelength is 1.5 Å, estimate the number of reflections measured and the corresponding resolution for the three data sets.

References

1. Reynolds CD et al (1974) J Cryst Mol Struct 4:213
2. Cruickshank DWJ (1957) Acta Crystallogr 10:504 [This study used the non-standard space group $P2_1/a$, equivalent to $P2_1/c$, with **a** and **c** interchanged and **b** changed to $-$**b**]
3. Keeling RO (1957) Acta Crystallogr 10:209
3a. Glusker JP et al (1968) Acta Crystallogr B24:359
4. Ladd M, Palmer R (2003) Structure determination by X-ray crystallography, 4th edn. Kluwer Academic/Plenum, New York
5. Palmer HT, Palmer RA (1969) Acta Crystallogr B25:1090
6. Duwell EJ, Baenziger NC (1995) Acta Crystallogr 8:705

7. Woolfson MM (1956) Acta Crystallogr 9:804
8. Sim GA (1960) Acta Crystallogr 13:511
9. Brünger AT (1992) Nature 355:472
10. Phenix Tutorial8. http://www.phenix-online.org/documentation/refinement-examples.htm
11. Wilson AJC (1976) Acta Crystallogr A32:994
12. Cruickshank DWJ (1949) Acta Crystallogr 2:154
13. idem. ibid. (1999) D55:583
14. Blow DM (2002) Acta Crystallogr D58:792
15. Ladd MFC (1967) Z Kristallogr 124:64
16. Wilson AJC (1950) Acta Crystallogr 3:397–398
17. Rossman MG, Blow DM (1961) Acta Crystallogr 14:641
18. Blow DM, Crick FHC (1959) Acta Crystallogr 12:794
19. Dickerson RE, Weintzierl JE, Palmer RA (1968) Acta Crystallogr B24:997
20. Hamilton WC (1965) Acta Crystallogr 18:502
21. Rogers D (1981) Acta Crystallogr A37:734
22. Flack HD (1983) Acta Crystallogr A39:876
23. http://en.wikipedia.org/wiki/Stereochemistry
24. Palmer RA, Potter BS, Palmer HT, Frampton CS (2007) J Chem Cystallogr 37:1
25. Palmer RA, Potter BS, Leach MJ, Jenkins TC, Chowdhry BZ (2010) Med Chem Commun 1:45
26. Barnes CL (1997) J Appl Crystallogr 30:568 [Based on ORTEP-III (v 1.0.3) by CK Johnson and MN Burnett]
27. Merrit EA, Bacon DJ (1997) Methods Enzymol 277:505 [Implemented in WinGX (qv)]
28. Tanczos AC, Palmer RA, Potter BS, Saldanha JW, Howlin BJ (2004) Comput Biol Chem 28(5–6):375
29. Spencer J et al (2012) CrystEngComm 14(20), 6441
30. Peerdeman AF, Bijvoet JM (1956) Acta Crystallogr 9:1012
31. Oszlányi G and Sütü A (2004) Acta Cryst A60:134
32. Palatinus L (2004) Acta Cryst A60:604
33. http://escher.epfl.ch/eCrystallography/ (Sect. 7.4); http://user.web.psi.ch/powder08/downloads/Hinrichsen-Charge%20flipping.pdf
34. Baerlocher Ch et al (2007) Science 315:1113
35. McCusker LB, Baerlocher Ch (2007) Zeit für Kristallogr 222:47
36. Sisak D et al (2012) J Appl Crystallogr (in press)
37. Van Smaalen S (2007) Incommensurate Crystallography, Oxford University Press
38. Ladd MFC (1968) Z Kristallogr 126:147
39. Ladd MFC (1998) Introduction to physical chemistry, 3rd edn. Cambridge University Press, Cambridge

Bibliography: Introductory Structure Analysis

Bragg WL (1949) A general survey, vol I, The crystalline state. Bell, London
Hahn T (ed) (2002) International tables for crystallography, vol A, 5th edn. Kluwer Academic, Dordrecht
Henry NFM, Lonsdale K (eds) (1965) International tables for X-ray crystallography, vol I. Kynoch Press, Birmingham

General Structure Analysis

Buerger MJ (1959) Vector space. Wiley, New York
Dunitz JD (1995) X-ray analysis and the structure of organic molecules. Verlag Helvetica Chimica Acta, Basel
Woolfson MM (1997) An introduction to X-ray crystallography, 2nd edn. Cambridge University Press, Cambridge

Protein Crystallography

Blundell TL, Johnson LN (1977) Protein crystallography. Academic, New York
Eisenberg D (1970) in The Enzymes, vol I, 3rd edition (ed Boyer PD), Academic, New York
Phillips DC (1966) Adv Struct Res Diffraction Methods 2:75

Crystallographic Computing

Ahmed FR et al (eds) (1970) Crystallographic computing. Munksgaard, Copenhagen
SHELX-97 manual. http://shelx.uni-ac.gwdg.de/SHELX/shelx.pdf
CCP4. http://www.ccp4.ac.uk/download/downloadman.php

Chemical Data

Sutton LE (ed) (1958; suppl 1965) Tables of interatomic distances and configuration in molecules and ions. The Chemical Society, London