DIFFRACTION PATTERNS AND TRIAL STRUCTURES

Part II

This page intentionally left blank

The diffraction pattern obtained

5

In this chapter we will describe those factors that control the intensities of Bragg reflections and how to express them mathematically so that we can calculate an electron-density map. The Bragg reflections have intensities that depend on the arrangement of atoms in the unit cell and how X rays scattered by these atoms interfere with each other. Therefore the diffraction pattern has a wide variety of intensities in it (see Figure 3.8a). Measured X-ray diffraction data consist of a list of the relative intensity $I(hkl)$, its indices (h, k, and l), and the scattering angle 2θ (see Chapter 4), for each Bragg reflection. All the values of the intensity $I(hkl)$ are on the same relative scale, and this entire data set describes the "diffraction pattern." It is used as part of the input necessary to determine the crystal structure.

As already indicated from a study of the diffraction patterns from slits and from various arrangements of molecules (Figures 3.1 and 3.9 especially), the angular positions (2θ) at which scattered radiation is observed depend only on the dimensions of the crystal lattice and the wavelength of the radiation used, while the intensities $I(hkl)$ of the different diffracted beams depend mainly on the nature and arrangement of the atoms within each unit cell. It is these two items, the unit-cell dimensions of the crystal and its atomic arrangement, that comprise what we mean by "the crystal structure." Their determination is the primary object of the analysis described here.

Representation of the superposition of waves

As illustrated in Figure 1.1b and the accompanying discussion, and mentioned again at the start of Chapter 3, X rays scattered by the electrons in the atoms of a crystal cannot be recombined by any known lens. Consequently, to obtain an image of the scattering matter in a crystal, the "structure" of that crystal, we need to simulate this recombination, which means that *we must find a way to superimpose the scattered waves, with the proper phase relations between them, to give an image of the material that did the scattering, that is, the electrons in the atoms.* We call this image

an "electron-density map." It shows approximately zero values at sites in the unit cell where there are no atoms, and positive values at sites of atoms. The electron-density values are higher for heavier atoms than for lighter atoms (an effect expressing the number of electrons in each atomic nucleus) so that this electron-density map may permit discrimination between atoms that have different atomic numbers.

How then can the superposition of waves be represented? There are several ways. Electromagnetic waves, such as X rays, may be regarded as composed of many individual waves. When this radiation is scattered with preservation of the phase relationships among the scattered waves, the amplitude of the resultant beam in any direction may be determined by summing the individual waves scattered in that direction, taking into account their relative phases (see Figure 3.2). We use a cosine wave (or a sine wave, which differs from it by a phase change of $\pi/2$ radians or 90°). The phase for this cosine wave may be calculated by noting the position of some point on it, such as a maximum. This is measured relative to an arbitrarily chosen origin (see Figures 1.2 and 5.1a).

There are several ways of representing electromagnetic waves so that they can be summed to give information on the nature of the combined wave.

Graphical representation

The usual way to represent electromagnetic waves graphically is by means of a sinusoidal function. Unfortunately, graphical superposition of waves of the type illustrated in Figure 3.2 is not convenient with a digital computer. Therefore, for speed and convenience in computing, other representations are preferred.

Algebraic representation

When we represent a wave by a trigonometric (cosine) function, we use the following algebraic expressions for the vertical displacements (x_1 or x_2) of two waves at a particular moment in time:

$$x_1 = c_1 \cos(\phi + \alpha_1) \tag{5.1}$$

$$x_2 = c_2 \cos(\phi + \alpha_2) \tag{5.2}$$

Here c_1 and c_2 are the amplitudes of the two waves (their maximum vertical displacements). The value of ϕ is, at a given instant, proportional to the time (or distance) for the traveling wave and is the same for all waves under consideration; α_1 and α_2 are the phases, expressed relative to an arbitrary origin. We will assume here that the wavelengths of the scattered waves are identical, inasmuch as the X rays used in structure analyses are generally monochromatic (only one wavelength). Because the wavelengths are the same, the phase difference between the two scattered waves ($\alpha_1 - \alpha_2$), remains constant (assuming that no change in the phase of either wave has taken place during scattering).

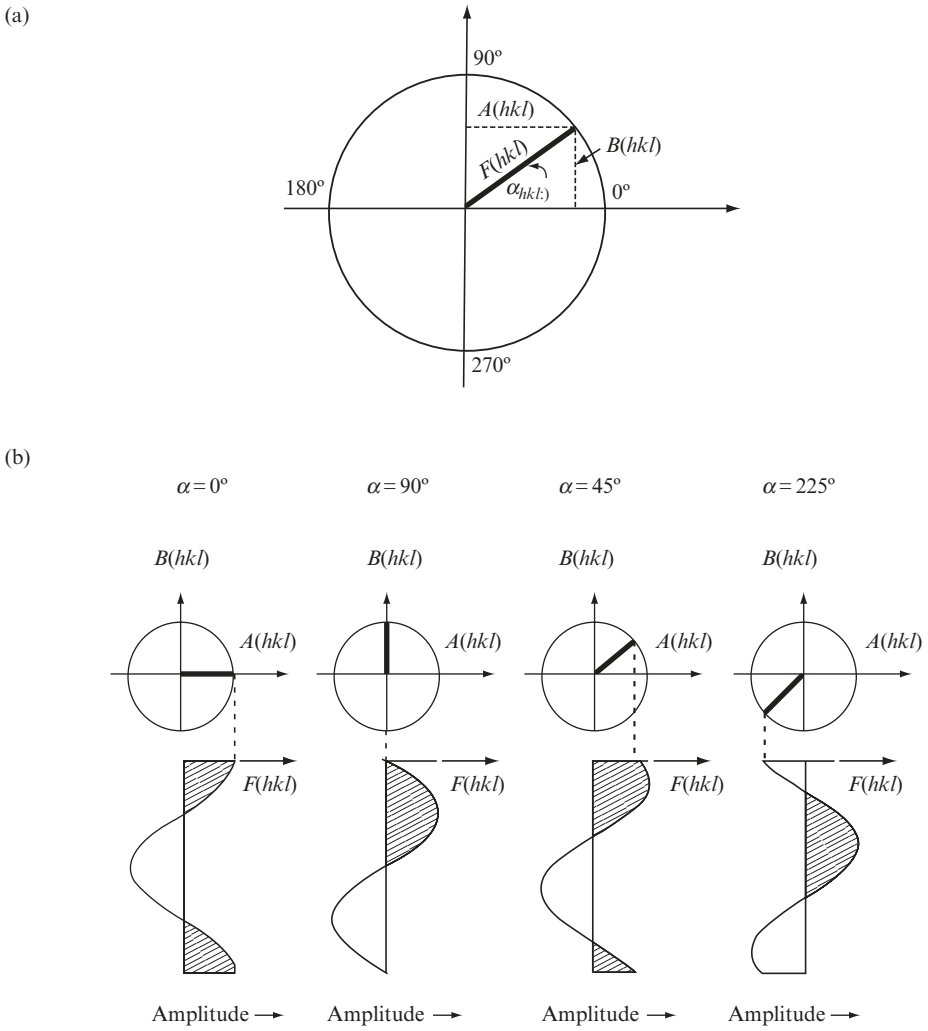


Fig. 5.1 The representation of sinusoidal waves.

(a) Graphical representation of a sinusoidal cosine wave, amplitude $F(hkl)$ represented by the radius of the circle, and phase α_{hkl} represented by the counterclockwise angle measured at the center of the circle. (b) Four examples of a phase angle, represented as shown in (a), and the cosine wave it represents. Note the relationship of the phase angle in the circular representation to the location of the peak of the cosine wave.

When the waves are superimposed, the resulting displacement, x_r, is, at any time, simply the sum of the individual displacements, as shown earlier in a graphical manner in Figure 3.2:

$$x_r = x_1 + x_2 = c_1 \cos(\phi + \alpha_1) + c_2 \cos(\phi + \alpha_2) \tag{5.3}$$

which, since $\cos(A + B) = \cos A \cos B - \sin A \sin B$, may be rewritten as

$$x_r = c_1 \cos\phi \cos\alpha_1 - c_1 \sin\phi \sin\alpha_1 + c_2 \cos\phi \cos\alpha_2 - c_2 \sin\phi \sin\alpha_2 \tag{5.4}$$

or

$$x_r = (c_1 \cos \alpha_1 + c_2 \cos \alpha_2) \cos \phi - (c_1 \sin \alpha_1 + c_2 \sin \alpha_2) \sin \phi \tag{5.5}$$

If we define the amplitude, c_r, and phase, α_r, of the resulting wave such that

$$c_r \cos \alpha_r = c_1 \cos \alpha_1 + c_2 \cos \alpha_2 = \sum_j c_j \cos \alpha_j \tag{5.6}$$

and

$$c_r \sin \alpha_r = c_1 \sin \alpha_1 + c_2 \sin \alpha_2 = \sum_j c_j \sin \alpha_j \tag{5.7}$$

then we can rewrite Eqn. (5.5) as

$$x_r = c_r \cos \alpha_r \cos \phi - c_r \sin \alpha_r \sin \phi = c_r \cos(\phi + \alpha_r) \tag{5.8}$$

Thus the resultant of adding two waves of equal wavelength is a wave of the same frequency, with a phase α_r (relative to the same origin) given by Eqns. (5.6) and (5.7) or, more compactly, by the following equation:

$$\tan \alpha_r = \frac{c_r \sin \alpha_r}{c_r \cos \alpha_r} = \frac{\sum_j c_j \sin \alpha_j}{\sum_j c_j \cos \alpha_j} \tag{5.9}$$

The amplitude of the resultant wave, c_r, is given by

$$\begin{aligned} c_r &= [(c_r \cos \alpha_r)^2 + (c_r \sin \alpha_r)^2]^{1/2} \\ &= \left[\left(\sum_j c_j \cos \alpha_j \right)^2 + \left(\sum_j c_j \sin \alpha_j \right)^2 \right]^{1/2} \end{aligned} \tag{5.10}$$

Vectorial representation

These relationships can all be expressed alternatively in terms of two-dimensional vectors, as illustrated in Figures 5.1a and b. You will remember that a vector has a magnitude (measure), direction (angle from the horizontal), and sense (where it starts and ends) (see Glossary). The length of the jth vector is its amplitude, c_j, and the angle that it makes with the arbitrary zero of angle (usually taken as the direction of the horizontal axis pointing to the right, with positive angles measured counterclockwise) is the phase angle α_j. This is shown in Figures 5.1a and 5.2a, where c_j is represented as $F(hkl)$, the structure factor. The components of the vectors along orthogonal axes are just $A = c_j \cos \alpha_j$ and $B = c_j \sin \alpha_j$ and the components of the vector resulting from addition of two (or more) vectors are just the sums of the components of the individual vectors making up the sum, a result expressed in Eqns. (5.6) and (5.7). The relationship of the vector representation of a

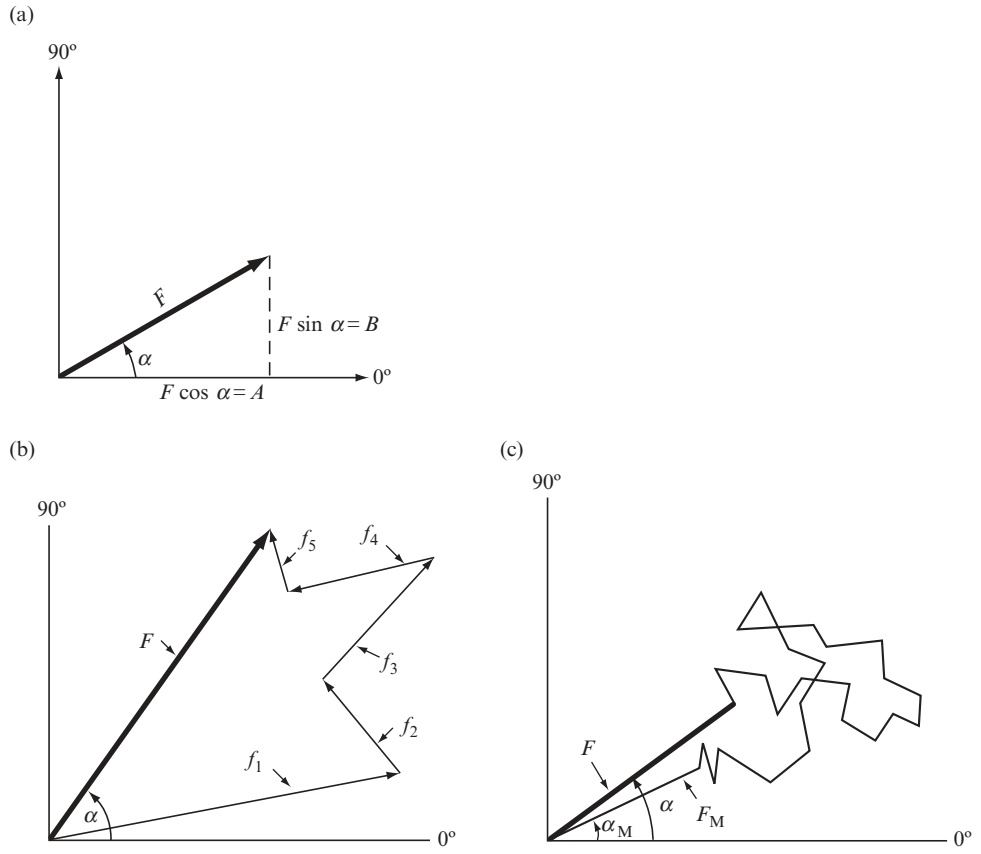


Fig. 5.2 Vector representation of structure factors.

(a) The relation of the vector **F** to A and B.
(b) The vector addition of the contribution of each atom to give a resultant **F**.
(c) If a "heavy" atom (M) has a much higher atomic number, and hence a much longer vector in a diagram (like that in Figure 5.2b) than any of the other atoms present, then the effect on the vector diagram for **F** is normally as if a short-stepped random walk had been made from the end of $\mathbf{F}_M$. Since the steps or f-values for the lighter atoms are relatively small, there is a reasonable probability that the angle between **F** and $\mathbf{F}_M$ will be small and an even higher probability that α (for the entire structure) will lie in the same quadrant as α_M (for the heavy atom alone). Thus the heavy-atom phase, α_M, may be used as a first approximation to the true phase, α.

wave to its sinusoidal appearance and phase angle is shown in Figure 5.1b. When there are several atoms in the unit cell, the various component scattering vectors can be added, as shown in Figures 5.2b and c.

Exponential representation (complex numbers)

For computational convenience, vector algebra is an improvement over graphical representation, but an even simpler notation is that involving so-called "complex" numbers, often represented as exponentials. The *exponential representation* is particularly simple because multiplication of

exponentials involves merely addition of the exponents. Equations (5.6) and (5.7) express the components of the resulting wave; Eqn. (5.10) expresses the amplitude of the resulting wave as the square root of the sum of the squares of its components, which we will now abbreviate as A and B. Equations (5.6), (5.7), and (5.10) may be rewritten as

$$A = c_r \cos \alpha_r = \sum_j c_j \cos \alpha_j \tag{5.11}$$

$$B = c_r \sin \alpha_r = \sum_j c_j \sin \alpha_j \tag{5.12}$$

and

$$c_r = (A^2 + B^2)^{1/2} \tag{5.13}$$

We will, as is conventional, let i represent $\sqrt{-1}$, an "imaginary" number. A complex number, C, is defined as the sum of a "real" number, x, and an "imaginary" number, iy (where y is real),

$$C = x + \mathrm{i}y \tag{5.14}$$

The magnitude of C, written as $|C|$, is defined as the square root of the product of C with its complex conjugate C^* (which is defined as $x - \mathrm{i}y$) so that

$$|C| \equiv [CC^*]^{1/2} = [(x + \mathrm{i}y)(x - \mathrm{i}y)]^{1/2} = [x^2 - \mathrm{i}^2 y^2]^{1/2} = [x^2 + y^2]^{1/2} \tag{5.15}$$

Comparison of Eqns. (5.14) and (5.15) with Eqns. (5.10)–(5.13) shows that the *vector representations of a wave and the complex number representations are parallel*, provided that we identify the vector itself as $A + \mathrm{i}B$. The result is that c_r of Eqn. (5.13) is identified with $|C|$ of Eqn. (5.15), and hence the vector components A and B are identified with x and y, respectively. A and B [as given by Eqns. (5.11) and (5.12)] represent components along two mutually orthogonal axes (called, with enormous semantic confusion, the "real" and "imaginary" axes, although both are perfectly real). The magnitude of the vector is given, as is usual, by the square root of the sum of the squares of its components along orthogonal axes, $(A^2 + B^2)^{1/2}$, as in Eqns. (5.13) and (5.15).

One advantage of the complex representation follows from the identity

$$\mathrm{e}^{\mathrm{i}\alpha} \equiv \cos \alpha + \mathrm{i} \sin \alpha \tag{5.16}$$

(which can easily be proved using the power-series expansions for these functions). We then have our expression for the total scattering as

$$A + \mathrm{i}B = c_r \cos \alpha_r + \mathrm{i}c_r \sin \alpha_r \equiv c_r \mathrm{e}^{\mathrm{i}\alpha_r} \tag{5.17}$$

Note that the amplitude of this scattered wave is c_r and its phase angle is α_r, as before, with $\alpha_r = \tan^{-1}(B/A)$, as in Eqn. (5.9).

Thus Eqn. (5.17) provides a mathematical means that is computer-usable for summing values of A and iB. It is often said, when this representation of the result of the superposition of scattered waves is used,

that A is the "real" component and B the "imaginary" component, a terminology that causes considerable uneasiness among those who prefer their science firmly founded and not flirting with the unreal or imaginary. It cannot be stressed too firmly that *the complex representation is merely a convenient way of representing two orthogonal vector components in one equation*, with a notation designed to keep algebraic manipulations of the components in different directions separate from one another. Each component is entirely real, as is evident from Figures 5.1 and 5.2.

Scattering by an individual atom

Electrons are the only components of the atom that scatter X rays significantly, and they are distributed over atomic volumes with dimensions comparable to the wavelengths of X rays used in structure analysis. The amplitude of scattering for an atom is known as the "atomic scattering factor" or "atomic form factor", and is symbolized as f. It is the scattering power of an atom measured relative to the scattering by a single electron under similar conditions. If the electron density is known for computed atomic orbitals (see Hartree, 1928; James, (1965); Stewart et al., 1965; Pople, 1999), then atomic scattering factors can be calculated from this electron density as shown in Figure 5.3. The electron densities of the atomic orbitals form the basis of the scattering

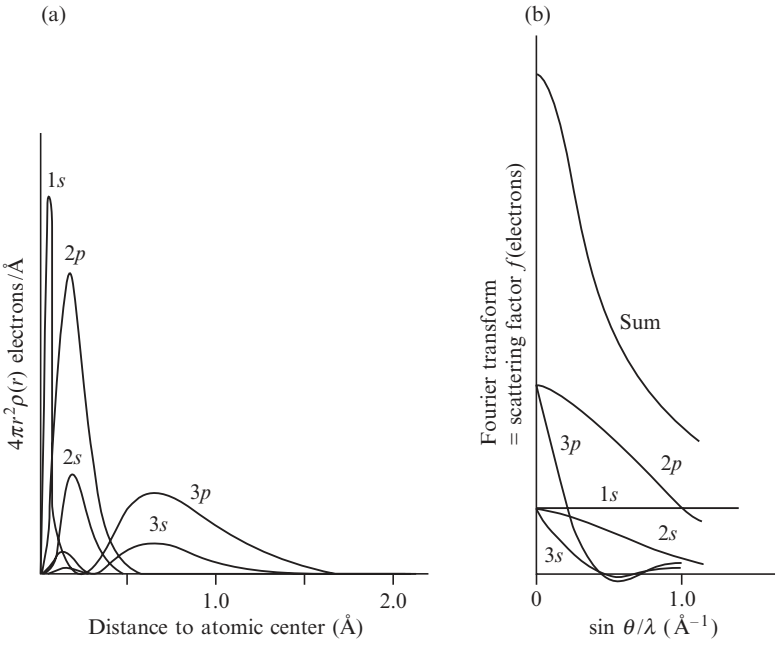


Fig. 5.3 Atomic scattering factors.

(a) Radial electron density distribution in atomic orbitals from theoretical calculations and (b) the scattering factors derived from them. The scattering curves in Figure 5.4 are similar to the uppermost curve (marked "Sum") in (b) here.

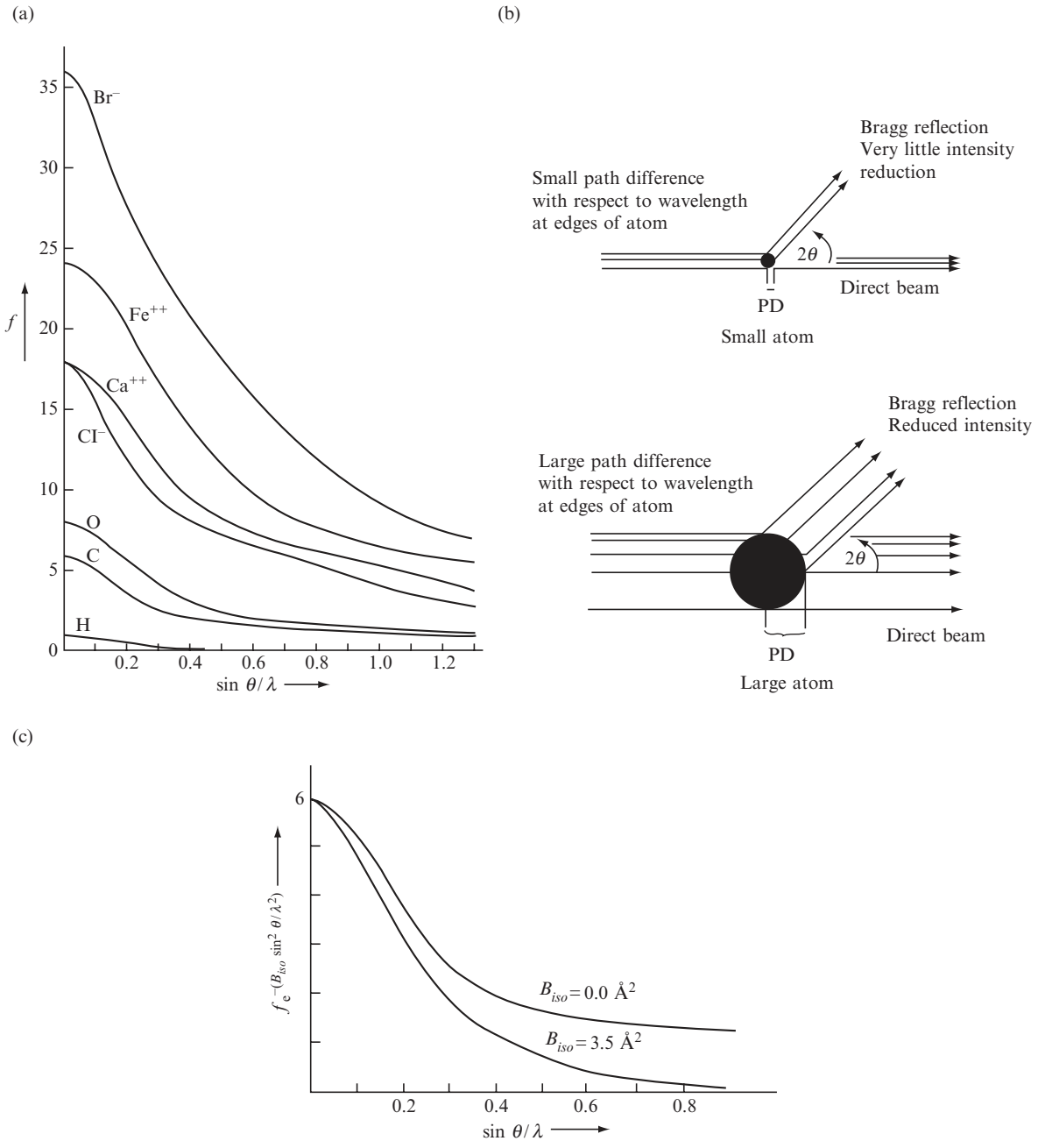


Fig. 5.4 Atomic-scattering-factor curves.

(a) Some atomic-scattering-factor curves for atoms, given as a function of $\sin\theta/\lambda$ so that they will be independent of wavelength. (Remember that 2θ is the deviation of the diffracted beam from the direct X-ray beam, wavelength λ.) The scattering factor for an atom is the ratio of the amplitude of the wave scattered by the atom to that of the wave scattered by a single electron. At $\sin\theta/\lambda = 0$ the value of the scattering factor of a neutral atom is equal to its atomic number, since all electrons then scatter in phase. Note that calcium (Ca^{++}) and chloride (Cl^{-}) are isoelectronic; that is, they have the same number of extranuclear electrons.

factors as a function of $\sin\theta/\lambda$; they are published and available in *International Tables*. Some values are given in Appendix 5.

For most purposes in structure analysis it is adequate to assume that atoms themselves are spherically symmetrical, but, with some of the best data now available, small departures from spherical symmetry (attributable to covalent bonding, lone pairs of electrons, and nonspherical orbitals, for example) are detectable. However, in our discussions, unless stated otherwise, we will assume spherical symmetry of atoms. This means that the scattering by an assemblage of atoms—that is, by the structure—can be very closely approximated by summing the contributions to each scattered wave from each atom independently, taking appropriate account of differences in the phase angles of each wave. Some atomic scattering factors, plotted as a function of $\sin\theta/\lambda$, are shown in Figure 5.4a. Since the diffraction pattern is the sum of the scattering from all unit cells, and this can be represented by the average contents of a single one of these unit cells, vibrations or disorder may be considered the equivalent of the smearing out of the electron density, so that there is a greater fall-off in the intensity of the diffraction pattern at a higher $\sin\theta/\lambda$ values (cf. the optical analogy in Figure 3.1: the wider the slit, the narrower the diffraction pattern). This modification of the fall-off by atomic vibration, motion or disorder, which results in a larger apparent atomic size as shown in Figure 5.4b, increases the falloff in scattering power as a function of scattering angle (Figure 5.4c). This fall-off may be isotropic (equal in all directions) or anisotropic (greater in certain directions in the unit cell than in others). Information obtained from an analysis of such atomic motion or disorder is discussed in Chapter 12. It leads, in nearly all crystal structures, to a model with anisotropic displacement parameters representing an inexact register of atomic positions from unit cell to unit cell. By contrast to X-ray scattering, neutrons are scattered by atomic nuclei, rather than by electrons around a nucleus, and hence, since the nucleus is so small (equivalent to a "point atom"), the neutron scattering for a nonvibrating nucleus is almost independent of scattering angle.

The positively charged calcium ion pulls electrons closer to the nucleus than does the chloride ion, which is negatively charged and has a lower atomic number. The resulting "narrower atom" for Ca^{++} will, for reasons shown in Figure 3.1, give a broader diffraction pattern. This is shown at high values of $\sin\theta/\lambda$ by higher values of f for Ca^{++} than for Cl^{-}.

(b) When radiation is scattered by particles that are very small relative to the wavelength of the radiation, such as neutrons, the scattered radiation has approximately the same intensity in all directions. When it is scattered by larger particles, the radiation scattered from different regions of the particle will still be in phase in the forward direction, but at higher scattering angles there is interference between radiation scattered from various parts of the particle. The intensity of radiation scattered at higher angles is thus less than for that scattered in the forward direction. This effect is greater the larger the size of the particle relative to the wavelength of the radiation used.

(c) The effects of isotropic vibration on the scattering by a carbon atom. Values are shown for a stationary carbon atom (B_{iso} of $0.0\,Å^2$) and for one with a room temperature isotropic displacement factor (B_{iso} of $3.5\,Å^2$) that corresponds to a root-mean-square amplitude of vibration of 0.21 Å. Vibration and disorder result in an apparently relatively greater size for the atoms (since we are considering an average of millions of unit cells), and consequently a decrease in scattering intensity with increasing scattering angle. If B_{iso} is large, no Bragg reflections may be detectable at high values of 2θ; that is, a narrower diffraction pattern is obtained from the "smeared-out" electron cloud of a vibrating atom (cf. Figure 3.1).

Scattering by a group of atoms (a structure)

The X radiation scattered by one unit cell of a structure in any direction in which there is a diffraction maximum has a particular combination of amplitude and relative phase, known as the *structure factor* and symbolized by F or $F(hkl)$ (Sommerfeld, 1921). It is the ratio of the amplitude of the radiation scattered in a particular direction by the contents of one unit cell to that scattered by a single electron at the origin of the unit cell under the same conditions. The intensity of the scattered radiation is proportional to the square of the amplitude, $|F(hkl)|^2$. In the manner just discussed [see Eqn. (5.17)], the structure factor can be represented either exponentially or as an ordinary complex number:

$$F(hkl) = |F(hkl)|e^{i\alpha(hkl)} = A(hkl) + iB(hkl) \tag{5.18}$$

with $|F|$ or $|F(hkl)|$ representing the amplitude of the scattered wave, and $\alpha(hkl)$ its phase relative to the chosen origin of the unit cell.* As before (Figure 5.1), $\alpha = \tan^{-1}(B/A)$ and $c_r = |F(hkl)| = (A^2 + B^2)^{1/2}$. The quantities A and B, representing the components of the wave in its vector representation (see Figure 5.2), can be calculated, if one knows the structure, merely by summing the corresponding components of the scattering from each atom separately. These components are [by Eqns. (5.6) and (5.7)] the products of the individual atomic-scattering-factor amplitudes, f_j, and the cosines and sines of the phase angles, α_j, of the waves scattered from the individual atoms:

$$A(hkl) = \sum_j f_j \cos \alpha_j \tag{5.19}$$

and

$$B(hkl) = \sum_j f_j \sin \alpha_j \tag{5.20}$$

But how do we calculate α_j for each atom?

If an atom lies at the origin of the unit cell and if other atoms lie one or several unit-cell translations (a) from it, then this grating of atoms will give a series of Bragg reflections $h00$ on diffraction. If there is another atom between two of them, at a distance xa from the origin (where x is less than 1), radiation scattered by this atom will interfere with the other resultant Bragg reflection by an amount that depends on the value of x. This can be generalized so that, for each $h00$ Bragg reflection, the phase difference (interference) will depend on the value of hx as illustrated in Figures 5.5 and 5.6. We show in Appendix 6 that the phase of the wave scattered in the direction of a reciprocal lattice point (hkl) by an atom situated at a position x, y, z in the unit cell (where x, y, and z are expressed as fractions of the unit-cell lengths a, b, and c, respectively) is just $2\pi(hx + ky + lz)$ radians, relative to the phase of the wave scattered in the same direction by an atom at the origin. This is important because it defines the effect of the location of an atom in the unit cell. The "relative phase angle" for an atom at x,y, z, where these numbers are defined

* The structure factor **F** may be represented as a vector, but it is not conventionally written in bold face, so we, as is common, will use F for the vector and $|F|$ for its amplitude.

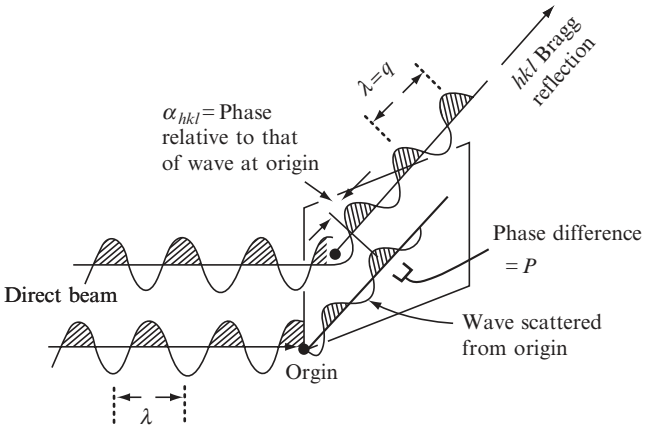


Fig. 5.5 The meaning of "relative phase."

The relative phase angle $\alpha(hkl)$ of a Bragg reflection *hkl* is the difference between the phase of a wave scattered by an atom (shown as a black circle within the unit cell) and the phase of a wave scattered in the same direction by an imaginary atom at the chosen origin of the unit cell.

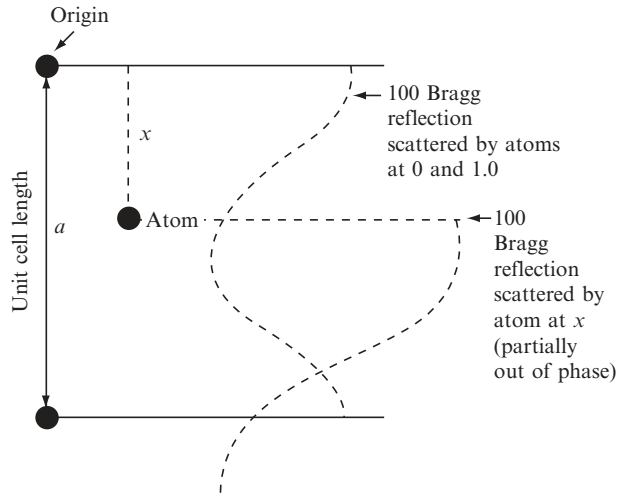


Fig. 5.6 The relative phase angle on diffraction.

If a one-dimensional structure with a repeat distance a has an atom at 0.0 (and this is repeated from unit cell to unit cell by other atoms at 1.0, 2.0, etc.) and an atom at x/a, the phase difference between the atom at 0.0 and the atom at x/a is $2\pi hx$ radians. Suppose that the atom at 0.0 is at the chosen origin of the system. Its phase angle for a cosine function is 0°. The phase angle of the atom at x is $2\pi hx$ radians. This is the difference of its phase with that of the atom at the origin, and hence the radiation scattered by the atom at x is considered to have a relative phase of $2\pi hx$ radians.

with respect to a chosen origin at 0, 0, 0, is $2\pi(hx + ky + lz)$ radians. If the location of the chosen origin is changed, the relative phase will also be changed. Each structure factor is the sum of the scattering from all atoms j in the unit cell. Thus Eqns. (5.19) and (5.20) (for all atoms j) can be rewritten as

$$A(hkl) = \sum_j f_j \cos 2\pi(hx_j + ky_j + lz_j) \tag{5.21}$$

$$B(hkl) = \sum_j f_j \sin 2\pi(hx_j + ky_j + lz_j) \tag{5.22}$$

where the value of f_j chosen is that corresponding to the value of $\sin\theta/\lambda$ for the Bragg reflection in question, modified to take into account any thermal vibration of the atom. A comparison with Eqns. (5.19) and (5.20) shows that we now know the phase α_j. The magnitude of $|F(hkl)|$ depends only on the relative positions of the atoms in the unit cell, except to the extent that f_j is a function of the scattering angle. The size and shape of the unit cell do not appear as such in the expressions for A and B. In Figure 5.2, F is represented as a vector. Note that a shift in the chosen origin of the unit cell will add a constant to the phase angle of each atom [see Eqns. (5.21) and (5.22)]; that is, it will rotate the phase diagrams in Figure 5.2 relative to the coordinate axes, but will leave the length of $|F(hkl)|$, and hence the values of $|F(hkl)|^2$ and the intensity, unchanged.

Effects of atomic vibration and displacements on atomic scattering

Atomic vibrations in a crystal, that is, displacements from equilibrium positions, have a frequency of the order of 10^{13} per second. This is much slower than the frequencies of X rays used to study crystals; these are of the order of 10^{18} per second. Therefore a vibrating atom will appear stationary to X rays but displaced in a random manner within the vibration amplitude. Atoms in other unit cells will also exhibit this random deviation from their equilibrium positions, different for each such atom in the various unit cells throughout the crystal. Because minor static displacements of atoms appear similar to displacements caused by atomic vibrations, it is usual to use the term "atomic displacement parameter" rather than "atomic temperature factor" for the correction factor. When $2\theta = 0$, all electrons in the atom scatter in phase, and the scattering power of an atom at this angle, expressed relative to the scattering power of a free electron, is just equal to the number of electrons present (the atomic number for neutral atoms).

However, an atom has size (relative to the wavelength of the X rays used), with the result that X rays scattered from one part of an atom interfere with those scattered from another part of the same atom at all angles of scattering greater than 0°. This causes the scattering to

fall off with increasing scattering angle or, more precisely, increasing values of $\sin\theta/\lambda$, as indicated in Figure 5.4b. The fall-off in intensity with higher scattering angle (Figure 5.4c) increases as the vibrations of atoms become greater, and these vibrations in turn increase in extent with rising temperature. Atoms are in motion in the crystalline state, however, even when the temperature is reduced to near absolute zero. This vibration, coupled with displacements of some atoms, leads to a significant reduction in intensity that can be approximated by an exponential function that has a large effect at high 2θ values (illustrated in Figure 5.4c); this indicates, as noted by Peter J. W. Debye and Ivar Waller, that atomic motion endows a larger "apparent size" to atoms (Debye, 1914: Waller, 1923). Effectively, since atoms are displaced different amounts from unit cell to unit cell at the given instant in time that measurement occurs, atoms appear to have become smeared in the average of all the unit cells in the crystal. If the displacement amplitude is sufficiently high, essentially no diffracted intensity will be observed beyond some limiting value of the scattering angle; that is, the "slit" is effectively widened by the vibration and so the "envelope" is narrow (Figure 3.5a). If the displacements are nearly isotropic—that is, do not differ greatly in different directions—the exponential factor can be written as $\exp(-2B_{\text{iso}}\sin^2\theta/\lambda^2)$, with B_{iso} called the *atomic displacement factor*.** B_{iso} is equal to $8\pi^2 <u^2>$, where $<u^2>$ is the mean square amplitude of displacement of the atom from its equilibrium position. The type of disorder found in a crystal may be static, with the atom in one site in one unit cell and a different site in another unit cell. Alternatively, it may be dynamic, which implies that the atom moves from one site to another. The overall effect in both cases is a reduction in the scattering factors of the atoms involved as $\sin\theta/\lambda$ increases (see Willis and Pryor, 1975).

** Many crystallographers omit the subscript "iso," relying on the context to avoid confusion with the quantity B defined in $F = A + iB$.

If the motion or disorder is anisotropic, it is necessary to replace B_{iso} by six terms. This is usually necessary for all atoms except hydrogen atoms; these have only weak scattering power. Atoms in crystals rarely have isotropic environments. The six parameters define the orientations of the principal axes of the ellipsoid that represents the anisotropic displacements and the magnitudes of the displacements along these axes. The results are often displayed in an ORTEP† diagram, in which the atomic displacement factors are drawn as ellipsoids (Johnson, 1965). If the anisotropy is severe, the ellipsoid representing the displacement probability and its direction may be abnormally extended in shape and may be better represented as disorder in two positions.

† ORTEP = Oak Ridge Thermal Ellipsoid Plot (Johnson, 1965).

Macromolecules, such as proteins, show interesting thermal and displacement effects. While their structures are generally, but not always, measured at a lower resolution than for small molecules, anisotropic displacement parameters are rarely determined, but isotropic displacements give information on the motion and flexibility of various portions in the molecule. One domain of the molecule may appear to move in a hingelike manner with respect to another part of the same molecule. Also, side chains at the surface of the macromolecule may have

alternate atomic positions from unit cell to unit cell as they interact with the various water molecules that fill nearly half of the crystal volume.

Calculating a structure factor

With a method for expressing a structure factor by means of an equation (Eqn. 5.18), and information on the components of this equation, it is possible to obtain a calculated value for the structure factor. This can be compared with the experimental value derived from $I(hkl)$. The data needed in order to calculate a structure factor include the values of x, y, and z for each atom; h, k, l, and $\sin\theta/\lambda$ for the Bragg reflection under consideration; and the scattering factor f_j for each atom at that value of $\sin\theta/\lambda$, modified by atomic displacement factors. Then it is necessary to calculate $2\pi(hx_j + ky_j + lz_j)$ and its sine and cosine for each atom and the Bragg reflection for which $F(hkl)$ is being calculated. This gives all the information necessary to sum the results for each atom and obtain $A(hkl)$ and $B(hkl)$ according to Eqns. (5.21) and (5.22). These lead to $F(hkl)$, that is, $(A^2 + B^2)^{1/2}$, and the relative phase angle $\alpha(hkl)$, that is, $\tan^{-1}(B/A)$, for the Bragg reflection with indices h, k, and l when all the atomic coordinates are known. This process has to be repeated for all of the other Bragg reflections. It demonstrates how important computers are to the X-ray crystallographer.

Information on the electron-density map will have to wait until we know the phase of the structure factor (so that we can determine the atomic positions x, y, and z). All we have so far are the experimentally measured structure amplitudes, $|F(hkl)|$, but we can calculate $F(hkl) = A(hkl) + \mathrm{i}B(hkl)$ (including its relative phase angle $\alpha = \tan^{-1}(B(hkl)/A(hkl))$, see Eqns. 5.21 and 5.22) if we have x, y, and z for a model in a unit cell of known dimensions and space group.

Summary

When X rays are diffracted by a crystal, the intensity of X-ray scattering at any angle is the result of the combination of the waves scattered from different atoms and the manner in which they modify this intensity by various degrees of constructive and destructive interference. A structure determination involves a matching of the observed intensity pattern to that calculated from a postulated model, and it is thus imperative to understand how this intensity pattern can be calculated for any desired model. The combination of the scattered waves can be represented in various ways:

(1) The waves may be drawn *graphically* and the displacements (ordinates, vertical axis) at a given position (abscissae, horizontal axis) summed.

(2) A wave may be represented *algebraically* as

$$x_j = c_j \cos(\phi + \alpha_j) \quad \text{(for the } j\text{th wave)}$$

and the displacements, x_j, of several such waves summed to give a resultant wave.

(3) The waves may be expressed as *two-dimensional vectors* in an orthogonal coordinate system, amplitude c_j, with the relative phase angle α_j measured in a counterclockwise direction from the horizontal axis. This is the equivalent of representing one complete wavelength as 360°, so that the periodicity of the wave is expressed. The phase relative to some origin is given as a fraction of a revolution. The vectors may then be summed by vectorial addition of their components.

(4) The waves may be represented in *complex notation*

$$A_j + \mathrm{i}B_j = c_j \mathrm{e}^{\mathrm{i}\alpha_j}$$

which is merely a convenient way of representing two orthogonal vector components (at 0° and 90°) in one equation. By convention A is the component at 0° and B the component at 90°.

X rays are scattered by electrons. The extent of scattering depends on the atomic number of the atom and the angle of scattering, 2θ, and is represented by an atomic scattering factor f. For a group of atoms, the amplitude (relative to the scattering by a single electron) and the relative phase of the X rays scattered by one unit cell are represented by the structure factor $F(hkl) = A(hkl) + \mathrm{i}B(hkl)$ for each Bragg reflection. For a known structure with atoms j at positions x_j, y_j, z_j, this may be calculated from

$$A(hkl) = \sum_j f_j \cos 2\pi(hx_j + ky_j + lz_j)$$

and

$$B(hkl) = \sum_j f_j \sin 2\pi(hx_j + ky_j + lz_j)$$

where the summation is over all atoms in the unit cell. The relative phase angle $\alpha(hkl)$ is $\tan^{-1}(B/A)$ and the structure factor amplitude $|F(hkl)|$ is $\{(A(hkl)^2 + B(hkl)^2\}^{1/2}$. The value of $F(hkl)$ may be reduced as a result of thermal vibration and atomic displacement so that if F_{novib} is the value for a structure containing stationary atoms, the experimental values will correspond to $F(hkl) = F_{\text{novib}} \exp(-B_{\text{iso}} \sin^2\theta/\lambda^2)$, where B_{iso}, the atomic displacement parameter, is a measure of the amount of vibration and/or displacement ($B_{\text{iso}} = 8\pi^2 < u^2 >$, where $< u^2 >$ is the mean square amplitude of displacement). With precise experimental data, it is possible to measure the anisotropy of vibration and displacement.

6 The phase problem and electron-density maps

In order to obtain an image of the material that has scattered X rays and given a diffraction pattern, which is the aim of these studies, one must perform a three-dimensional Fourier summation. The theorem of Jean Baptiste Joseph Fourier, a French mathematician and physicist, states that a continuous, periodic function can be represented by the summation of cosine and sine terms (Fourier, 1822). Such a set of terms, described as a Fourier series, can be used in diffraction analysis because the electron density in a crystal is a periodic distribution of scattering matter formed by the regular packing of approximately identical unit cells. The Fourier series that is used provides an equation that describes the electron density in the crystal under study. Each atom contains electrons; the higher its atomic number the greater the number of electrons in its nucleus, and therefore the higher its peak in an electron-density map. We showed in Chapter 5 how a structure factor amplitude, $|F(hkl)|$, the measurable quantity in the X-ray diffraction pattern, can be determined if the arrangement of atoms in the crystal structure is known (Sommerfeld, 1921). Now we will show how we can calculate the electron density in a crystal structure if data on the structure factors, including their relative phase angles, are available.

Calculating an electron-density map

The Fourier series is described as a "synthesis" when it involves structure amplitudes and relative phases and builds up a picture of the electron density in the crystal. By contrast, a "Fourier analysis" leads to the components that make up this series. The term "relative" is used here because the phase of a Bragg reflection is described relative to that of an imaginary wave diffracted in the same direction at a chosen origin of the unit cell (see Figure 6.1). The number of electrons per unit volume, that is, the electron density at any point x, y, z, represented by $\rho(xyz)$, is given by the following expression (for an electron-density map,

a Fourier synthesis):

$$\rho(xyz) = \frac{1}{V_c} \sum_{\text{all } hkl} \sum \sum F(hkl) \exp[-2\pi i(hx + ky + lz)] \qquad (6.1)$$

Here V_c is the volume of the unit cell, and $F(hkl)$ is the structure factor for the Bragg reflection with indices h, k, and l. The triple summation is over all values of the indices h, k, and l. This summation, first calculated in 1925, represents a mathematical analogy to the process effected physically in the microscope (Duane, 1925; Havighurst, 1925; Waser, 1968). As described in Chapter 4, the amplitude of $F(hkl)$, that is, $|F(hkl)|$, is easily derived [Eqn. (4.3)] from the intensity of the Bragg reflection. The phase of that same Bragg reflection $\alpha(hkl)$, however, is not.

We will simplify the following equations by putting

$$\phi = 2\pi(hx + ky + lz) \qquad (6.2)$$

We then abbreviate $A(hkl)$ and $B(hkl)$ to A and B, respectively,* and $F(hkl) = |F(hkl)|e^{i\phi}$ to $F = A + iB$. This leads to Eqn. (6.3) (from Eqns. (5.16) to (5.18) for $|F(hkl)| = F(hkl)e^{-i\phi}$. In this equation, $e^{-i\phi} = \cos\phi - i\sin\phi$ and $i^2 = -1$:

$$Fe^{-i\phi} = (A + iB)(\cos\phi - i\ \sin\phi) = A\cos\phi + B\sin\phi - i(A\sin\phi - B\cos\phi) \qquad (6.3)$$

Because the summation in Eqn. (6.1) is over all values of the indices h, k, and l, it includes, in addition to every Bragg reflection hkl, the corresponding one with all indices having the opposite signs, $-h$, $-k$, $-l$ (also denoted $\bar{h}$, $\bar{k}$, $\bar{l}$). The *magnitude* of each term ($A(hkl)$, $B(hkl)$, $\cos\phi$, and $\sin\phi$) is normally the same** for a Bragg reflection with indices hkl as for that with indices $-h$, $-k$, $-l$. The *sign* of the term will change for these pairs of Bragg reflections if the term involves sine functions [since $\sin(-x) = -\sin x$], but will remain unchanged if it involves cosine functions [since $\cos(-x) = \cos x$]. Both $A(hkl)$ [the sum of cosines, by Eqn. (5.19)] and $\cos\phi$ have the same sign for hkl as for $-h$, $-k$, $-l$, whereas $B(hkl)$ [the sum of sines, by Eqn. (5.20)] and $\sin\phi$ have opposite signs for this pair of Bragg reflections. Therefore, when Eqn. (6.3) is substituted in Eqn. (6.1) and the summation is made, the $i(A\sin\phi - B\cos\phi)$ terms cancel for each pair of Bragg reflections hkl and $\overline{hkl}$ and vanish completely. The remaining terms, $A\cos\phi$ and $B\sin\phi$, need be summed over only half of the Bragg reflections. All those with any one index (for example, h) negative are omitted and a factor of 2 is introduced to account for this. Therefore we may write, by Eqns. (6.1) and (6.3),

$$\rho(xyz) = \frac{1}{V_c}\left\{|F(000)| + 2\sum_{\substack{h\geq 0,\ \text{all } k,l \\ \text{excluding } F(000)}}^{\infty}\sum\sum(A\cos\phi + B\sin\phi)\right\} \qquad (6.4)$$

* Note that the exponential terms in the expressions for F (the structure factor) and ρ (the electron density) are opposite in sign; $F = \Sigma f e^{i\phi}$ and $\rho = (1/V)\Sigma F e^{-i\phi}$. This is because these are Fourier transforms of each other (Glasser, 1987a,b; Carslaw, 1930).

** This implies that "Friedel's Law" $|F(hkl)|^2 = |F(\bar{h}\bar{k}\bar{l})|^2$ is obeyed (Friedel, 1913); deviations from this law are considered in Chapter 10.

Since $A = |F| \cos \alpha$ and $B = |F| \sin \alpha$ [by Eqn. (5.17), where α is the relative phase angle of $F(hkl)$, and $\cos X \cos Y + \sin X \sin Y = \cos(X - Y)$, the above expression for the electron density (Eqn. (6.4)) may be rewritten†

$$\rho(xyz) = \frac{|F(000)|}{V_c} + \frac{2}{V_c} \sum_{\substack{h \geq 0, \text{ all } k,l \\ \text{excluding } F(000)}} \sum_{\infty} \sum_{\infty} \sum_{\infty} (|F| \cos(\phi - \alpha)) \tag{6.5}$$

† A schematic example of the calculation of the function described in Eqn. (6.5) is shown in Figure 6.2.

This may be alternatively expressed as

$$\rho(xyz) = \frac{|F(000)|}{V_c} + \frac{2}{V_c} \sum_{\substack{h \geq 0, \text{ all } k,l \\ \text{excluding } F(000)}} \sum_{\infty} \sum_{\infty} \sum_{\infty} |F(hkl)| \cos[2\pi(hx + ky + lz) - \alpha(hkl)]$$

Remembering that $\phi = 2\pi(hx + ky + lz)$, an inspection of Eqn. (6.5) shows that we need both the magnitudes $|F(hkl)|$ and the relative phases $\alpha(hkl)$ of the radiation that has been diffracted in different directions. These are necessary for us to be able to form an image of the scattering matter, $\rho(xyz)$. *If we knew $|F(hkl)|$ and $\alpha(hkl)$, we could then calculate the Fourier summation in Eqn. (6.5) and plot the values of $\rho(xyz)$, thereby obtaining a three-dimensional electron-density map*. By assuming that atoms lie at the centers of peaks in this map, we would then know the atomic structure of the crystal.

However, as we have already stressed many times, we can normally obtain only the structure factor amplitudes $|F(hkl)|$ and not the relative phase angles $\alpha(hkl)$‡ directly from the experimental measurements. We must *derive* $\alpha(hkl)$, either from values of $A(hkl)$ and $B(hkl)$ that are computed from structures we have deduced in various ways ("trial structures"), or by purely analytical methods. The problem of getting estimates of the phase angles so that an image of the scattering matter can be calculated is called the *phase problem* and is the central one in X-ray crystallography. Chapters 8 and 9 are devoted to methods used to solve the phase problem, either by deriving a trial structure and then calculating approximate values of $\alpha(hkl)$ for each Bragg reflection, or by trying to find values of $\alpha(hkl)$ directly. Recall that, for the third-order Bragg reflection, the path difference between waves scattered one repeat unit (a) apart (that is, by equivalent atoms in adjacent unit cells) is three wavelengths. The important fact for the reader to understand is that each resultant wave should be traced back and its phase compared with that of an imaginary wave being scattered at the origin of the repeat unit (with a relative phase angle of 0°); that is why we call it a "relative phase," the origin being in a position chosen by the investigator (see Nyburg, 1961).

‡ Under certain conditions, when two-beam diffraction occurs, some phase information may be derived from experimental measurements (see Chapter 10).

How do we derive the relative phases of the density waves, that is, their phases relative to a chosen origin? We attempt to show, in Figure 6.1, how the X rays scattered from different atoms are summed to give the resultant X ray beams of various amplitudes (and hence inten-

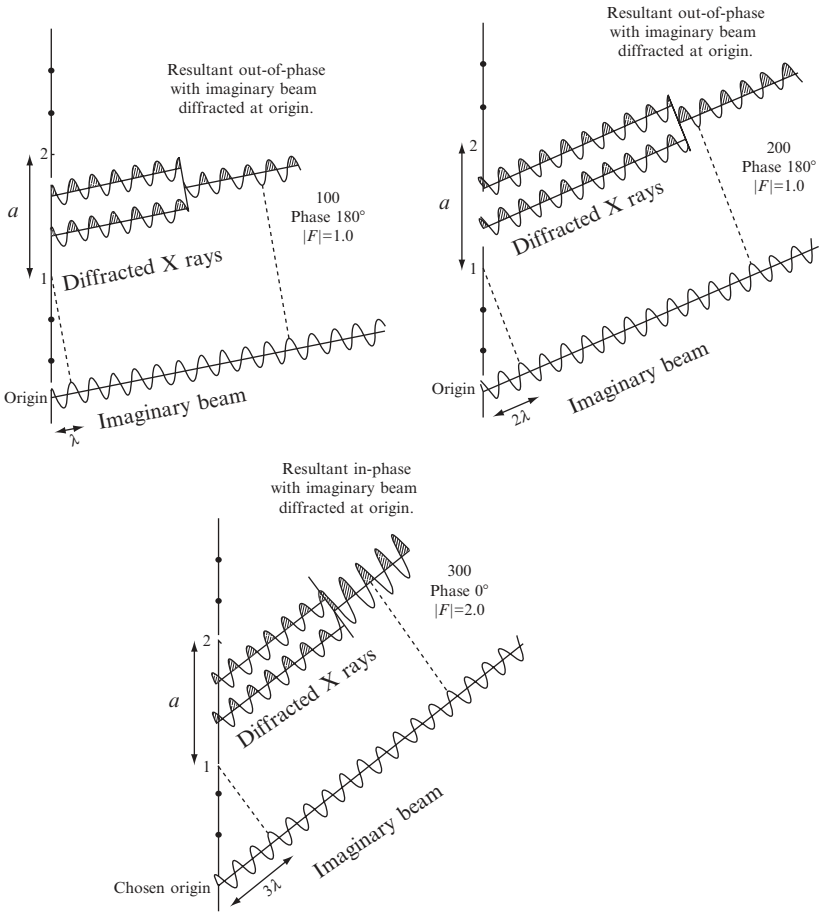


Fig. 6.1 Scattered waves and their relative phases.

A one-dimensional crystal with two atoms in the unit cell, one at $x = 1/3$ and the other at $x = 2/3$. Shown are the Bragg reflections (a) 100, (b) 200, and (c) 300 and their relationships to an imaginary wave scattered at the chosen origin of the unit cell (which leads to the "relative phase angle"). Note that the most intense of these three is the 300 Bragg reflection.

sities). A unit cell containing two atoms, one at $x = 1/3$ and the other at $x = 2/3$, is used to illustrate how relative phases are derived. Compared with an imaginary atom at the origin, the atom at $x = 1/3$ scatters for a third order reflection with a path difference of one wavelength and the atom at $x = 2/3$ scatters with a path difference of two wavelengths. Thus both scatter in phase with the wave scattered at $x = 0$. However, for the second order, the atom at $x = 1/3$ scatters X rays with a path difference of 0.67 wavelengths from that scattered by the imaginary atom at the origin, and the atom at $x = 2/3$ scatters with a path difference of 1.33 wavelengths from the wave scattered at the origin. The resultant wave is then $(0.33 + 0.67)/2 = 0.50$ wavelengths out of phase with the wave scattered by the imaginary atom at the origin. Thus, in summing

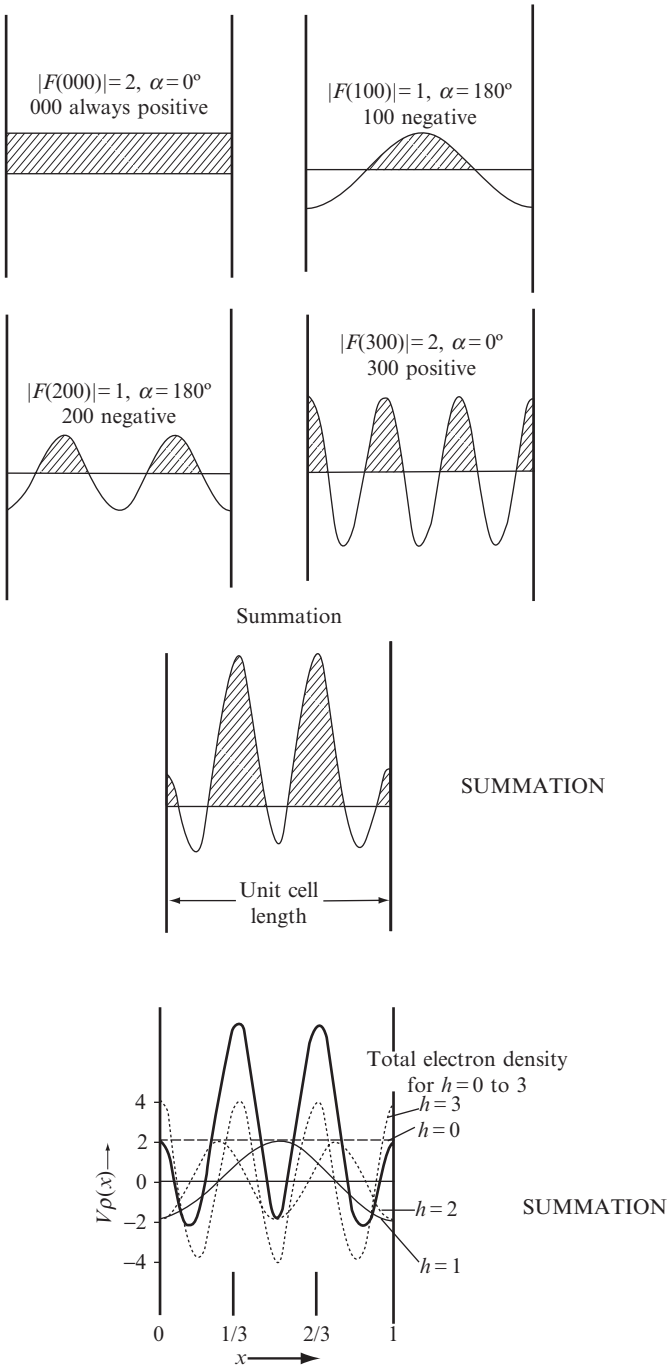


Fig. 6.2 Fourier synthesis of the Bragg reflections from Figure 6.1.

The Fourier summation of density waves to give an electron-density map with peaks at $x = \pm 1/3$. At any point x, y, z in the unit cell, volume V_c, the electron density $\rho(xyz)$ may be calculated by use of Eqn. (6.5). The following data have been used for this one-dimensional example:

density waves, as shown in Figure 6.2, the 300 wave has a relative phase angle of 0° and the 200 density wave has a relative phase angle of 180°.

Fourier transforms

We have shown that the electron density $\rho(xyz)$ (Eqn. 6.1) can be expressed in an equation that involves the structure factors $F(hkl)$ as coefficients,

$$\rho = (1/V)\Sigma\ F\,\mathrm{e}^{-\mathrm{i}\phi} \tag{6.6}$$

It is also possible to express the structure factors in terms of the electron density:

$$F\ =\ \Sigma\ f\mathrm{e}^{\mathrm{i}\phi} \tag{6.7}$$

The relationship between these two is referred to as a Fourier transform or Fourier inversion. These equations show that the structure factor is the Fourier transform of the scattering density (electrons in the molecule) sampled at the reciprocal lattice point *hkl*, while the electron density is the Fourier transform of the structure factors (which contain their relative phases). The intensity at a particular point of the diffraction pattern of an object (a set of relative $|F(hkl)|^2$ values) is proportional to the square of the Fourier transform of the object (with the distribution of matter in the object described by $\rho(x, y, z)$). Examples of Fourier transforms are shown in Figure 6.3, with electron density and density waves on the left and structure factors, with their relative phases, shown on the right as positive or negative ($\alpha = 0°$ or 180°). Equation (6.6) or (6.7) (whichever

h	**−3**	**−2**	**−1**	**0**	**1**	**2**	**3**
$\lvert F(hkl)\rvert$	2	1	1	2	1	1	2
$\alpha(hkl)(°)$	0	180	180	0	180	180	0
$\cos[2\pi(hx-\alpha)]$	$+\cos 6\pi x$	$-\cos 4\pi x$	$-\cos 2\pi x$	+1	$-\cos 2\pi x$	$-\cos 4\pi x$	$+\cos 6\pi x$

Therefore $\rho(x) = |F(000)|/V_c + (4\cos 6\pi x - 2\cos 4\pi x - 2\cos 2\pi x + 2)/V_c$.

When $h = 0$, the function does not depend on x and so is a straight line (but drawn with half its amplitude to conform to the electron-density map equation with positive and negative values of h). The phase angle of this is necessarily 0°. The function for $h = 1$ is $-\cos 2\pi x$, the negative sign resulting from the relative phase angle of 180°, and so forth. These functions are summed for each value of x to give the result shown by the heavy solid line. It has peaks at $x = \pm 1/3$. Clearly, unless the phases were known, it would not be possible to sum the waves correctly. This kind of calculation must be made, with thousands of Bragg reflections, at each of many thousands of points to give a complete electron-density map in three dimensions. Therefore high-speed computers are essential. For a three-dimensional electron-density map it is not possible to plot heights of peaks (because we have no fourth spatial dimension), and therefore contours of equal electron density (or height) are drawn on sections through the three-dimensional map. Atomic centers appear at the centers of areas of high electron density, which look like circular mountains on a topographical map. The larger values of F dominate the Fourier summation.

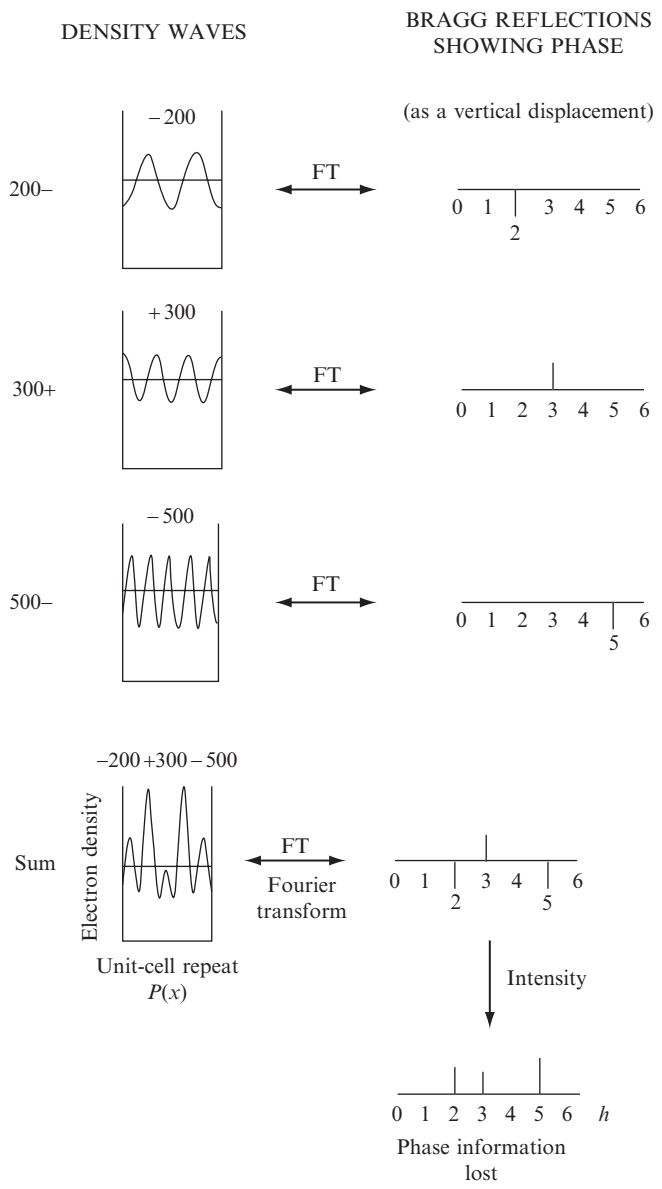


Fig. 6.3 Summing Fourier transforms.

Density waves for the 200, 300, and 500 Bragg reflections and their Fourier transforms. When the columns are summed, the density waves (on the left) give the electron density map, while their Fourier transforms (on the right) give the phases of the individual density waves. When intensities are measured, the phase information is lost. Note: This is a different structure from that in Figures 6.1 and 6.2.

is most appropriate) is used for the transformations. As will be seen later, it is convenient to be able to move readily between real (electron density) and reciprocal (structure factor) space, and this is how it is done. For example, one may want to modify an electron-density map

and calculate a new data set of structure factors for comparison with experimental values or may calculate the theoretical electron density for an atom or ion and then examine the atomic scattering factors relevant to this density (as mentioned in Chapter 5). Alternatively, one may want to change some (or all) of the structure factors and investigate the effect of this on the electron-density map. The Fourier transform equations make this possible.

Summing density waves to obtain an electron-density map

We have emphasized the analogy between the action of a lens in collecting and refocusing radiation to give an image of the scattering matter, and the process of Fourier summation, a mathematical technique for forming an image by use of information about the amplitudes and relative phases of the scattered waves. Fourier summation techniques can be applied even when the waves cannot be refocused, as in the X-ray experiment. With a lens the light waves are (ideally) brought together with the same phases that they had when they left the object; in the X-ray diffraction experiment these phases are usually not measurable, although if they can be found in some way, then it is possible to calculate an electron-density map as shown in Figures 6.3 and 6.4.

The individual waves in Eqn. (6.5) that are summed to give the electron-density map are referred to, for convenience in this book, as "density waves" (see Bijvoet et al., 1948). In other words, each term $|F(hkl)|\cos[(2\pi(hx+ky+lz)-\alpha(hkl)]$, calculated as a function of x, y, and z, is a density wave, as illustrated in Figure 6.4. In effect, Eqn. (6.5) could be rewritten to say that *the electron density* $\rho(x, y, z)$ *at a point in space* x, y, z *is equal to the sum of these density waves*. Thus each Bragg reflection with its relative phase can be considered to produce a density wave in the crystal, with an amplitude that can be derived from the intensity of the Bragg reflection; the superposition of these density waves, once their phases are known, produces the electron-density map for the crystal:

$$\rho(xyz) = \frac{1}{V_c}\left\{ \left|F(000)\right| + 2\sum_{\text{all density waves}}\sum^{\infty}\sum\left|F(hkl)\right|\cos[2\pi(hx+ky+lz)-\alpha(hkl)]\right\} \qquad (6.8)$$

The determination of the phases of these density waves is the subject of much of the rest of this book. But what is the wavelength of a density wave and how is it related to the order (h, k, l) of the diffracted beam? Their wavelengths depend on h, k, and l, not the wavelength of the X rays that caused each Bragg reflection. A close examination of Eqn. (6.5) shows that $|F(hkl)|$ is modified by a cosine function

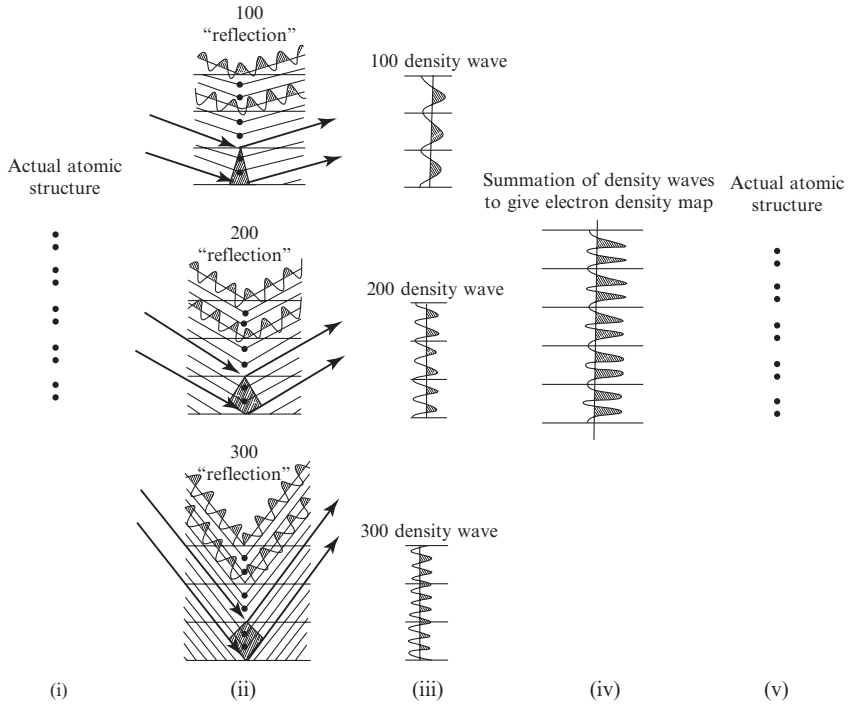


Fig. 6.4 Overview of X-ray diffraction.

Summary of the diffraction experiment, showing (i) the atomic structure (one-dimensional in this case); (ii) diffraction of X rays by the crystal structure; (iii) density waves; (iv) summation of the density waves to give the electron-density map; and the result is an image of the actual structure (v), which is the same as (i).

of $(2\pi(hx + ky + lz) - \alpha(hkl))$; thus it becomes a periodic function of h, k, and l. In the simple case (Figure 6.2) where k and l are both zero, $\cos 2\pi(hx)$ is at a maximum value when $x = 1/h$; that is, this cosine term has an apparent wavelength of a/h (where a is the unit-cell length in the x direction and x is expressed as a fraction of this dimension a).

In summary, the wavelengths of the density waves are $d_{hkl} = \lambda/2 \sin\theta$, their amplitudes are $|F(hkl)|$, and their phases are $\alpha(hkl)$. For example, the wavelength of the 1 0 0 density wave is the repeat distance $a\,(= d_{100})$ (see Figure 6.2), the wavelength of the 2 0 0 density wave is $a/2$ because the second order of diffraction occurs at a $\sin\theta$ value twice that of the first order, and so forth. For the 1 0 0 reflection, phase π, Eqn. 6.8 gives the function $\cos[2\pi x + \pi]$ which is maximal at $x = 1/2$ (see Figure 6.4). These are the density waves that are summed to give the electron-density map shown in Figure 6.2. "High resolution" implies a high value of $\sin\theta$ and thus a small value for the effective wavelength of the density wave; as we shall see later, high-resolution Bragg reflections (short wavelength density waves) are needed to provide high-resolution images of molecules.

The density waves, derived by arguments such as these, are summed as shown in Figures 6.2 and 6.4 to give the electron density of the

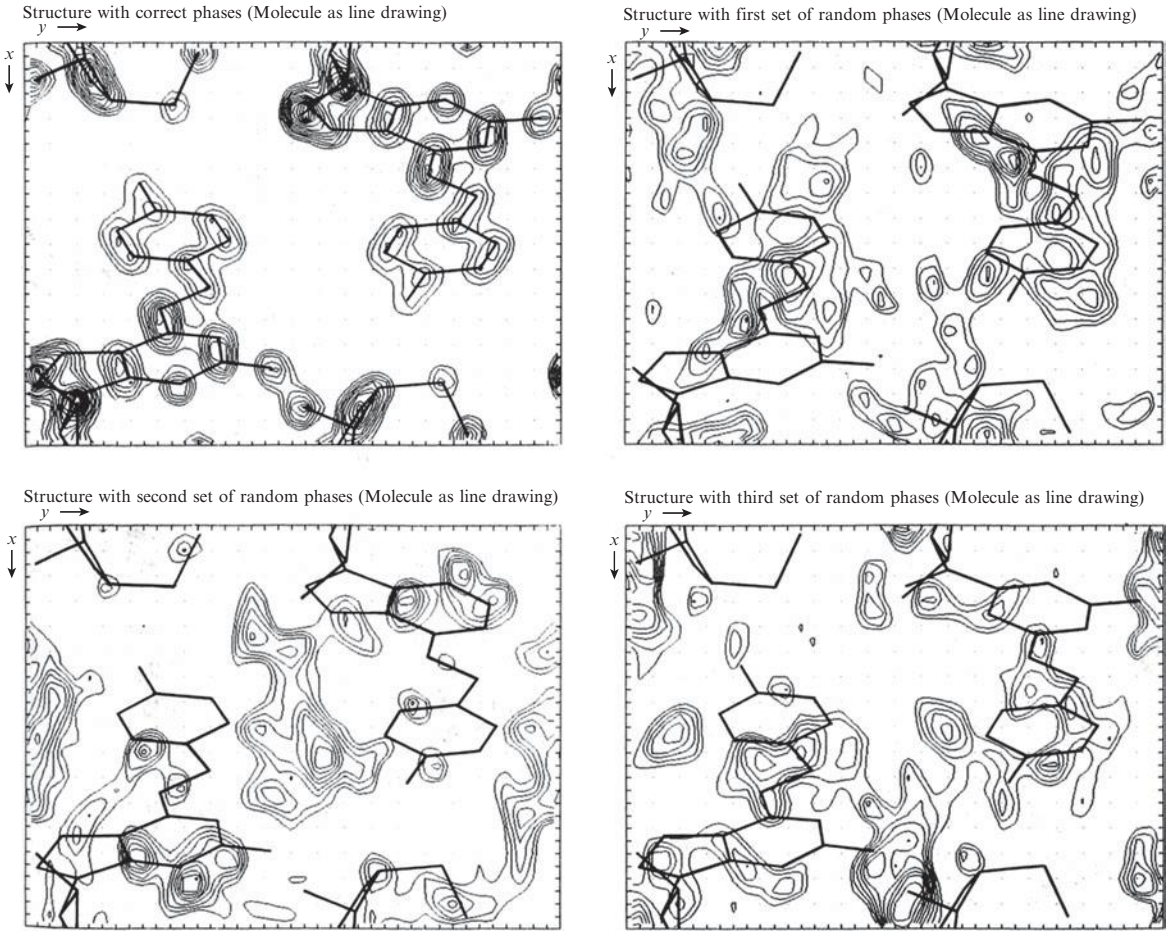


Fig. 6.5 Comparison of electron-density maps when the phases are correct and when they are incorrect and random.

In the computation of all maps shown here, the same $|F(hkl)|$ values but different phases were used. The upper left electron-density map (a) is the correct result; the other three maps ((b) to (d)) have incorrect relative phases, and provide an incorrect electron-density map. The phases of these three "random phase" maps were found by a computer program for random number generation. Since the structure is noncentrosymmetric, the phase for each Bragg reflection could have any value between 0° and 360°. In each case, the molecular skeleton is shown by solid lines in the correct position, but it is clear that only the first map (top left) correctly represents the true structure.

Courtesy H. L. Carrell.

structure, and the peaks in such a map correspond to the centers of atoms. The importance of the phases in determining a structure is illustrated in Figure 6.5. Each of the four electron-density maps in this figure has the same values of $|F|$, but differs in the phases used in the calculation. For clarity, the true crystal structure is indicated by a line diagram. As can be seen, only the first map correctly gives peaks at atomic positions. An electron-density map with correct phases much more nearly approximates the correct structure than does an electron-density map with incorrect phases, even if each has the correct magnitudes for the $|F(hkl)|$ values. The analysis of electron-density and Patterson maps has

benefited greatly from the improvements in computer graphics so that now it is possible to view the three-dimensional map on a computer screen and rotate and move it at will in order to obtain structural information. However, automatic fitting of a three-dimensional model structure to the electron-density map is now possible by computer without any need for display (Lamzin et al. 2001; Oldfield, 2003).

An initial trial structure

At the start of a structure determination one does not know the positions of all the atoms in the structure (for if one did, the structure would probably not need to be investigated), but one can often deduce an approximation to the correct structure. The calculated phases for this initial (approximate) trial structure will provide a starting point for structure determination. This trial structure may be one that completely fills the unit cell or else it may be only a partial structure (even, for example, one heavy atom). It is possible to calculate an approximation to the true electron density by a three-dimensional Fourier summation of the observed structure factor amplitudes, $|F_o|$, with phases calculated from an initial trial structure which may be only partially complete. It has been found that the general features of an electron-density map depend much more on the phase angles than on the structure factor amplitudes. Therefore a map calculated with only approximately correct phases will be an imperfect representation of the structure. However, it is biased toward the correct structure because the observed structure amplitudes $|F_o|$ were used in the calculation. By comparison with a similar synthesis using the calculated amplitudes $|F_c|$, or even more simply by computing the difference $(|F_o| - |F_c|)$ to obtain a "difference synthesis", one can deduce the changes in the model needed to give better agreement with observation. The positions of some hitherto unrecognized atoms may be indicated, and shifts in the positions of some atoms already included will normally be suggested as well.

Correctness of the trial structure

Once the approximate positions and identities of all the atoms in the asymmetric unit are known (that is, when the true crystal structure is known), the amplitudes and phases of the structure factors can readily be calculated (see Chapter 5). These calculated amplitudes, $|F(hkl)_c|$, may be compared with the observed amplitudes, $|F(hkl)_o|$. If the structural model is a correct one and the experimentally observed data are reasonably precise, the agreement should be good. The situation is different for phases. The phases calculated for a trial structure cannot be compared with observed phases, because normally *phases are not*

observed; they depend on where the origin of the unit cell was chosen to be.

One measure of the correctness of a structure is the so-called discrepancy index (or reliability index or conventional residual), R, defined as

$$R = \frac{\sum \left| (|F_o| - |F_c|) \right|}{\sum (|F_o|)} \tag{6.9}$$

It is a measure of how closely the experimentally observed structure factor amplitudes are matched by the values calculated for a proposed trial structure. At present, R values in the range of 0.02 to 0.06 (alternatively described as 2 percent to 6 percent) are being quoted for the most reliably determined structures of small molecules. An R value of 0.83 corresponds to a random centrosymmetric structure; that is, with proper scaling a randomly incorrect structure with a center of symmetry would give an R value of about 0.83 (0.59 for a noncentrosymmetric crystal structure) (Wilson, 1950). A refinable trial structure may have an R value between 0.25 and 0.35, or even somewhat higher. This value will (hopefully) be decreased by methods described in Chapter 11 to a much lower value. If one atom of high atomic number is present, the initial trial value of R may be much lower because the position of this atom can usually be determined reasonably well even at an early stage, and a heavy atom normally dominates the scattering, as illustrated in the atomic scattering factors in Figure 5.4a. If the trial structure is a reasonable approximation to the correct structure, the R value goes down appreciably as refinement proceeds.

The discrepancy index R is, however, only one measure of the precision (but not necessarily the accuracy) of the derived structure. It denotes how well the calculated model fits the observed data. Many complications can cause errors in the observed or calculated structure factors or both—for example, absorption of the X-radiation by the crystal, or atomic scattering factors and temperature factors that do not adequately describe the experimental situation. The fit of the calculated structure factors to the observed ones may then be good, but if the observations are systematically in error, the *accuracy* of the derived structure may be low, despite an apparently high *precision*. Hence care must be taken in interpreting R values. In general, the lower the R value the better the structure determination, but if one or more very heavy atoms are present, they may dominate the structure factor calculation to such an extent that the contributions from light atoms may not have noticeable effects on R, especially if the structure has not been refined extensively. The positions of the light atoms may then be significantly in error. Also the resolution of the data (i.e., the maximum value of $\sin\theta/\lambda$) must be taken into account in assessing the meaning of an R value. A few grossly incorrect trial structures have been refined to R values as low as 0.10. Fortunately this situation is not common.

Resolution of a crystal structure

The variation of resolving power with scattering angle in structural diffraction studies has a direct analogy with the resolution of an ordinary microscope image (Abbé, 1873; Porter, 1906). If some of the radiation scattered by an object under examination with a microscope escapes

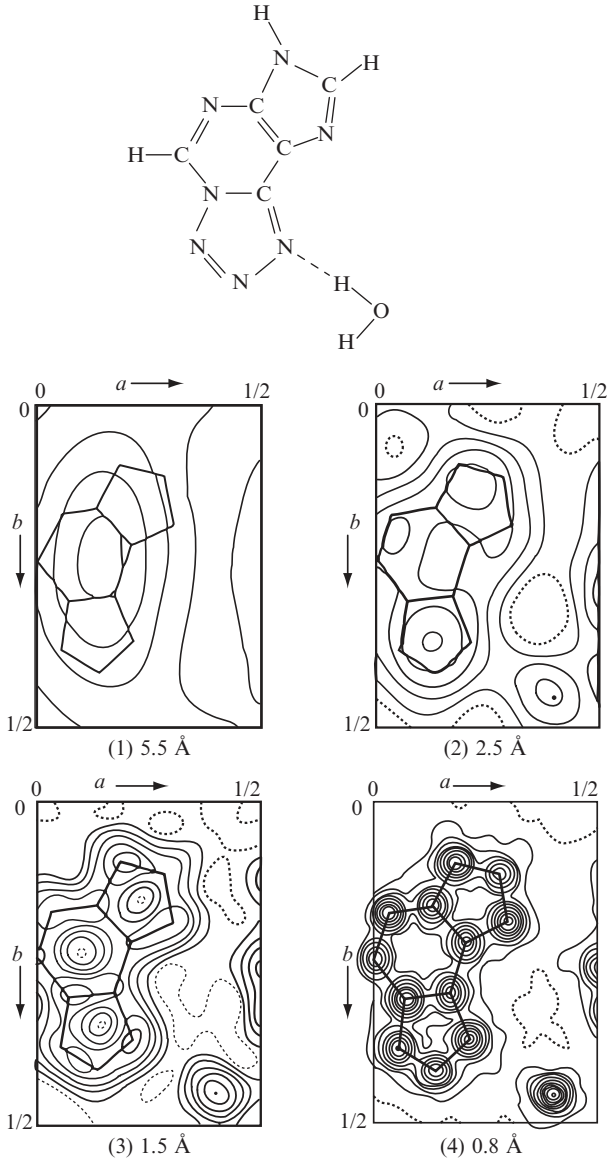


Fig. 6.6 Different stages of resolution for a given crystal structure.

The electron-density maps shown were calculated after eliminating all observed $|F(hkl)|$ measured beyond a given 2θ value. The "resolution" obtained is usually expressed in terms of the interplanar spacings $d = \lambda/(2\sin\theta)$ corresponding to the maximum observed 2θ values ($\lambda = 1.54$ Å for copper radiation in this example).

rather than being recombined to form an image (as shown in Figure 1.1), the image that is formed will be, to some degree, an imperfect representation of the scattering object. More particularly, fine detail will remain unresolved. Similarly, with X rays, if the diffraction pattern for the customary wavelengths is observed only out to a relatively small scattering angle, the resolution of the corresponding image reconstructed from it will be low. Furthermore, the resolution will be limited by the wavelength chosen even if the entire pattern is observed. Some examples of electron-density maps calculated with data out to a listed resolution are shown in Figure 6.6. As can be seen, lower numbers, indicating higher resolution, give more detailed pictures of the molecule. As in any process of image formation by recombination of scattered radiation, detail significantly smaller than the wavelength used cannot be distinguished by *any* scheme. On the other hand, the positions of well-resolved objects of known shape can be measured with high precision, and fortunately *all* interatomic distances are well resolved in three dimensions with the X rays we generally use. Hence, the positions of the resolved atoms can be measured and the details of molecular geometry calculated quite precisely.

The basic data in X-ray crystal studies

It is important to stress here which are the experimental data in an X-ray or neutron diffraction experiment. *The experimental results are the intensities of the diffracted beams* (combined with their indices *hkl*) and their conversion to $|F(hkl)|$ values. The relative phases are generally not measured, but are derived by the methods described in Chapters 8 and 9; isomorphous replacement and Renninger reflection measurements may, however, give some initial phase information. *The electron-density maps that follow are generally not primary experimental data* but are the

	d(Å)	Maximum 2θ(°)	Relative number of Bragg reflections included in each calculation
(1)	5.5	16	7
(2)	2.5	36	27
(3)	1.5	62	71
(4)	0.8	162	264

In each of the maps, the skeleton of the actual structure from which the data were taken has been superimposed. The first stage (1) (upper left) is typical of those encountered early in the determination of a protein structure. For protein structures, a degree of resolution between (2) and (3) is generally as much as is possible. The detail shown in (4) is characteristic of a structure determination with good crystals of low-molecular-weight compounds with radiation from an X-ray tube with a copper target. These electron-density maps may be compared to views of an object through a microscope, each corresponding to a different aperture (from Glusker et al., 1968). Note the lower peak heights of the carbon atoms compared with the nitrogen atoms. Also note that in the high-resolution structure, hydrogen atom locations are indicated.

results of estimated phase angles and, as shown in Figure 6.5, may or may not be correct. Therefore, if the final structure is not as expected, rereview the methods used to obtain the phases and to refine the proposed trial structure.

Summary

The electron density at a point x, y, z in a unit cell of volume V_c is

$$\rho(xyz) = \frac{1}{V_c} \sum_{\text{all } hkl} \sum \sum \left|F(hkl)\right| \cos[2\pi(hx + ky + lz) - \alpha(hkl)]$$

[see Eqn. (6.5)]. Therefore, if we knew $|F(hkl)|$ and $\alpha(hkl)$ (for each h, k, l) we could compute $\rho(x, y, z)$ for all values of x, y, and z and plot the values obtained to give a three-dimensional electron-density map. Then, assuming atomic nuclei to be at the centers of peaks, we would know the entire structure. However, we can usually obtain only the structure factor amplitudes $|F(hkl)|$ and not the relative phase angles $\alpha(hkl)$ directly from experimental measurements. *This is the phase problem*. We must usually derive values of $\alpha(hkl)$ either from values of $A(hkl)$ and $B(hkl)$ computed from suitable "trial" structures or by the use of purely analytical methods. In practice, approximations to electron-density maps can be calculated with experimentally observed values of $|F(hkl)|$ and calculated values of $\alpha(hkl)$. If the trial structure is not too grossly in error, the map will be a reasonable representation of the correct electron-density map, and the structure can be refined to give a better fit of observed and calculated $|F(hkl)|$ values. The discrepancy index R is one measure of the correctness of a structure determination. However, it is at best a measure of the precision of the fit of the model used to the experimental data obtained, not a measure of the accuracy. Some structures with low R values have been shown to be incorrect.

Symmetry and space groups

7

A certain degree of symmetry is apparent in much of the natural world, as well as in many of our creations in art, architecture, and technology. Objects with high symmetry are generally regarded with pleasure. Symmetry is perhaps the most fundamental property of the crystalline state and is a reason that gemstones have been so appreciated throughout the ages. This chapter introduces some of the fundamental concepts of symmetry—symmetry operations, symmetry elements, and the combinations of these characteristics of finite objects (point symmetry) and infinite objects (space symmetry)—as well as the way these concepts are applied in the study of crystals.

An object is said to be symmetrical if after some movement, real or imagined, it is or would be indistinguishable (in appearance and other discernible properties) from the way it was initially. The movement, which might be, for example, a rotation about some fixed axis or a mirror-like reflection through some plane or a translation of the entire object in a given direction, is called a symmetry operation. The geometrical entity with respect to which the symmetry operation is performed, an axis or a plane in the examples cited, is called a symmetry element. *Symmetry operations are actions* that can be carried out, while *symmetry elements are descriptions* of possible symmetry operations. The difference between these two symmetry terms is important.

It is possible not only to determine the crystal system of a given crystalline specimen by analysis of the intensities of the Bragg reflections in the diffraction pattern of the crystal, but also to learn much more about its symmetry, including its Bravais lattice and the probable space group. As indicated in Chapter 2, the 230 space groups represent the distinct ways of arranging identical objects on one of the 14 Bravais lattices by the use of certain symmetry operations to be described below. The determination of the space group of a crystal is important because it may reveal some symmetry within the contents of the unit cell. Space group determination also vastly simplifies the analysis of the diffraction pattern because different regions of this pattern (and hence of the atomic arrangement in the crystal) may then be known to be identical. Furthermore, symmetry greatly reduces the number of required calcu-

lations because only the contents of the asymmetric portion of the unit cell (the asymmetric unit) need to be considered in detail. In summary, the concept of the unit cell reduces the amount of structural information that needs to be determined for a crystal. It is not necessary to determine the locations of millions of molecules in a crystal experimentally, only the locations of those in one unit cell. The concept of the space group further reduces the information required to the "asymmetric unit," which is a portion of the unit cell that is defined by the space group of the crystal structure. Once the locations of atoms in the asymmetric unit are known, it is possible to calculate the positions of all other atoms in the unit cell and also of all those in the entire crystal by application of the space-group symmetry operations. These have been meticulously tabulated and are readily available in *International Tables* (Hahn, 2005).

Scrutiny of diffraction patterns of crystals reveals that there are often systematically related positions where diffraction maxima might occur but where, in fact, the observed intensity is zero. For example, if molecules pack in a crystal so that there is a two-fold screw axis parallel to the a axis, this means that each atom is moved a distance $a/2$ and then rotated 180° about the screw axis (from x, y, z to $\frac{1}{2}+x, -y, \frac{1}{2}-z$). A result is that for every atom at position x there is another at $\frac{1}{2}+x$. As far as $h\ 0\ 0$ Bragg reflections are concerned, the unit-cell size has been halved (to $a/2$) and the reciprocal lattice spacing has doubled (to $2a^*$). Bragg reflections will then only be observed for *even* values of h. This situation is made evident by summing in Eqns. (5.21) and (5.22) for atoms at x and $\frac{1}{2}+x$ when k and l are zero:*

$$A(h00) = f\cos 2\pi(hx) + f\cos 2\pi(hx + h/4)$$
$$= 2f\cos 2\pi(hx + h/4)\cos 2\pi(h/4) \qquad (7.1)$$
$$B(h00) = f\sin 2\pi(hx) + f\sin 2\pi(hx + h/4)$$
$$= 2f\sin 2\pi(hx + h/4)\cos 2\pi(h/4) \qquad (7.2)$$

$A(h00)$ and $B(h00)$ are both zero if h is odd, and therefore no Bragg reflection is observed. By contrast, if h is even, values may be found for $A(h00)$ and $B(h00)$.

Most, but not all, combinations of symmetry elements give rise to systematic relationships among the indices of some of the systematically "absent reflections." The word "systematically" implies some numerical relationship between the indices hkl of the Bragg reflections. For example, the only $h\,k\,0$ Bragg reflections with a measurable intensity may be those for which $(h+k)$ is even. Such systematic relationships imply certain symmetry relations in the packing in the structure. Before continuing with an account of methods of deriving trial structures, we present a short account of symmetry and, particularly, its relation to the possible ways of packing molecules or ions in a crystal.

* $\cos x + \cos y = 2\cos[(x+y)/2]\cos[(x-y)/2]$

$\sin x + \sin y = 2\sin[(x+y)/2]\cos[(x-y)/2]$

$\cos 0$ and $\cos 2\pi = 1, \cos\pi = -1$

$\cos\frac{\pi}{2}$ and $\cos\frac{3\pi}{2} = 0$

Symmetry groups

Any isolated object, such as a crystal, can possess "point symmetry". This term means that *any symmetry operation*, such as a rotation of, say, 180°, when applied to the object, *must leave at least one point within the object fixed* (unmoved). "Space symmetry" is different, because it includes *translational symmetry* (which is not permitted in a point group because this requires one point to be fixed in space). A translation operation is a space-symmetry operation; it leaves no point unchanged, since it moves all points equal distances in parallel directions. For example, an infinite array of points, such as a crystal lattice (or an ideal unbounded crystal structure), has translational symmetry, since unit translation (motion in a straight line, without rotation) along any unit crystal lattice vector moves the crystal lattice into self-coincidence. Because most macroscopic crystals consist of 10^{12} or more unit cells, it is a fair approximation to regard the arrangement of atoms throughout most (if not all) of a real crystal as possessing translational symmetry.

Symmetry elements (defined above as descriptions of possible symmetry operations) can be classified into groups (Cotton, 1971). A group, in the mathematical sense, is a set of elements, one of which must be the identity element, and the product of any two elements must also be an element in that same group. In addition, the order in which symmetry elements are combined must not affect the resulting element, and, for every element in the group, there must be another that is its inverse so that when the two are multiplied together the identity element is obtained. Studies of crystal symmetry involve point groups (one point unmoved when symmetry operations are applied) that are used in descriptions of crystals, and space groups (which also allow translational symmetry) that are used in descriptions of atomic arrangements within crystals.

Point symmetry and point groups

The operations of rotation, mirror reflection, and inversion through a point are point-symmetry operations, since each will leave at least one point of the object in a fixed position. The geometrical requirements of crystal lattices restrict the number of possible types of point-symmetry elements that a crystal can have to these three:

(1) *n-fold rotation axes*. A rotation of $(360/n)^\circ$ leaves the object or structure apparently unchanged (self-coincident). The order of the axis is said to be n, where n is an integer. When $n = 1$ (that is, a rotation of 360°), the operation is equivalent to no rotation at all (0°), and is said to be the "identity operation." A four-fold rotation axis, 90° rotation at each step, is shown in Figure 7.1, and is denoted by the number 4. It may be proved that only axes of

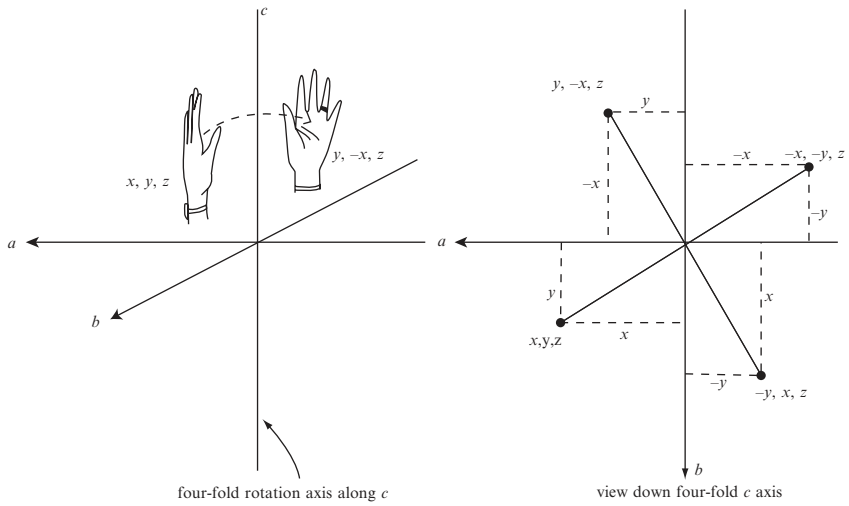
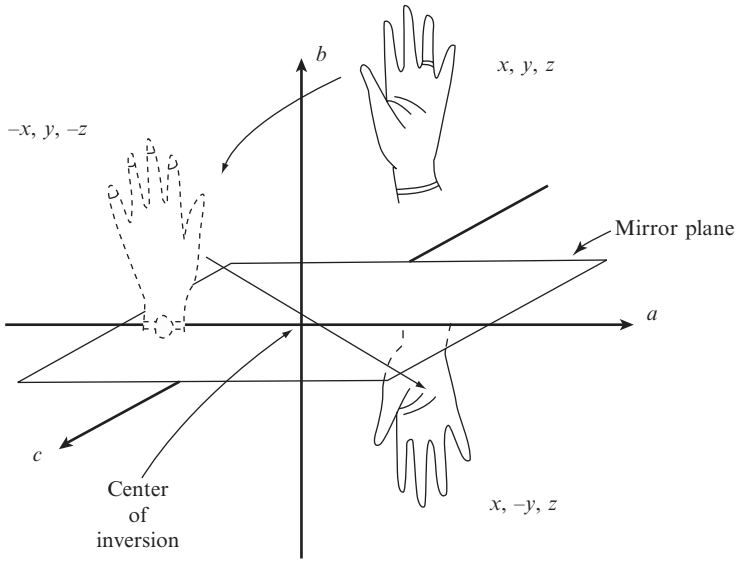


Fig. 7.1 A four-fold rotation axis.

In the figures in this chapter, in order to make the distinction of left and right hands clearer, a ring and watch have been indicated on the left hand but not on the right (even after reflection from the left hand). A four-fold rotation axis, parallel to c and through the origin of a tetragonal unit cell ($a = b$), moves a point at x, y, z to a point at $(y, -x, z)$ by a rotation of 90° about the axis. The sketch on the right shows all four equivalent points resulting from successive rotations; only two of these are illustrated in the left-hand sketch.

order 1, 2, 3, 4, and 6 are compatible with structures built on three-dimensional (or even two-dimensional) crystal lattices. Isolated molecules can have symmetry axes of other orders (5, 7, 8, or 17, for example), but when crystals are formed from a molecule with, for example, a five-fold axis of symmetry, *this five-fold axis cannot be a symmetry axis of the crystal*, although it can be a symmetry axis of the molecule. The molecule may still retain its five-fold symmetry in the crystal, but it can never occur at a position such that this symmetry is a necessary consequence of five-fold symmetry in the crystalline environment. In other words, five-fold symmetry is local and not crystallographic—that is, not required by any space group. This results from the requirement that there be no empty spaces in the packing in a crystal. Pentagonal tiles will not cover a floor without leaving untiled spaces.

(2) *n-fold rotatory-inversion axes*. The inversion operation, with the origin of coordinates as the "center of inversion," implies that every point x, y, z is moved to $-x, -y, -z$. From a point at x, y, z one could consider an imaginary straight line to proceed through the center of symmetry (at 0, 0, 0) and, further, an equal distance to $-x, -y, -z$. This inversion can also be augmented by a rotation to give an n-fold rotatory-inversion axis. This involves a rotation of $(360/n)^\circ$ (where n is 1, 2, 3, 4, or 6) followed by inversion through some point on the axis so that no apparent change in the object or structure occurs. The one-fold case, $\bar{1}$, is the inversion operation itself and is often merely called a center of symmetry. A two-fold rotatory-inversion axis, denoted $\bar{2}$, is



Two-fold rotatory-inversion axis *or* mirror plane

Fig. 7.2 A mirror plane.

The operation $\bar{2}$, a two-fold rotatory-inversion axis parallel to b and through the origin, converts a point at x, y, z to a point at $x, -y, z$. One way of analyzing this change is to consider it as the overall result of first a two-fold rotation about an axis through the origin and parallel to b (x, y, z to $-x, y, -z$) and then an inversion about the origin ($-x, y, -z$ to $x, -y, z$). This is the same as the effect of a mirror plane perpendicular to the b axis. Note that a left hand has been converted to a right hand. The hand illustrated by broken lines is an imaginary intermediate for the symmetry operation $\bar{2}$.

shown in Figure 7.2. In general, these axes are symbolized as $\bar{n}$. The rotatory-inversion operations differ from the pure rotations in an important respect; they convert an object into its mirror image. Thus a pure rotation can convert a left hand only into a left hand. By contrast, a rotatory-inversion axis will, on successive operations, convert a left hand into a right hand, then that right hand back into a left hand, and so on. Chiral objects that cannot be superimposed on their mirror images cannot possess any element of rotatory-inversion symmetry.

(3) *Mirror planes*. We are all familiar with mirrors. They convert a left-handed molecule into a right-handed molecule. As shown in Figure 7.2, a mirror plane is equivalent to a two-fold rotatory-inversion axis, $\bar{2}$, with the axis oriented perpendicular to the plane. The symbol m is more common for this symmetry element.

The point symmetry operations listed above (1, 2, 3, 4, 6, $\bar{1}$, $\bar{2}$ or m, $\bar{3}$, $\bar{4}$, and $\bar{6}$) can be combined together in just 32 ways in three dimensions to form the 32 three-dimensional *crystallographic point groups* (Phillips, 1963). There are, of course, other point groups, appropriate to isolated molecules and other figures, containing, for example, five-fold axes, but objects with such symmetry will have problems packing without gaps in three-dimensional space. The 32 crystallographic point groups or

symmetry classes may be applied to the shapes of crystals or other finite objects (Groth, 1906–1919); the point group of a crystal may sometimes be deduced by an examination of any symmetry in the development of faces. For example, a study of crystals of beryl shows that each has a six-fold axis perpendicular to a plane of symmetry ($6/m$), with two more symmetry planes parallel to the six-fold axis and at 30° to each other (mm). The corresponding point group is designated $6/mmm$. This external symmetry is a manifestation of the symmetry in the internal structure of the crystal. Frequently, however, the environment of a crystal during growth is sufficiently perturbed that the external form or morphology of the crystal does not reflect, to the extent that it might, the internal symmetry. Diffraction studies then help to establish the point group as well as the space group.

Space symmetry

A combination of the point-symmetry operations with translations gives rise to various kinds of space-symmetry operations, in addition to the pure translations.

(1) *n-fold screw axes*. A two-fold screw axis, 2_1, is shown in Figure 7.3. Screw axes result from the combination of translation (by distances such as $1/r$ of the repeat axis) and pure rotation (by an n-fold axis) and are symbolized by n_r. They involve a rotation of $(360/n^\circ)$ (where $n = 1, 2, 3, 4$ or 6) and a translation parallel to the axis by the fraction r/n of the identity period along that axis (where r is less than n and both are integers). If we consider a quantity $p = n - r$, then the axes n_r and n_p (such as 4_1 and 4_3 screw axes) are enantiomorphous; that is, they are mirror images of one another, like left and right hands. It is important, however, to note that it is only the *screw axes* that are enantiomorphous; structures built on them will not be enantiomorphous unless the objects in the structure are themselves enantiomorphous. Thus a left hand operated on by a 4_1 will give an arrangement that is the mirror image of that produced by the operation of a 4_3 on a right hand, but not, of course, the mirror image of that produced by the operation of a 4_3 on another left hand, as shown in Figure 7.4 (far left and far right).

(2) *Glide planes*. These symmetry elements result from the combination of translation with a mirror operation (or its equivalent, $\bar{2}$, normal to the plane), as illustrated in Figure 7.5. The glide must be parallel to some crystal lattice vector, and, because the mirror operation is two-fold, a point equivalent by a simple translational symmetry operation (a crystal lattice vector) must be reached after two glide translations. Thus these translations may be half of the repeat distance along a unit-cell edge, in which case the glide plane is referred to as an a-glide, b-glide, or c-glide, depending

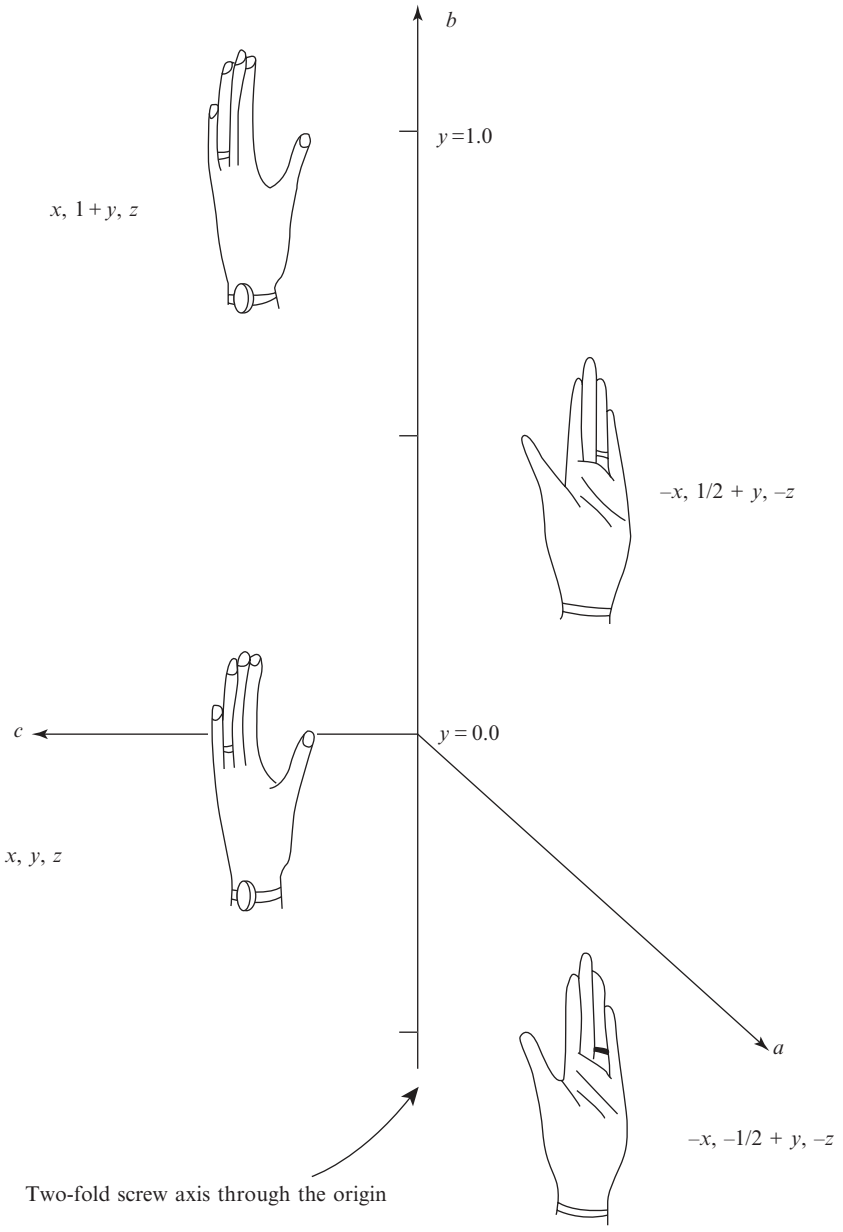


Fig. 7.3 A two-fold screw axis.

A two-fold screw axis, 2_1, parallel to b and through the origin, which combines both a two-fold rotation (x, y, z to $-x, y, -z$) and a translation of $b/2$ ($-x, y, -z$ to $-x, \frac{1}{2} + y, -z$). A second screw operation will convert the point $-x, \frac{1}{2} + y, -z$ to $x, 1 + y, z$, which is the equivalent of x, y, z in the next unit cell along b. Note that the left hand is never converted to a right hand by this screw axis.

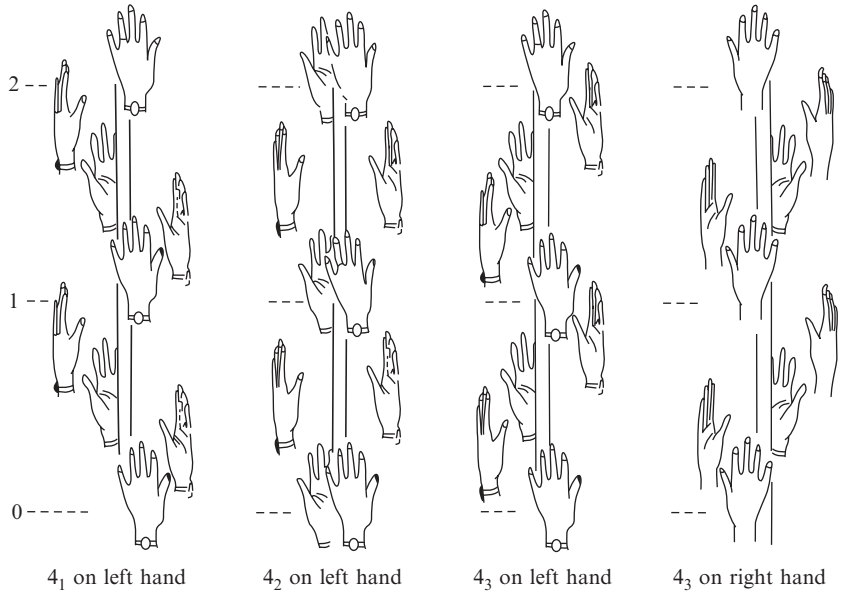
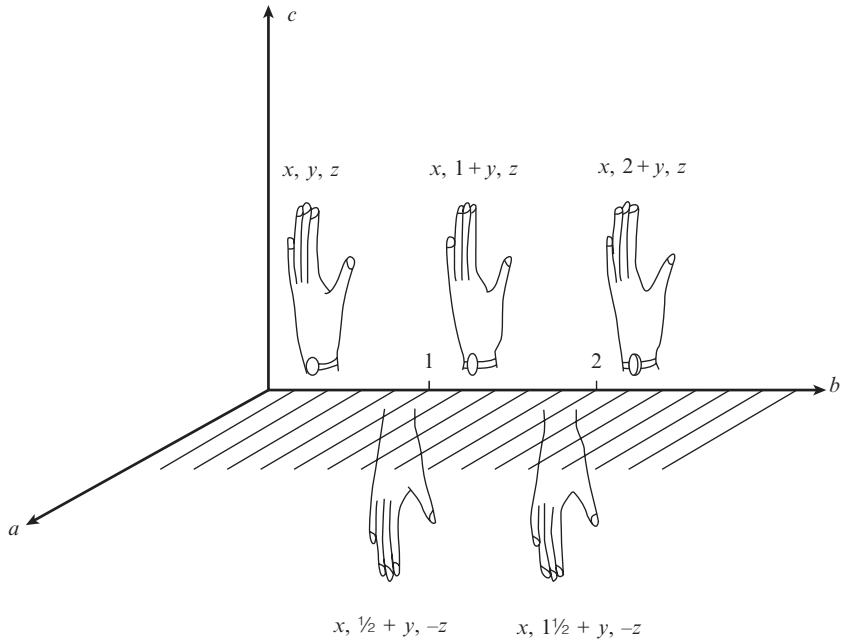


Fig. 7.4 A four-fold screw axis.

Some crystallographic four-fold screw axes showing two identity periods for each. Note that the effect of 4_1 on a left hand is the mirror image of the effect of 4_3 on a right hand.



b-glide plane through the origin and normal to *c*

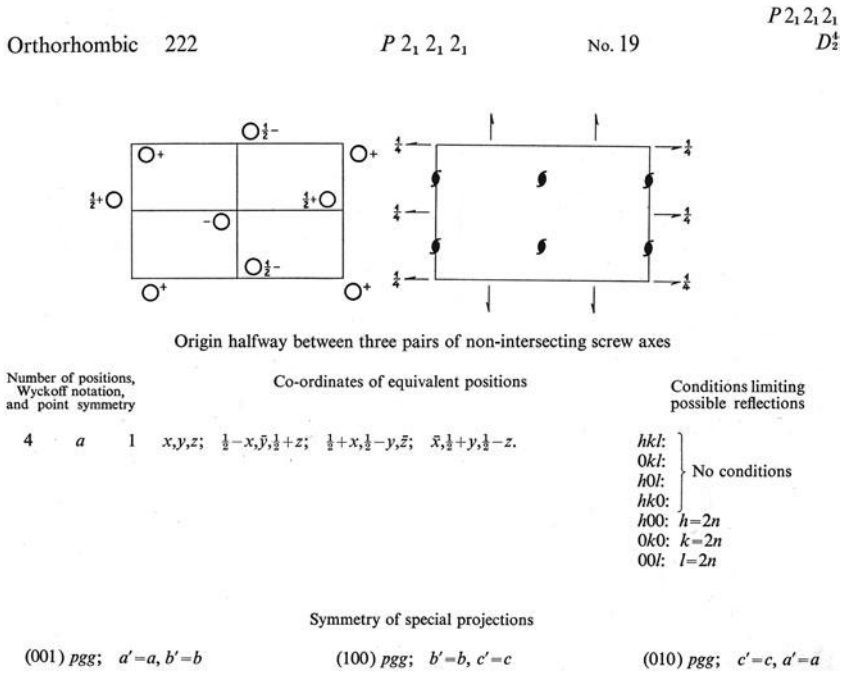
Fig. 7.5 A glide plane.

A *b*-glide plane normal to *c* and through the origin involves a translation of $b/2$ and a reflection in a plane normal to *c*. It converts a point at x, y, z to one at $x, \frac{1}{2} + y, -z$. Note that left hands are converted to right hands, and vice versa.

on the edge parallel to the translation. Alternatively, the glide may be parallel to a face diagonal. No glide operation involves fractional translational components other than 1/2 or 1/4, and the latter occurs only for glide directions parallel to a face diagonal or a body diagonal in certain nonprimitive space groups.

Space groups

We showed in Chapter 2 how an investigation of the symmetries of crystal lattices led to the seven crystal systems (triclinic, monoclinic, orthorhombic, tetragonal, hexagonal, rhombohedral, and cubic). These, when combined with unit-cell centering (face- or body-centering), gave the 14 Bravais lattices (see Appendix 2). If the 14 Bravais lattices are



Orthorhombic 222 $P\,2_1\,2_1\,2_1$ No. 19 $P2_12_12_1$ D_2^4

Origin halfway between three pairs of non-intersecting screw axes

Number of positions, Wyckoff notation, and point symmetry			Co-ordinates of equivalent positions	Conditions limiting possible reflections
4	*a*	1	x,y,z; $\frac{1}{2}-x,\bar{y},\frac{1}{2}+z$; $\frac{1}{2}+x,\frac{1}{2}-y,\bar{z}$; $\bar{x},\frac{1}{2}+y,\frac{1}{2}-z$.	*hkl*: No conditions 0*kl*: No conditions *h*0*l*: No conditions *hk*0: No conditions *h*00: $h=2n$ 0*k*0: $k=2n$ 00*l*: $l=2n$

Symmetry of special projections

(001) *pgg*; $a'=a, b'=b$ (100) *pgg*; $b'=b, c'=c$ (010) *pgg*; $c'=c, a'=a$

Fig. 7.6 Part of a page from *International Tables for X-Ray Crystallography*.

Information on the space group $P2_12_12_1$. The crystal is orthorhombic and there are three sets of mutually perpendicular nonintersecting screw axes. P denotes a primitive crystal lattice (that is, one lattice point per cell with no face- or body-centering) and 2_1 denotes a two-fold screw axis. The origin of the cell, chosen so that it lies halfway between these three pairs of nonintersecting screw axes, lies in the upper left-hand corner with the x direction down and the y direction across to the right; x is parallel to a and y is parallel to b. The symbol (◗) refers to a two-fold screw axis perpendicular to the plane of the paper. The symbol (¬) refers to a two-fold screw axis in a plane parallel to the plane of the paper; the fractional height of this plane above the plane $z = 0$ is shown (unless the screw axis is in the plane $z = 0$). The operations of the space group on the point (x, y, z) give three additional equivalent positions, whose coordinates are listed. Thus the screw axis parallel to c at $x = 1/4$, $y = 0$ converts an atom at x, y, z to one at $1/2 - x, -y, 1/2 + z$. Similar transformations are effected by the other two sets of screw axes (parallel to a and b, respectively). The diffraction patterns of crystals with this space group show systematic absences only for $h\ 0\ 0$ when h is odd, $0\ k\ 0$ when k is odd, and $0\ 0\ l$ when l is odd. Such crystals contain only molecules of one handedness (chirality). This diagram is from Volume 1 of *International Tables*. The current Volume A of *International Tables* contains the same information and more. Reproduced with permission of the International Union of Crystallography.

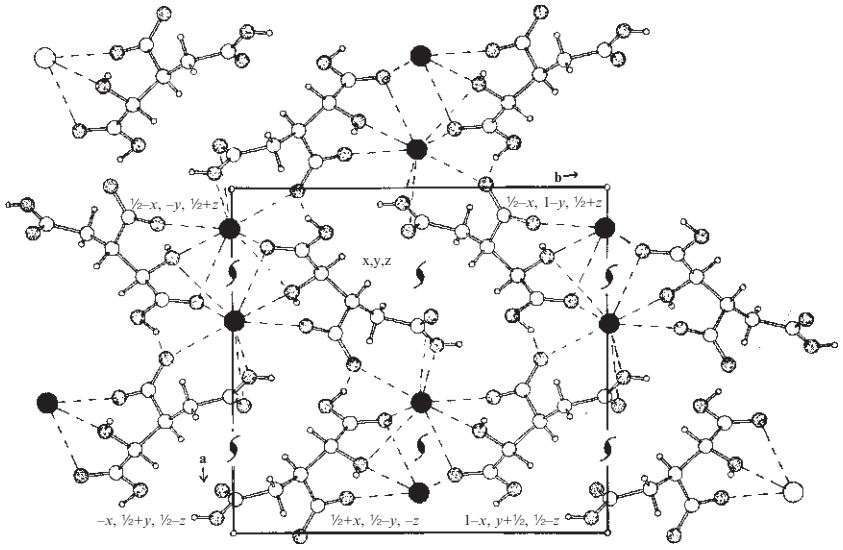


Fig. 7.7 A structure that crystallizes in the space group $P2_12_12_1$.

Contents of the unit cell of potassium dihydrogen isocitrate (van der Helm et al., 1968). The space group (see Figure 7.6) requires that, for each atom at x, y, z, there should be equivalent atoms at $\frac{1}{2} - x, -y, \frac{1}{2} + z$; $\frac{1}{2} + x, \frac{1}{2} - y, -z$; and $-x, \frac{1}{2} + y, \frac{1}{2} - z$. These are indicated on the diagram (oxygen stippled, potassium dark, hydrogen small). Interactions via hydrogen bonding and metal coordination are indicated by broken lines. This figure illustrates how anions cluster around a cation (dark spheres, K^+) and how this clustering, together with hydrogen bonding, is a major determinant of the structure.

combined with the symmetry elements of the 32 crystallographic point groups (involving reflection, rotation, and rotation-inversion symmetry), plus, in addition, the translational symmetry elements of glide planes and screw axes, the result is just 230 arrangements. These 230 space groups are compatible with the geometrical requirements of three-dimensional crystal lattices, that is, that the space-group symmetry should generate exactly the same arrangement of objects from unit cell to unit cell. There are thus 230 three-dimensional space groups, ranging from that with no symmetry other than the identity operation (symbolized by $P1$, the P implying primitive) to those with the highest symmetry, such as $Fm3m$, a face-centered cubic space group. These 230 space groups represent the 230 distinct ways in which objects (such as molecules) can be packed in three dimensions so that the contents of one unit cell are arranged in the same way as the contents of every other unit cell.

It is interesting to note that these 230 unique three-dimensional combinations of the possible crystallographic symmetry elements were derived independently in the last two decades of the nineteenth century by Evgraf Stepanovich Fedorov in Russia, Artur Moritz Schönflies in Germany, and William Barlow in England (Schoenflies, 1891; Fedorov, 1891; Barlow, 1894). It was not until several decades later that anything was known of the actual atomic structure of even the simplest crystalline solid. Since the introduction of diffraction methods for studying the structure of crystals, the space groups of many thousands of crystals

have been determined. It has been found that about 60% of the organic compounds studied crystallize in one of six space groups.**

All 230 space groups, and the systematically absent Bragg reflections found for them in the diffraction pattern, are listed in *International Tables*, Volume 1 or A, which is in constant use by X-ray crystallographers (Wyckoff, 1922; Astbury and Yardley, 1924; Hahn, 2005). Part of a specimen page from Volume 1 is shown in Figure 7.6. The symmetry operations in a space group must ensure that the next unit cell has the same contents as the original, and that it packs against the original unit cell with no gaps or spaces. Once the space group is determined from the systematically absent Bragg reflections in the X-ray diffraction pattern and by other means, if needed, *only the structure of the contents of the asymmetric unit*, not the entire unit cell, *need be determined*. The contents of the rest of the cell (and of the entire structure) are then known by application of the symmetry operations of the space group. An example is shown in Figure 7.7. An excellent way to obtain an introduction to space groups is to work one's way through the 17 plane groups listed just before the space groups in *International Tables for Crystallography*, Volume A, *Space-group Symmetry* (Hahn, 2005).

** The centrosymmetric space groups $P2_1/c$, $P\bar{1}$, $C2/c$, and $Pbca$ and noncentrosymmetric space groups $P2_12_12_1$ and $P2_1$.

Space group ambiguities

The principal method used to determine the space group of a crystal is that of determining which Bragg reflections are systematically absent in the space group. These are listed in *International Tables*, Volume A. As shown in an example at the beginning of this chapter, these systematic absences depend on the translational symmetry of the space group (screw axes, glide planes, face- or body-centering); that is, a two-fold screw axis resulted in systematic absences for $h00$ when h is odd. Therefore, space groups with the same translational symmetry elements (for example, $P2_1$ and $P2_1/m$)† will have the same systematic absences in their diffraction patterns, giving rise to an ambiguity in the determination of the space group.

† Equivalent positions for $P2_1$ are x, y, z and $-x, \frac{1}{2}+y, -z$. Equivalent positions for $P2_1/m$ are x, y, z; $-x, -y, -z$; $-x, \frac{1}{2}+y, -z$; and $x, \frac{1}{2}-y, z$.

However, there are ways of overcoming this problem. If the crystal contains only one enantiomorph of an asymmetric molecule, then the space group cannot contain a mirror or glide plane or a center of symmetry, since these symmetry elements convert one enantiomorph into the other. As a result, if the ambiguity involves a pair of space groups, one centrosymmetric and the other noncentrosymmetric (such as $P1$ and $P\bar{1}$ or $P2_1$ and $P2_1/m$), then a distinction can be made if the crystal contains molecules of only one chirality, since the crystal cannot then be centrosymmetric. In other cases the distinction can usually be made, as described in more detail in Chapter 8, by a consideration of the distribution of intensities in the diffraction pattern, since centrosymmetric structures have a higher proportion of Bragg reflections of very low intensity than do noncentrosymmetric structures. Other diagnostic methods involve tests of physical properties, including the

piezoelectric and pyroelectric effects (see Chapter 2). These effects are found only for noncentrosymmetric crystals. Still another method of distinguishing between space groups is to analyze the vectors in the Patterson map, described in Chapter 9. Finally, a consideration of the chemical identity of the contents of the unit cell may help resolve any space group ambiguity.

The following example of such an ambiguity may be of interest. The protein xylose isomerase, consisting of four identical subunits bound in a tetramer, crystallizes in the space group $I222$ or $I2_12_12_1$ with two molecules (eight subunits) in the unit cell. The systematic absences in Bragg reflections are, unfortunately, the same for both space groups, so here is an example of a space group ambiguity. This follows because the space-group absences for a body-centered unit cell are such that $h + k + l$ must be even, while the three screw axes require $h00$ with h even, $k00$ with k even, and $l00$ with l even; these last three requirements are included in the first condition, so that it does not make any difference to the systematic absences whether or not the screw axes are there. However, each unit cell for either space group contains eight asymmetric units, and therefore one subunit (one quarter of the molecule) must be the asymmetric unit (together with solvent, not considered here). If the space group were $I2_12_12_1$ the protein would be an infinite polymer, because of the requirements of the two-fold screw axes, contrary to physical evidence. Therefore the space group is $I222$, so that the subunits are related to each other by two-fold rotation axes rather than two-fold screw axes.

Chirality

Chirality is the handedness of a structure (Greek: *cheir* = hand); that is, if a structure cannot be superimposed on its mirror image it is said to be chiral or enantiomorphous. We are most familiar with this in the example of the asymmetric carbon atom—that is, a carbon atom connected to four different chemical groups so that two types of molecules, related to each other by a mirror plane, are found. This chirality, however, can also extend to the crystal structure itself. For example, silica crystallizes in a helical arrangement that has a handedness shown in the external shape of the crystal—small hemihedral[‡] faces appear in such a way as to give crystals that are mirror images of each other. The observation of such hemihedry was used by Louis Pasteur in 1848 to separate sodium ammonium tartrate into its left- and right-handed enantiomers (Pasteur, 1848; Patterson and Buchanan, 1945). Solutions of these pure enantiomers rotate the plane of polarization of light in opposite directions. When such resolution[§] occurs the space group must contain no mirror planes, glide planes, or centers of inversion (i.e., any symmetry operation that would convert a left-handed structure into a right-handed structure). Such crystals also exhibit pyroelectric and piezoelectric properties as a result of their asymmetry. Pasteur's

[‡] Called "hemihedral" because only half the number of faces expected for a centrosymmetric structure is observed.

[§] The term "resolution" is used in a different sense from that in the caption to Figure 6.6. Here it is used to mean the separation of enantiomers. The term is also used to describe the process of distinguishing individual parts of an object, as when viewing them through a microscope.

resolution of sodium ammonium tartrate was possible because the space group was one for a noncentrosymmetric structure, so the two crystal forms looked different. If an asymmetric molecule crystallizes in a centrosymmetric space group, then there are equal numbers of left- and right-handed molecules in the crystal structure. This will be discussed further in Chapter 10.

Space groups of chiral objects

Chiral molecules, such as proteins and nucleic acids, cannot crystallize in space groups with centers of symmetry, mirror planes, or glide planes, because otherwise molecules with the opposite chirality would also be required. There are 65 space groups that are suitable for such chiral molecules (see Appendix 7). In all, there are three types of space groups. 90 space groups are centrosymmetric and contain equal numbers of both enantiomers (left-handed and right-handed species) in the crystal. There are, however, 75 other space groups that are neither centrosymmetric nor chiral; that is, while the space group is noncentrosymmetric, the unit cell still contains equal numbers of both enantiomers (see Appendix 7). Some space groups among those for chiral molecules are designated as "polar space groups." They do not have a defined origin, for example, because, as in the space group $P2_1$, there is only one screw axis and it moves an atom at x, y, z to $-x$, $\frac{1}{2}+y$, $-z$ (by convention along the b axis). So y can have any value for the first atom in a list of atomic coordinates; its value then defines the origin that has been selected (but this is not defined by the space group), so that all other atoms are correctly related in space to the first atom. The polar space groups are indicated in Appendix 7. This polar property of the crystal must be remembered when atomic coordinates are being refined, as described later in Chapter 11. It also implies that opposite crystal faces perpendicular to the b axis may have different physical characteristics, as will be described in Chapter 10.

Summary

Symmetry in the contents of the unit cell is revealed to some extent by the symmetry of the diffraction pattern and by the systematically absent Bragg reflections (see Appendix 2). The probable space group of the crystal can be deduced from this information about the diffraction pattern. Knowledge of the space group may also give information on molecular packing, even before the structure has been determined.

(1) There are 14 distinct three-dimensional lattices (the Bravais lattices), corresponding to seven different crystal systems.

(2) Point-symmetry operations leave at least one point within an object fixed in space. Those characteristic of crystals consist of:

(a) n-fold rotation axes (1, 2, 3, 4, 6) and
(b) n-fold rotatory-inversion axes ($\bar{1}$, $\bar{2}$ or m, $\bar{3}$, $\bar{4}$, or $\bar{6}$).

(3) These point-symmetry operations can be combined in 32 and only 32 distinct ways to give the three-dimensional crystallographic point groups.

(4) Combination of point-symmetry operations with translations gives space-symmetry operations by way of:
(a) n-fold screw axes, n_r, and
(b) glide planes.

(5) All these operations may act on a given motif in the asymmetric portion of the structure. They can be combined in just 230 distinct ways, giving the space groups which can be used to describe crystal structures composed of multiple unit cells, each with identical structural components within them.

8

The derivation of trial structures. I. Analytical methods for direct phase determination

As indicated at the start of Chapter 4, after the diffraction pattern has been recorded and measured, the next stage in a crystal structure determination is solving the structure—that is, finding a suitable "trial structure" that contains approximate positions for most of the atoms in the unit cell of known dimensions and space group. The term "trial structure" implies that the structure that has been found is only an approximation to the correct or "true" structure, while "suitable" implies that the trial structure is close enough to the true structure that it can be smoothly refined to give a good fit to the experimental data. Methods for finding suitable trial structures form the subject of this chapter and the next. In the early days of structure determination, trial and error methods were, of necessity, almost the only available way of solving structures. Structure factors for the suggested "trial structure" were calculated and compared with those that had been observed. When more productive methods for obtaining trial structures—the "Patterson function" and "direct methods"—were introduced, the manner of solving a crystal structure changed dramatically for the better.

We begin with a discussion of so-called "direct methods." These are analytical techniques for deriving an approximate set of phases from which a first approximation to the electron-density map can be calculated. Interpretation of this map may then give a suitable trial structure. Previous to direct methods, all phases were calculated (as described in Chapter 5) from a proposed trial structure. The search for other methods that did not require a trial structure led to these phase-probability methods, that is, direct methods. A direct solution to the phase problem by algebraic methods began in the 1920s (Ott, 1927; Banerjee, 1933; Avrami, 1938) and progressed with work on inequalities by David Harker and John Kasper (Harker and Kasper, 1948). The latter

authors used inequality relationships put forward by Augustin Louis Cauchy and Karl Hermann Amandus Schwarz that led to relations between the magnitudes of some structure factors (see Glossary). These proved very useful, enabling them to derive relationships between the relative phases of different structure factors, and therefore to determine the crystal structure of decaborane (Kasper et al., 1950). This provided a previously unthought-of chemical structure for this molecule and greatly augmented our understanding of the structure and chemistry of the boron hydrides. Many scientists and mathematicians worked on the derivation of phase relationships in direct methods from this time on.* David Sayre provided an important equation that led to his demonstration of the structure of hydroxyproline (Sayre, 1952), while Herbert Hauptman and Jerome Karle worked on the probabilistic basis of direct methods (Karle and Hauptman, 1950; Hauptman and Karle, 1953). These, and the studies of many others, led to the equations of direct methods that are used today, and to the production of computer programs to do the analysis (Germain et al., 1971, for example) together with initially much-needed teaching on how to interpret the results of their use correctly.

"Direct methods" make use of two important facts: (1) that the intensities of Bragg reflections contain the structural information that peaks (representing atoms) are well resolved from each other (the principle of *atomicity*), and (2) that the background is fairly flat, and that this background should not be negative, because this would imply a negative electron density (the principle of *positivity*). These two conditions are true for X-ray diffraction, where atoms generally scatter by an amount that depends on their atomic number. The basic assumption that atoms are resolved from each other results in a requirement of high resolution, usually 1.1 Å or better, for direct methods. In the case of neutron diffraction, the electron-density map may have negative peaks because atoms, such as hydrogen, with a negative scattering factor for neutrons, are present. In spite of this, direct methods appear to work for neutron structures as well (Verbist et al., 1972). Centrosymmetric structures (with the positional coordinates of each atom at x, y, z, matched by those of an equivalent atom at $-x$, $-y$, $-z$) are considered first here, because the problems presented by noncentrosymmetric structures are more formidable. Techniques other than "direct methods" for deriving trial structures and the principles upon which they are based are discussed in Chapter 9.

It is possible to derive relations among the phases of different Bragg reflections. The basic assumption of direct methods is that the intensities in the X-ray diffraction pattern contain phase information (because the phases are constrained to give atomic peaks and positive electron density and this limits their values). It means that direct methods can be viewed as a mathematical problem—the control of the phase angles of density waves because of the principles of atomicity and positivity. How can the many density waves be aligned (as required by their

* There have been many involved in the development of direct methods, in the programming of methods to use them, and in teaching people how to do it. These include (in alphabetical order), in the earlier stages, William Cochran, Joseph Gillis, David Harker, Herbert Hauptman, Isabella Karle, Jerome Karle, John S. Kasper, Peter Main, David Sayre, George Sheldrick, Michael Woolfson, and William H. Zachariasen. Many others also merit our appreciation of the ease with which crystal structures can generally be determined.

individual phases) so that the resultant electron-density map shows peaks or a flat background with no negative areas? What must their phases be to satisfy these conditions?

In crystal structures with a center of symmetry at the origin and no appreciable anomalous-dispersion effects, each structure factor has a phase angle of 0° or 180°, so that $\cos\alpha$ is just +1 or −1 and $\sin\alpha = 0$. Therefore, in a centrosymmetric structure, $F = |F|\cos\alpha = +|F|$ or $-|F|$, and one often speaks of the *sign of a structure factor*; when the phase angle α is 0° we write "+" and when it is 180° we write "−". If N Bragg reflections have been observed for the structure, 2^N electron-density maps would need to be calculated, representing all possible combinations of signs for all N independent structure factors. One of these 2^N maps must represent the true electron density, but how could one tell which one it is? For even as few as twenty Bragg reflections, more than one million different maps would need to be calculated (2^{20} = 1,048,576), and most structures of interest have of the order of 10^3–10^6 unique Bragg reflections. Since the contributions from Bragg reflections with high values for the structure factor amplitude will tend to dominate any electron-density map calculated, only the most intense Bragg reflections need be considered initially when one is trying to obtain an approximation to the correct map. However, even with as few as ten terms, the number of possible maps is 1024, much too high a number to make any simple trial-and-error method practicable. With a noncentrosymmetric crystal structure, a phase angle may be anywhere between 0° and 360° and one would have to calculate an impossibly large number of maps to ensure having at least approximately correct phase angles for even ten Bragg reflections.

Relationships can be found among the signs of the structure factors, and these relationships involve the magnitudes of the larger structure factors normalized (that is, modified) in a certain way, as will be described in this chapter. If you want to know what a given structure factor of known relative phase contributes to the overall electron density in a unit cell (its density wave), it is easy to plot this. Suppose that $F(1\ 0\ 0)$ for a centrosymmetric structure is large (see Figure 8.1). If this Bragg reflection has a positive sign (phase angle of 0°), then the computed electron-density map has a peak near the origin at $x = 0$ and a hole at $x = \frac{1}{2}$. By contrast, if this Bragg reflection has a negative sign, there is a peak at $x = \frac{1}{2}$ and a hole at $x = 0$. Therefore the fact that this Bragg reflection is intense in a centrosymmetric structure implies that there must be a peak in the electron-density map near either $x = 0$ or $x = \frac{1}{2}$, whatever the sign (phase) to be associated with $F(1\ 0\ 0)$. If $F(2\ 0\ 0)$ is considered, it can be seen in Figure 8.1 that a peak at either 0 or $\frac{1}{2}$ implies a positive sign for $F(2\ 0\ 0)$. Consequently, *if* $F(1\ 0\ 0)$ *and* $F(2\ 0\ 0)$ *are intense*, $F(2\ 0\ 0)$ *is probably positive* no matter whether the sign of $F(1\ 0\ 0)$ is positive or negative. Figure 8.1 shows also that when only these terms are summed, a positive sign for $F(2\ 0\ 0)$ results in an "electron density" that has a shallower negative trough than does

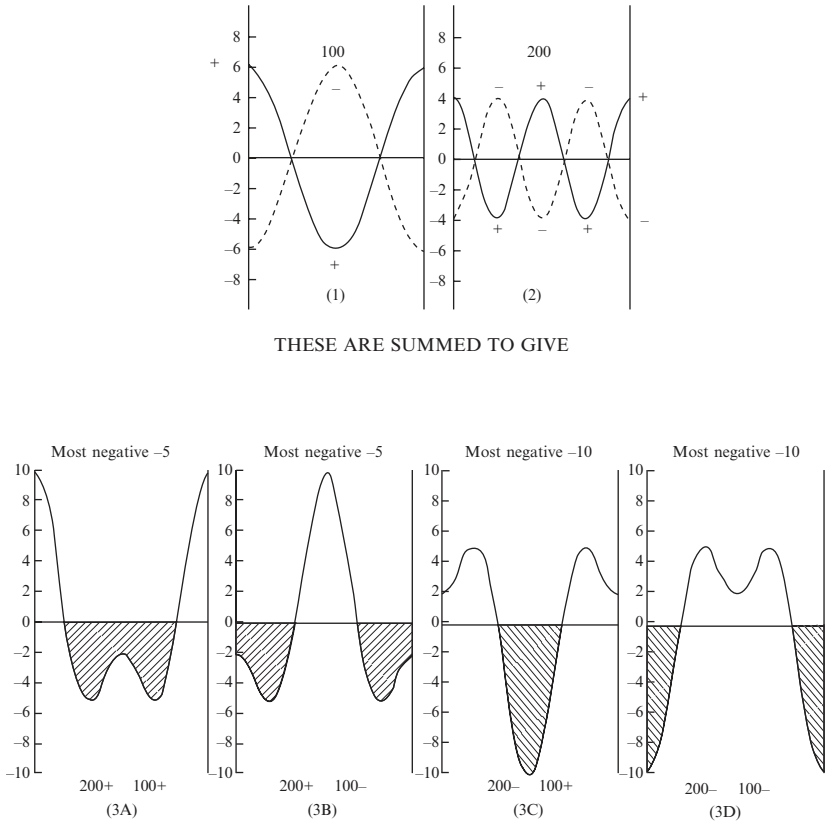


Fig. 8.1 Summing density waves.

In centrosymmetric structures, the phase angle of any structure factor $F(hkl)$ is either 0° or 180°. "Electron-density maps" based on one structure factor ("density wave") are shown for $F(1\ 0\ 0)$ and $F(2\ 0\ 0)$. In general, for centrosymmetric structures, if $F(hkl)$ is large, whatever its sign, and $F(2h2k2l)$ is also large, then the latter is probably positive (a phase of 0°). (1) Possible situations for $F(1\ 0\ 0)$: solid line—F positive, phase 0°; dotted line—F negative, phase 180°. (2) Possible situations for $F(2\ 0\ 0)$; solid line—F positive, phase 0°; dotted line—F negative, phase 180°. (3) Summations for the four combinations of possible situations in (1) and (2), showing the deep negative areas obtained when $F(2\ 0\ 0)$ is given a phase of 180° (C, D). The $F(0\ 0\ 0)$ term, which has been omitted, is always positive and therefore when it is included the sum is always more positive at each given point (see Figure 6.2).

Areas of negative electron density are shaded. The inferences on the position of an atom (at x) from these electron-density maps are: 3A, $x = 0$; 3B, $x = 1/2$; 3C, 3D, $x = 1/4, 3/4$. The last two have more negative troughs and so are excluded. Therefore we conclude 200 is + (phase angle 0°) and x is 0 and/or $1/2$.

the electron density that results when a negative sign is assigned to $F(2\ 0\ 0)$ (regardless of the sign of $F(1\ 0\ 0)$). Thus the phase of $F(2\ 0\ 0)$ is probably +.

The principle of positivity of electron density may be extended to three dimensions. For example, David Sayre noted that the functions $\rho(r)$ and $\rho^2(r)$ in a crystal composed of identical atoms are similar in appearance. From analyses of the relationship between these two

functions and of their Fourier transforms, he showed that

$$\sum \sum_{K} \sum F(K)F(H-K) = V\mathrm{s}F(H) \tag{8.1}$$

This is the equation that bears his name (Sayre, 1952; see also Viterbo, 1992; Shmueli, 2007). In this equation $H = h, k, l$ and $K = h', k', l'$; V is the unit cell volume; s is the sign of the *hkl* Bragg reflection; and the summations are over all values of K.

If one considers probabilities (denoted $\approx$), rather than certainties (denoted =), it can be shown that, for a centrosymmetric structure, one obtains a *triple product*

$$\mathrm{s}F(H)\,\mathrm{s}F(K) \approx \mathrm{s}F(H+K) \tag{8.2}$$

where sF means the "sign of F" and $F(H)$, $F(K)$, and $F(H+K)$ are *all* intense Bragg reflections. The symbol $\approx$ means "is probably equal to." It should be noted that a special case of Eqn. (8.2) is

$$\mathrm{s}(2h\ 0\ 0) \approx [\mathrm{s}(h\ 0\ 0)]^2 \approx + \tag{8.3}$$

because whatever the sign of $F(h00)$, its square is positive. This is in agreement with our qualitative argument for $F(2\,0\,0)$ and $F(1\,0\,0)$ above and in Figure 8.1. In Figure 8.2 it is shown that if $F(3\ 0\ 0)$ is known to be positive and $F(2\ 0\ 0)$ is known to be negative, then, if all three are strong Bragg reflections, $F(5\ 0\ 0)$ is probably (but not definitely) negative. Again, this is shown to be consistent with the principle of positivity of electron density. Two types of sets of triple products of phases (see Eqn. 8.2) merit attention at this point. A "structure invariant" is a linear combination of the phases that is totally independent of the choice of origin; even if the origin is changed, the invariant remains unchanged. The same is true for "structure seminvariants" except that the origin change must be one that is allowed by space-group symmetry constraints. The identification of structure invariants and seminvariants helps to fix an origin and enantiomorph for the structure under study.

In practice, these analytical methods of phase determination are carried out on "normalized structure factors"—that is, values of the structure factor $|F(hkl)|$ modified to remove the fall-off in the individual scattering factors f with increasing scattering angle 2θ (see Figures 5.4 and 8.3). A normalized structure factor, $E(hkl)$, represents the ratio of a structure factor $F(hkl)$ to $(\Sigma f_j)^{1/2}$, where the sum is taken over all atoms in the unit cell at the value of $\sin\theta/\lambda$ appropriate to the values of h, k, and l for the Bragg reflection and includes an overall vibration factor. This sum, $(\Sigma f_j)^{1/2}$, represents the root-mean-square value that all $|F(hkl)|^2$ measurements would have (at that value of $\sin\theta/\lambda$) if the structure were a random one, composed of equal atoms (see the discussion of the Wilson plot at the end of Chapter 4):

$$\left|E(hkl)\right| = \frac{\left|F(hkl)\right|}{\left(\varepsilon \sum f_j\right)^{1/2}} \tag{8.4}$$

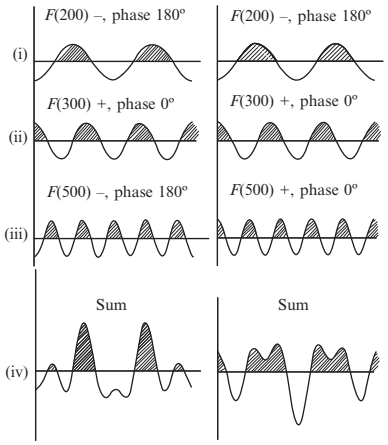


Fig. 8.2 Aiming for nonnegative electron density.

If $|F(2\,0\,0)|$, $|F(3\,0\,0)|$, and $|F(5\,0\,0)|$ are all large they must contribute significantly to the final electron-density map (via "density waves"). Suppose that it is found that $F(2\ 0\ 0)$ has a negative sign and $F(3\ 0\ 0)$ has a positive sign; the areas in which each then contributes in a positive manner to the electron-density map are shaded in (i) and (ii) on the left. The regions in which these areas overlap, near $x = \pm 0.3$, correspond to regions to which $F(5\ 0\ 0)$ contributes positively only if the sign of the term $F(5\ 0\ 0)$ is negative, that is, a phase of 180°, as indicated in (iii). On summation of these terms with the indicated signs the background is reduced, as in (iv); if $F(5\ 0\ 0)$ has a positive sign, that is, a phase of 0°, the map is far less satisfactory. The relation among these signs may then be written (where s means "the sign of")

$$\mathrm{s}(5\ 0\ 0) \approx \mathrm{s}(2\ 0\ 0)\mathrm{s}(3\ 0\ 0)$$

which is a special case of Eqn. (8.2). This follows from the discussion in the text since deep negative troughs (areas of negative electron density) are not satisfactory or physically meaningful. With proper phasing, the background is reduced to a value closer to zero. Thus in (iv) on the left the most negative value of the electron density is −4 e/Å, while for (iv) on the right, which has a less satisfactory set of phases, the most negative value of the electron density is −9 e/Å. The addition of data for $F(000)$ will probably result in an almost nonnegative map if $F(500)$ has a phase of 180°.

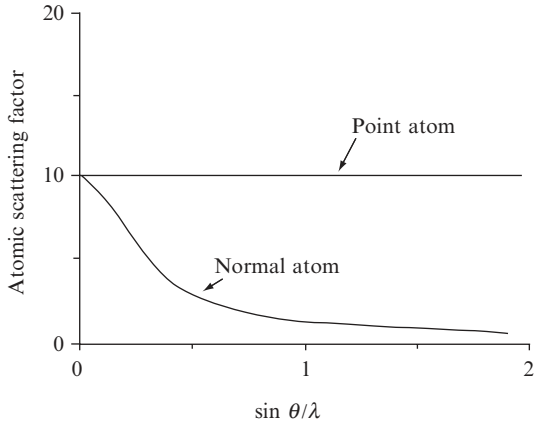


Fig. 8.3 X-ray scattering by point atoms and normal atoms.

Theoretical point atoms have no width and no vibrational amplitude. As a result there is no fall-off in value as $\sin\theta/\lambda$ increases. By contrast, real normal atoms have width and vibrational amplitude and their atomic scattering factors fall off at high $\sin\theta/\lambda$.

where ε is a constant (an epsilon factor)**, 1, 2, or 4, depending on the crystal class, and the summation is from $j = 1$ to N. This use of $E(hkl)$ values is approximately equivalent to considering each atom to be a point atom (an extremely sharp peak occupying a very small volume in the electron-density map). As a result, high-order Bragg reflections (high $\sin\theta/\lambda$), which normally are weaker because of the intensity fall-off of atomic scattering factors f with $\sin\theta$ values, may have large $|E|$ values that would play an appropriate role in the structure determination (rather than being ignored because of their low values when $|F|$ is used).

** Contained in the expression for $E(hkl)$ is a factor, ε, that corrects for the fact that Bragg reflections in certain reciprocal lattice zones or rows (for example, $0\,k\,0$, $h\,k\,0$, etc.) have higher average intensities in certain space groups than do general Bragg reflections (hkl). It is an integer, 1, 2, or 4, depending on the crystallographic point group and the type of Bragg reflection (h, k, and/or $l = 0$).

Information on significant features of the structure is contained in the very intense and very weak Bragg reflections; these have different distributions when the structure is centrosymmetric and when it is noncentrosymmetric (Wilson, 1949). The centrosymmetric distribution has a higher proportion of Bragg reflections with very low intensities. An analysis of the $E(hkl)$ values in the diffraction pattern (the "distribution of $E(hkl)$ values") shows that they contain (as, of course, do the $F(hkl)$ values as well) information on whether the structure is centrosymmetric or noncentrosymmetric. For example, the mean value of $E(hkl)$ is 0.798 for a centrosymmetric structure and 0.886 for a noncentrosymmetric structure. The value that the crystallographer more commonly uses for this test is $|E^2 - 1|$, which is theoretically 0.968 for a centrosymmetric structure and 0.736 for a noncentrosymmetric structure. These values are calculated from the diffraction data that have just been measured, and they probably will indicate which symmetry the crystal has.

Once a table of $|E(hkl)|$ values has been prepared, it is usual to rank these $E(hkl)$ values in decreasing order of magnitude. Usually one works with the strongest 10 percent of them. Then one looks for groups of three Bragg reflections that satisfy the condition that their indices

are numerically related in the manner described in Eqn. (8.2); this selection of "triple products" $[E(hkl), E(h', k', l'), E(h + h', k + k', l + l')]$ is generally made by computer. If each of the three Bragg reflections in a triple product has a high $E(hkl)$ value, the product of their three signs is probably positive. This listing is called the "Σ_2" or "sigma 2" listing (see Eqn. (8.2) and Figure 8.2). The summation symbol Σ is used in this naming because, in the probability formula, summations are involved. The "Σ_1" relations (see Eqn. (8.3) and Figure 8.1) are simpler because they involve only pairs of intense Bragg reflections related by $E(hkl)$ and $E(2h, 2k, 2l)$ and contain the implication that the sign of $E(2h, 2k, 2l)$ is probably positive in a centrosymmetric structure (see Figure 8.1 and Eqn. (8.3)). The general equation used is

$$\mathrm{s}[E(hkl)] \approx \left[\sum_{h',k',l'} E(h', k', l') E(h - h', k - k', l - l') \right] \tag{8.5}$$

The probability aspects of these sign relationships are very important. If we replace h, k, l by H and h', k', l' by K, the probability that a triple product is positive in a centrosymmetric structure (that is, $\mathrm{s}_H \approx \mathrm{s}_K \mathrm{s}_{H-K}$) is[†]

$$\mathrm{P}_+ = \frac{1}{2} + \frac{1}{2} \tanh \left(\frac{E_H E_K E_{H-K}}{N^{1/2}} \right) \tag{8.6}$$

(where N is the number of equal atoms in the unit cell[‡]). Furthermore, the probability that $E(hkl) \equiv E_H$ is positive is

$$\mathrm{P}_+ = \frac{1}{2} + \frac{1}{2} \tanh \left(|E_H| \sum_K \frac{E_K E_{H-K}}{N^{1/2}} \right) \tag{8.7}$$

where the summation Σ is over all values of $K = (h', k', l')$, and $\mathrm{P}_+ = 1$ indicates a sign of +1, while $\mathrm{P}_+ = 0$ indicates that it is 0.[§] These probability aspects of direct methods result in a requirement for a large amount of diffraction data.

[†] tanh, the hyperbolic tangent of x, is $\{(\mathrm{e}^x - \mathrm{e}^{-x})/(\mathrm{e}^x + \mathrm{e}^{-x})\}$.

[‡] We advance from Eqn. (8.5) to Eqn. (8.6) by incorporating a commonly used abbreviation that has developed in the literature of direct methods: $H \equiv (h, k, l)$, $K \equiv (h', k', l')$, and hence $H + K \equiv (h + h', k + k', l + l')$. Note that, since $-K$ and K have the same E's (in sign and magnitude) in a centrosymmetric structure, then a relation between H and K and a relation between H and $-K$ are equivalent.

[§] For unequal atoms, $(1/N^{1/2})$ in Eqns. (8.6) and (8.7) is replaced by $\sigma_3 \sigma_2^{-3/2}$, where $\sigma_n = \Sigma Z_j^n$, the summation being from 1 to N, and N is the number of atoms with atomic number Z_j for the jth atom.

Solving the structure of a centrosymmetric structure

We will now describe the steps in the determination of a centrosymmetric structure by direct methods. When the list of "triple products" $[E(hkl), E(h'k'l')$, and $E(h + h', k + k', l + l')]$ has been prepared, the derivations of their signs requires some initial choices of signs. Initially, in three dimensions, one has a choice of the signs of three Bragg reflections for many centrosymmetric space groups; these choices determine which of the possible positions is used for the origin of the unit cell. The choice does not alter the structure, it just defines where the unit-cell origin is. In selecting three origin-fixing Bragg reflections,

An illustration of the successive application of Eqn. (8.2) to derive the signs of some of the strongest Bragg reflections for a structure. Values of $|E|$ are derived from those for $|F|$, chiefly to eliminate the effects of thermal vibration and to treat each atom as if all its electrons were concentrated at a point. Values of $|E|$ are used in deriving sign relationships because their magnitudes depend only on the arrangement and relative atomic numbers of the atoms.

Monoclinic example:

$$|F(hkl)| \neq |F(\bar{h}\,kl)|$$

$k+l$ even

$$s(hkl) = s(h\bar{k}\bar{l}) = s(\bar{h}k\bar{l}) = s(\bar{h}\bar{k}\bar{l})$$

$$s(\bar{h}\,kl) = s(\bar{h}\bar{k}l) = s(hk\bar{l}) = s(h\bar{k}\bar{l})$$

$k+l$ odd

$$s(hkl) = s(\bar{h}\bar{k}\bar{l}) = -s(h\bar{k}l) = -s(\bar{h}k\bar{l})$$

$$s(\bar{h}kl) = s(h\bar{k}\bar{l}) = -s(hk\bar{l}) = -s(\bar{h}\bar{k}l)$$

The compound studied is 2-keto-3-ethoxybutyraldehyde-bis (thiosemicarbazone), space group $P2_1/c$. The above sign relationships for this space group are to be found in *International Tables*, Volume A.

Relation to be used [Eqn. (8.2)]:

$$s(h+h', k+k', l+l') \approx s(h,k,l)\, s(h',k',l')$$

h k l	*E*	
3 3 1	3.74	+
$\bar{9}$ 6 7	3.25	+
$\overline{13}$ 1 4	2.92	+

Signs chosen arbitrarily fixing the origin. (If one or all of these signs had been negative another allowable origin would have resulted.)

h k l	*E*	*Relationships used*	*Sign found*	*(Notes)*
12 0 0	4.35	(6 0 0)(6 0 0)	+	(+)(+) = (−)(−) = +
6 0 0	2.80		?	
$\overline{25}$ 1 4	3.49	($\overline{12}$ 0 0)($\overline{13}$ 1 4)	+	++ = +
$\overline{22}$ 4 5	2.22	(3 3 1)($\overline{25}$ 1 4)	+	++ = +
6 4 2	2.86		a	An additional undetermined sign is chosen and is temporarily designated a
18 4 2	2.92	(12 0 0)(6 4 2)	a	+a = a
9 7 3	2.07	(3 3 1)(6 4 2)	a	+a = a
$\overline{22}$ 6 1	2.30	($\overline{13}$ $\bar{1}$ 4)($\bar{9}$ 7 $\bar{3}$)	−a	−(+a) = −a
$\overline{19}$ 3 2	2.84	(3 $\bar{3}$ 1)($\overline{22}$ 6 1)	−a	+(−a) = −a
		($\overline{13}$ $\bar{1}$ 4)($\bar{6}$ 4 $\bar{2}$)		−(+a) = −a
$\bar{7}$ 3 2	2.14	(12 0 0)($\overline{19}$ 3 2)	−a	+(−a) = −a
25 1 0	2.03	(18 4 2)(7 $\bar{3}$ $\bar{2}$)	−	a(−a) = −
		(6 4 2)(19 $\bar{3}$ $\bar{2}$)		a(−a) = −

Eventually the sign of some Bragg reflection could be found both in terms of a and independently of it; this established the fact that the sign of a was probably +. If this had not happened it would have been necessary to calculate two maps, one with the sign of a positive and one with the sign of a negative.

Fig. 8.4 Numerical use of Eqn. (8.2) to derive phases of a crystal structure.

it is essential for them to be different with respect to the evenness or oddness of their individual indices, and h, k, and l must not all be even. In the numerical example in Figure 8.4, arbitrary signs were chosen for F(3 3 1) (odd, odd, odd), $F(\bar{9}\ 6\ 7)$ (odd, even, odd), and $F(\overline{13}\ 1\ 4)$ (odd, odd, even) at the start.

The reader may ask where negative signs for phases come from. The relationships of signs of Bragg reflections with negative values of h, k, and/or l to that of a Bragg reflection with all indices positive are listed for each space group in *International Tables*, Volume A. For example, in the space group $P2_1/c$, if $k + l$ is odd, then $F(hkl) = -F(h\bar{k}\,l) = -F(\bar{h}\,k\bar{l})$. Negative signs are introduced into the sign relationships in this way. It is essential to have some negative terms in the calculation of the E-map, because an E-map with all signs positive will give a high peak at the origin, a rarely observed feature in complex structures; in fact this E-map with all signs positive resembles a Patterson function (to be described in the next chapter), but cannot be interpreted as an electron-density map.

From the list of "triple products" it should be possible to derive, for the set of $E(hkl)$ values, a set of signs that have been determined with acceptable probabilities (see the example in Figures 8.4 and 8.5). If difficulties occur, it may be necessary to choose another set of origin-fixing Bragg reflections. It may also be necessary to assign symbolic signs ("a", "b", etc.) to certain Bragg reflections and generate the signs of other Bragg reflections in terms of these symbols with the hope that eventually the actual signs of these symbols may become clear. This process is referred to as "symbolic addition" (Zachariasen, 1952; Karle and Karle, 1966). For example, in Figure 8.4 it is deduced that the sign of symbol "a" is positive. If n symbols have been used but their signs cannot be determined in this way, it will be necessary to compute 2^n E-maps.

The derivation of signs for a monoclinic centrosymmetric structure is shown in Figure 8.4. Some sign relationships could be immediately deduced from a knowledge of the monoclinic space group relationships among $F(hkl)$ appropriate for this structure. Others were then deduced from these new signs and some arbitrarily chosen signs. This process was continued until the signs of 836 out of the 872 strongest terms were found with no sign ambiguity (although it is usually not necessary to work with this many terms). Part of the resulting Fourier synthesis computed using $E(hkl)$ values (an E-map) is shown in Figure 8.5.

The crystal structure of hexamethylbenzene was reported by Kathleen Lonsdale in 1928 and showed that the benzene ring is planar and has six-fold symmetry (Lonsdale, 1928). The arrangement of atoms in

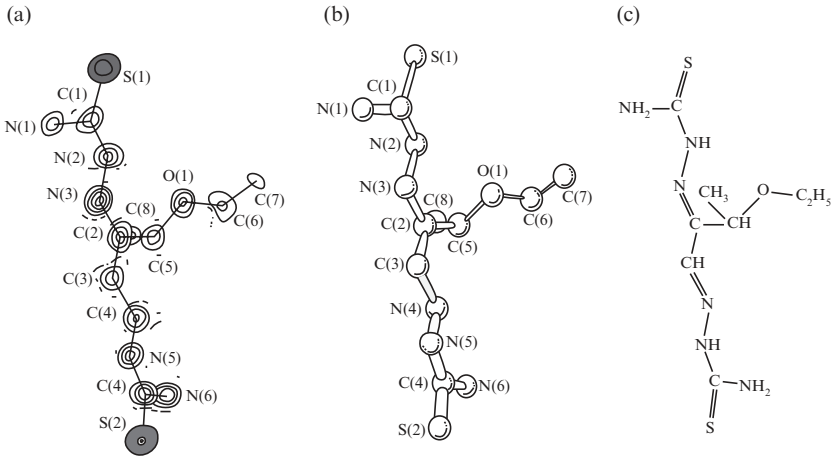


Fig. 8.5 An excerpt from an E-map.

(a) A three-dimensional map calculated with phases derived as in Figure 8.4 and $|E|$ values rather than values of $|F|$ as amplitudes (see Gabe et al., 1969). This is a composite map; each peak has been drawn as it appears in the section in which it has the highest value. It is a simple matter to pick out the entire molecule, 2-keto-3-ethoxybutyraldehyde-bis(thiosemicarbazone), from this map. The molecular skeleton and the presumed identity of each atom have been added to the peaks. (b) A ball-and-stick drawing of the molecule. (c) The chemical formula of the molecule.

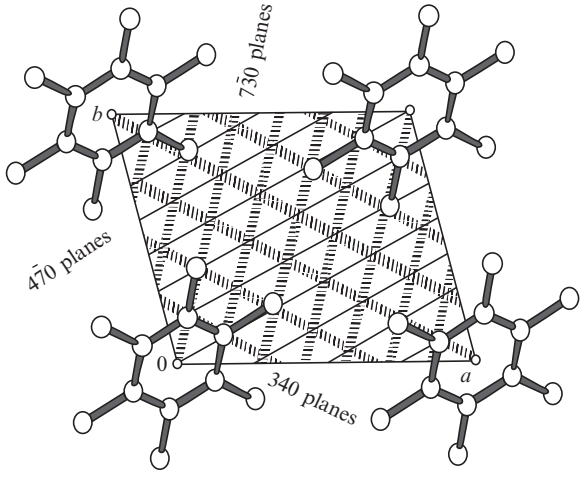


Fig. 8.6 A triple product in diffraction by hexamethylbenzene.

Shown is the crystal structure of hexamethylbenzene (Phase II) (Lonsdale, 1928). Maxima of the density waves of three intense Bragg reflections, 340, 7 $\bar{3}$ 0, and 4 $\bar{7}$ 0, are diagrammed with the molecular structure superimposed. Note that all of the carbon atoms lie at the intersection of maxima of the density waves of these three Bragg reflections.

the unit cell is shown in Figure 8.6. Three strong Bragg reflections, 340, 7 $\overline{3}$ 0 and 4 $\overline{7}$ 0, form a triple product (inspect the indices) and, although direct methods were not used in 1928, they illustrate the principle. The maxima of the density waves of these three Bragg reflections are shown in this figure. Note how the carbon atoms each lie on the intersections of three density-wave maxima.

The final stage is the calculation of an E-map. This is an electron-density map calculated with $E(hkl)$ values rather than $F(hkl)$ values (so that atoms are sharper, corresponding to point atoms) (see Figure 8.5a). If all has gone well, the structure will be clear in this map. Sometimes only part of the structure is revealed in an interpretable way and the rest may be found from successive electron-density maps or different electron-density maps. Sometimes the general orientation and connectivity of the molecule are found, but the positioning in the unit cell is wrong because some subsets of signs are in error. This problem is usually recognizable when distances between atoms are calculated and some nonbonded atoms are too close to others. In this case, the development of signs must be done again, this time following some new path, such as selecting origin-fixing Bragg reflections or assigning symbols to a different set of Bragg reflections.

Solving the structure of a noncentrosymmetric crystal

The derivation of phases for noncentrosymmetric structures is more complicated because the values for the phases are not simply 0° or 180°. For noncentrosymmetric crystal structures, an additional formula may be used to derive approximate values for the phase angle ϕ_H:

$$\phi_H \approx \sum_K (\phi_{H-K} + \phi_K) \tag{8.8}$$

where, as before, $H \equiv h, k, l$; $K \equiv h', k', l'$; ϕ is the phase angle of the structure factor; and the brackets refer to an average over all values of K, where $H = (K) + (H - K)$. The so-called "tangent formula,"

$$\tan \phi_H = \frac{\sum_K (|E_K||E_{H-K}| \sin(\phi_K + \phi_{H-K}))}{\sum_K (|E_K||E_{H-K}| \cos(\phi_K + \phi_{H-K}))} = \text{tangent formula} \tag{8.9}$$

is used extensively to calculate and also to refine phases for noncentrosymmetric structures. The probability function for

noncentrosymmetric structures is more complicated than that given in Eqns. (8.6) and (8.7). It is:

$$P(\phi_H) = \frac{\exp[-4x\cos(\phi_H - \phi_K - \phi_{H-K})]}{\int_0^{2\pi} \exp(4x\cos\gamma)\, d\gamma} \tag{8.10}$$

where $x = |E_H E_K E_{H-K}|/N^{1/2}$ and γ is a dummy variable. Higher-order structure invariants and seminvariants (quartets, quintets, for example) are also used in structure determination by direct methods. These are sets of more than three Bragg reflections with indices that have a zero sum. A so-called "negative quartet,"

$$\phi_H + \phi_K + \phi_L + \phi_{-H-K-L} = \pi \tag{8.11}$$

has a phase sum that is probably near 180° rather than 0° (Hauptman, 1974). It is useful not only in phase determination, but also for finding the correct solution if there are several possibilities. Some important computer programs currently in use for determining crystal structures include SHELXS and SHELXD (Sheldrick, 2008), Shake-and-Bake (Miller et al., 1993; Miller et al., 1994), SIR (Burla et al., 2005), and SUPERFLIP (Palatinus and Chapuis, 2007), but there are many more. The reader is advised to consult the World Wide Web for the most suitable program for use for a current problem. In addition, much useful information is provided in the various volumes of *International Tables* (see the reference list).

Dual-space algorithms, which involve iterative cycles of Fourier transforms between real and reciprocal space with changes at each step, have proved very useful in the determination of the structures of macromolecules. Some important methods of phase improvement for proteins involve "density modification." If the boundaries between solvent and protein have been determined in the electron-density map, the relative phases can be improved by "flattening" the solvent area ("solvent flattening"). Improved phases are then obtained by a Fourier transform (Hoppe and Gassmann, 1968; Wang, 1985; Leslie, 1987). Another example is provided by "real-space averaging" or "non-crystallographic symmetry averaging," in which electron densities (in real space) of two units are averaged. "Histogram matching" can also be applied to protein structures; in this, the initial electron densities are modified to conform to an expected distribution. The Shake-and-Bake (SnB) program is a phase-determining procedure for solving crystal structures by direct methods, and it has been incorporated into SHELXD (Schneider and Sheldrick, 2002). It alternates phase refinement in reciprocal space by use of the minimal principle (the shake) with real-space constraints through some form of electron-density modification (the bake) (Miller et al., 1993; Miller et al., 1994). The minimal

principle involves a residual described as a "minimum-variance, phase invariant." All phases are initially determined by computation from a random atomic arrangement and are refined by minimizing this residual. They are then Fourier transformed, and peaks are selected from the resulting electron-density map and used for a new trial structure for the next cycle of the method. While used successfully for small structures, the SnB program has also provided *ab initio* solutions (meaning no preliminary experimentally determined phase information but good resolution of 1.1 Å or higher) for protein structures involving as many as 1000 independent nonhydrogen atoms. Another dual-space structure determination method is the charge-flipping algorithm, which is an iterative process that requires a complete set of diffraction intensities to atomic resolution, but does not require any information on the symmetry or chemical composition of the crystal structure (Oszlányi and Süto, 2005; Palatinus and Chapuis, 2007). A random set of phases is assigned to the measured structure factors and their Fourier transform is calculated. All electron densities that fall below a selected positive value (to be selected by the user) are inverted (the charge-flipping step). The modified electron-density map is then Fourier transformed and the new phases from the charge-flipped map are combined with the original observed data, the structure amplitudes $|F(hkl)|$. Then follows a new iteration cycle. The process is repeated until a satisfactory structure is obtained. Results can be checked, for example, by "random omit maps," in which a selected proportion, say one third, of the highest peaks in an electron-density map are deleted, and the remaining atoms are used to calculate new phases and start a new cycle (Bhat and Cohen, 1984; Bhat, 1988). Then one can check if the deleted atoms appear in the new electron-density map.

In an attempt to improve the resolution of a measured data set, the "free lunch algorithm" (also called "nonmeasured reflection extrapolation") was introduced by Eleanor Dodson, developed by Carmelo Giacovazzo, and named by George Sheldrick "since one is apparently getting something for nothing" (Caliandro et al., 2005; Yao et al., 2005). It extends the resolution of the measured data significantly by simply inventing the missing data. This is done by Fourier transformation of the existing experimental F_o map using the Wilson plot to obtain the overall scale factor for the made-up data. Unexpectedly, it was found that the introduction of unmeasured structure amplitudes produced phases that improved the resulting electron-density map and led, in many cases, to a structure solution (Usón et al., 2007). For some reason random structure amplitudes are better than zero structure amplitudes for those high-resolution data that cannot be measured experimentally (Caliandro et al., 2007; Dodson and Woolfson, 2009).

Overview

Direct methods for both centrosymmetric and noncentrosymmetric structures have been programmed for many high-speed computers. Since the equations involve probabilistic rather than exact relations, uses of direct methods are most successful when care is taken initially in the choice of the origin-fixing and symbolically assigned phases. These are used to determine the phases of a good number of intense Bragg reflections. In many structure analyses a reasonable approximate ("trial") structure has been recognizable from an E-map calculated with only 5 or 10 percent of the observed Bragg reflections, although often larger fractions are used, as in the example illustrated in Figures 8.4 and 8.5. Generally these "direct methods" result in a structure that can be refined (Chapter 11), and so the structure may be considered to be determined. A variety of excellent computer programs generally ensure a correct structure. For several reasons, however, such success may be elusive with some structures. There are many possible problems that can arise in using these methods, such as a poor choice of origin-fixing Bragg reflections, the derivation of too few triple products so that some signs are generated with lower probabilities than one would like, and a preponderance of positive signs for the derived signs so that the resulting E-map has a huge peak at the origin even though there is no heavy atom in the structure. However, with care and experience these problems can usually, although not always, be overcome.

Summary

There are limits to the possible phase angles for individual Bragg reflections in both centrosymmetric and noncentrosymmetric structures. This follows from the constraints on the electron density; it must be nonnegative throughout the unit cell and it must contain discrete, approximately spherical peaks (atoms). For three intense related Bragg reflections in a centrosymmetric structure, the signs are related by

$$\mathrm{s}F(H) \approx \mathrm{s}F(K)\mathrm{s}F(H+K)$$

where s means "sign of"; $H \equiv h, k, l$; $K \equiv h', k', l'$; $H + K \equiv h + h', k + k', l + l'$; and F is a structure factor or E value. From such relationships it is often possible to derive phases for almost all strong Bragg reflections and so to determine an approximation to the structure (a "trial" structure) from the resulting electron-density map. Similar methods are available for noncentrosymmetric structures.

The steps in the determination of a structure by "direct methods" consist of:

(1) Making a list of E values in decreasing order of magnitude and working with the highest 10 percent or so.

(2) Analysis of the statistical distribution of E values to determine if the structure is centrosymmetric or noncentrosymmetric. This is important if there is an ambiguity in the space group determined from systematically absent Bragg reflections.

(3) Derivation of triple products among the high E values.

(4) Selection of origin-fixing Bragg reflections.

(5) Development of signs or phases for as many E values as possible using triple products and probability formulae.

(6) Calculation of E-maps and the selection of the structure from the peaks in the map.

All of these steps are now incorporated into computer programs in wide use.

9 The derivation of trial structures. II. Patterson, heavy-atom, and isomorphous replacement methods

The two methods to be described here, the Patterson method* and the isomorphous replacement method, have made it possible to determine the three-dimensional structures of large biological molecules such as proteins and nucleic acids. In addition, the Patterson function is still useful for small-molecule studies if problems are encountered during the structure analysis. If a crystal structure determination proves to be difficult, the Patterson map should be determined to see if it is consistent with the proposed trial structure.

The Patterson method involves a Fourier series in which only the indices (h, k, l) and the $|F(hkl)|^2$ value of each diffracted beam are required (Patterson, 1934, 1935). These quantities can be obtained directly by experimental measurements of the directions and intensities of the Bragg reflections. The Patterson function, $P(uvw)$, is defined in Eqn. (9.1). It is evaluated at each point u, v, w in a three-dimensional grid with axes u, v, and w that are coincident with the unit-cell axes x, y, z; the grid fills a space that is the size and shape of the unit cell:

$$P(uvw) = \frac{1}{V_c} \sum_{\text{all } h,k,l} \sum \sum \left|F(hkl)\right|^2 \cos 2\pi(hu + kv + lw) \qquad (9.1)$$

No phase information is required for this map, because $|F(hkl)|^2$, unlike $F(hkl)$, is independent of phase. *There is only one Patterson function for a given crystal structure*. For reasons that we explain shortly, a plot of this function is often called a vector map. Appendix 8 gives some useful background information and further details.

The Patterson function at a point u, v, w may be thought of as the convolution** of the electron density with itself in the following manner:

* In the three decades from the mid 1930s to the mid 1960s, the most powerful method of analysis of the diffraction pattern of a crystal was the Patterson method. It revolutionized structure determination because no longer was it necessary to propose a correct trial structure before analysis. For the first time it provided a means for solving most structures if good diffraction data were available.

** See the Glossary.

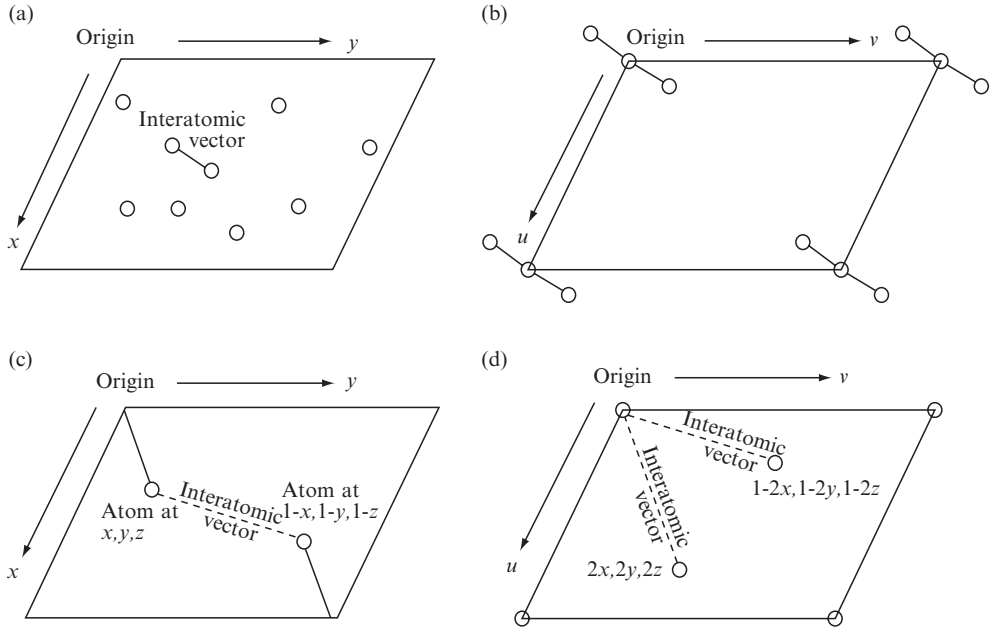


Fig. 9.1 Peaks in a Patterson (vector) map.

A Patterson map represents all interatomic vectors in a crystal structure, positioned with one end of the vector at the origin of the Patterson map. (a) Atoms in a crystal structure showing one interatomic vector, which will appear as shown in (b) in the Patterson map. (c) Two atoms related by a center of symmetry in a crystal structure. (d) The corresponding Patterson map showing vector coordinates.

$$P(uvw) = V_c \iiint_{\text{whole cell}} \rho(x, y, z)\rho(x + u, y + v, z + w)\, dx\, dy\, dz \qquad (9.2)$$

Equation 9.2 is obtained by multiplying the electron density at all points x, y, z in the unit cell (that is, $\rho(x, y, z)$) with the electron density at points $x + u$, $y + v$, and $z + w$ (that is, $\rho(x + u, y + v, z + w)$). This Patterson function, $P(u, v, w)$, can be thought of as the sum of the appearances of the structure when one views it from each atom in turn, a procedure illustrated in Figure 9.1. It is as if an atomic-scale elf sat on an atom, took a snapshot of his surroundings, then moved to the next atom and superimposed his second snapshot on the first, and so forth.† Essentially the Patterson map samples the crystal structure at all sites separated by a vector u_0, v_0, w_0 and notes if there is electron density at both ends of this vector; if this is so an interatomic vector has been localized. Therefore, if any two atoms in the unit cell are separated by a vector u_0, v_0, w_0 in the three-dimensional structure (or electron-density map), there will be peak in the Patterson map at the site u_0, v_0, w_0.

The Patterson map [Eqns. (9.1) and (9.2)] is flat, near zero, except for peaks that represent the orientation and length of every interatomic vector in the structure. The vector between any two atoms is the distance between them and the direction in space that a line connecting them would take. The heights of the peaks in the Patterson map are

† H. F. Judson, in *The Eighth Day of Creation* (Judson, 1996), uses the analogy of a cocktail party in describing the Patterson function. If there are one hundred guests at a party, there must have been one hundred invitations. The host would have to make almost five thousand introductions if he wanted to be sure everyone met each other, and this would involve ten thousand attempts to remember a new name. If the shoes of the guests are nailed to the floor, their handshakes must involve different lengths and directions of arms and different strengths of grip. This analogy may help some readers understand the meaning of the vectors in a Patterson map; they are interatomic vectors of different lengths and directions, with heights proportional to the product of the atomic numbers of the atoms at each end of the vector. If each partygoer could then recount every handshake and the direction, distance, and strength of it, then the location of every guest in the room would be known. Of course one would only use this very complicated method (five thousand vectors to locate one hundred people) if it were absolutely necessary.

proportional to the values of $Z_i Z_j$, where Z_i is the atomic number of the atom, i, at one end of the vector and Z_i is that of the atom, j, at the other end. The high peak that occurs at the origin of the Patterson function represents the sum of all the vectors between an individual atom and itself. It is important to note that a Patterson map is centrosymmetric whether or not the structure itself is centrosymmetric.‡ This is because a vector from atom B to atom A has the opposite direction but the same magnitude as a vector from atom A to atom B, so that these two vectors, A → B and B → A, are related by a center of symmetry. The symmetry of a Patterson map is generally not the same as that of the electron-density map for the same crystal structure, but is like the Laue symmetry. Symmetry elements containing translations (glide planes and screw axes) are replaced by mirror planes or simple rotation axes, respectively, and there is always the center of symmetry just described.

‡ This center of symmetry is evident in Figure 9.1c.

If there is a peak in the Patterson map at a position related to the origin of the map by a certain vector (with components u, v, w, corresponding to a certain distance and direction from the origin), then *at least one position of that particular vector in the corresponding crystal structure has both ends on atomic positions*. (Remember that a vector is characterized by a certain length and direction, but its origin may be anywhere). If there are many pairs of atomic positions related by a particular vector, or if there are only a few but the atoms involved have high atomic numbers, then the Patterson function will have a high peak at that particular position u, v, w. If the value of the Patterson function at a given position is very low, there is no interatomic vector in the structure that has that particular length and direction.

The Patterson map for a one-dimensional structure with identical atoms at $x = \pm 1/3$ is shown in Figure 9.2. The values of the function given by Eqn. (9.1) are designated $P(u)$, and positions in the one-dimensional map by u. An interesting feature of this map is that the same result would be obtained from a structure in which the atoms were at $x = \pm 1/6$. As shown in Figure 9.2, these two structures differ only in that the location of the origin of the unit cell has been changed; the *relative* positions of the atoms are the same in both solutions of the map.

Thermal motion and disorder of atoms will cause a broadening of the vector peaks and a lowering of their heights in the Patterson map. This broadening can be reduced by "sharpening" the peaks. One method of doing this is an artificial conversion of the atoms to point scatterers by dividing each $|F(hkl)|$ by the average scattering factor for the value of $\sin\theta/\lambda$ at which it was measured. Normalized structure factors $|E(hkl)|$ fit this criterion and are commonly used with unity subtracted from their square so that the origin Patterson peak will be removed. This means that the coefficients used to compute the map are modified from $|F(hkl)^2|$ to $|E(hkl)^2| - 1$. The resulting origin-removed, sharpened

Patterson map is

$$P(uvw) = \frac{1}{V_c} \sum_{\text{all } h,k,l} \sum \sum \left| E(hkl)^2 - 1 \right| \cos 2\pi(hu + kv + lw) \qquad (9.3)$$

There are areas in a Patterson map, called "Harker sections" or "Harker lines," where symmetry operators involving translational components (such as screw axes or glide planes) lead to useful information, especially if a heavy atom is present (Harker, 1936). Therefore if the space group lists atoms at x, y, z and $\frac{1}{2} - x$, $-y$, $\frac{1}{2} + z$, there will be peaks at $w = \frac{1}{2}$ in the Patterson map and they represent vectors

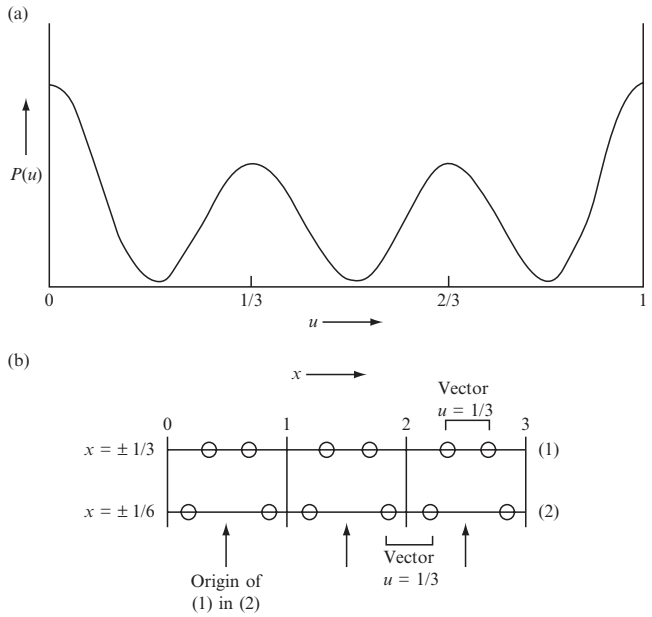


Fig. 9.2 The calculation of a Patterson map for a one-dimensional structure.

(a) The equation of the Patterson function in one dimension is

$$P(u) = \frac{1}{a} \sum_{\text{all } h} |F|^2 \cos 2\pi(hu)$$

The function plotted is $P(u)$ computed for a one-dimensional structure from the following hypothetical "experimental" data:

h	-3	-2	-1	0	1	2	3
$\lvert F \rvert^2$	4	1	1	4	1	1	4

(b) There are two structures consistent with this map, one with atoms at $x = \pm\frac{1}{3}$ and one with atoms at $x = \pm\frac{1}{6}$. As shown, these two structures are related simply by a change of origin.

between symmetry-related atoms at z and at $z + \frac{1}{2}$. Therefore a perusal of the Patterson map at $w = \frac{1}{2}$ for a structure with this particular space group may help solve the structure, especially if a heavy atom is present.

A problem with Patterson maps is that there are N^2 interatomic-vector peaks within a unit cell that contains N independent atoms. N of these peaks lie at the origin and, since the Patterson map has a center of symmetry, there are $(N^2 - N)/2$ independent vectors in the map. When N becomes at all large (even as small as 20), the $(N^2 - N)/2$ vector peaks in the Patterson map necessarily overlap one another, since they have about the same width as atomic peaks and occupy a volume equal to that occupied by the N atoms of the structure. For example, when $N = 20$ there are $20 \times 19/2 = 190$ Patterson peaks in the same volume that the 20 atomic peaks occupy in the electron-density map. With crystals of very large molecules, such as proteins, the overlap may become hopeless to resolve, except for the peaks arising from the interactions between atoms of very high atomic number, since a Patterson peak has a height proportional to the product of the atomic numbers of the two atoms involved in the vector it represents.

The structure shown in Figure 6.6, for which the Patterson map is shown in Figure 9.3, contains only 12 nonhydrogen (O, N, or C) atoms in the asymmetric unit. The great complexity of the Patterson map compared with the electron-density map is obvious. In this example, similar orientations of the six-membered rings in space-group-related molecules give rise to very similar sets of six interatomic vectors, the vectors in each set having nearly the same magnitude and direction, thereby giving a high peak in the Patterson map (see peaks B, C, and D in Figures 9.3a and b). Similarly oriented five-membered rings also lead to high peaks (peaks A and E). The slope of the ring system is clear in Figures 9.3c and d. This figure demonstrates the large amount of structural information available in a Patterson map. However, since all nonhydrogen atoms in the structure are similar in atomic number, and the chemical formula was unknown until the structure was determined, the Patterson map was too complicated to analyze when first obtained. Some Patterson maps that were much easier to interpret will be described later in this chapter.

Until the advent of computer-assisted direct methods in the late 1960s, analysis of Patterson maps was the most important method for getting at least a partial trial structure, especially for crystals containing one or a few atoms of atomic number much higher than those of the other atoms present. In principle, for all but the largest structures, a correct trial structure can always be found from the Patterson distribution, but it is often very difficult to unravel the map, especially when the chemical formula of the compound being studied is not known. Some people, however, find it a fascinating mental exercise to try to deduce at least part of a crystal structure from a Patterson map.

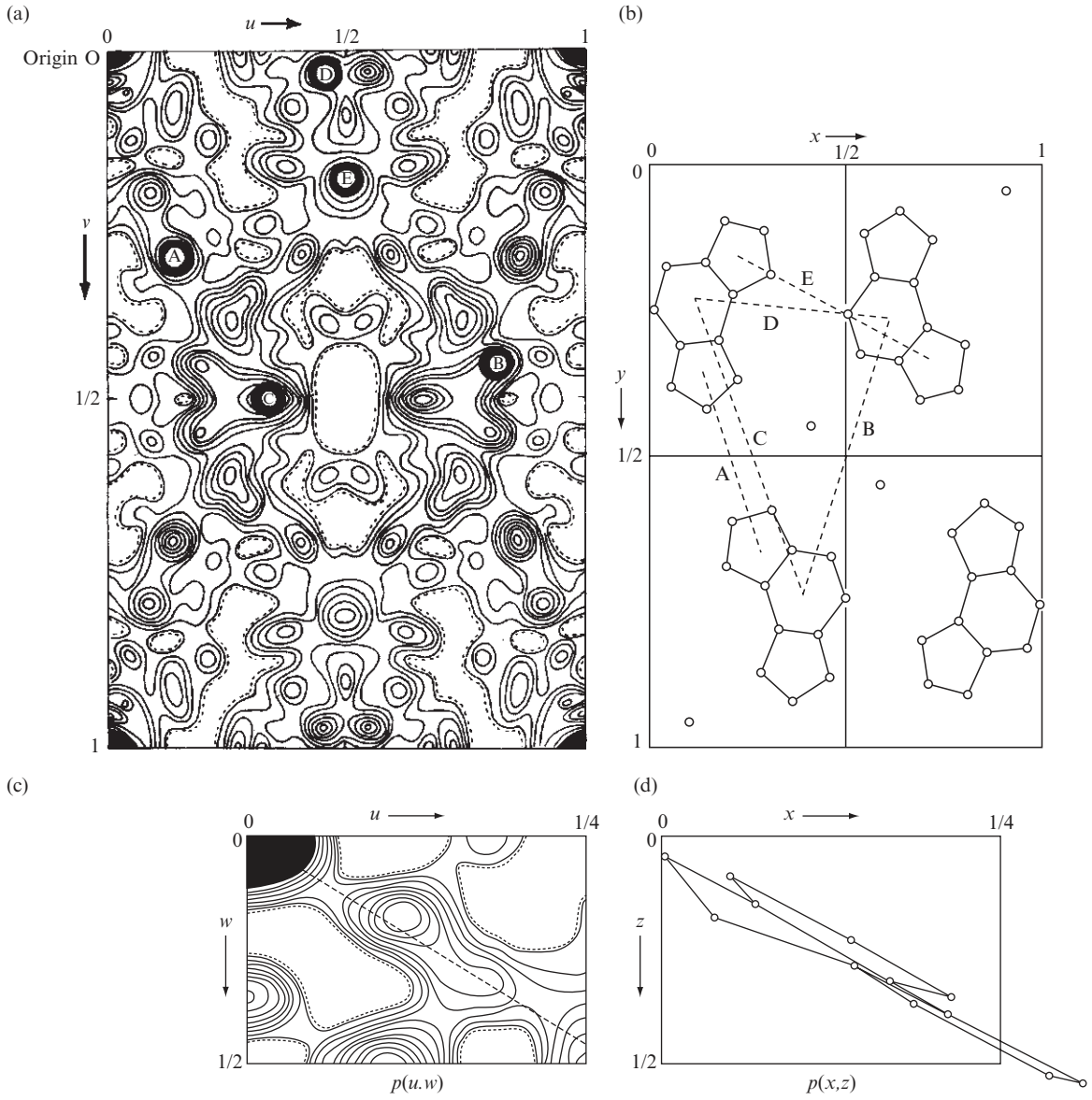


Fig. 9.3 The analysis of a Patterson map.

(a) A two-dimensional Patterson map, $P(u, v)$, a projection down the w axis, of an azidopurine is shown. The peaks in the $P(u, v)$ map that correspond to the multiple superposition of vectors from ring to ring are lettered A to E and are shown in both (a) and (b).
(b) The interpretation of the map shown in (a).
(c) The $P(u, w)$ map, a projection down the v axis, for the same structure, indicating the slope of the ring. The contour interval is arbitrary.
(d) One molecule shown for comparison with the Patterson map in (c).

Data from Glusker et al. (1968).

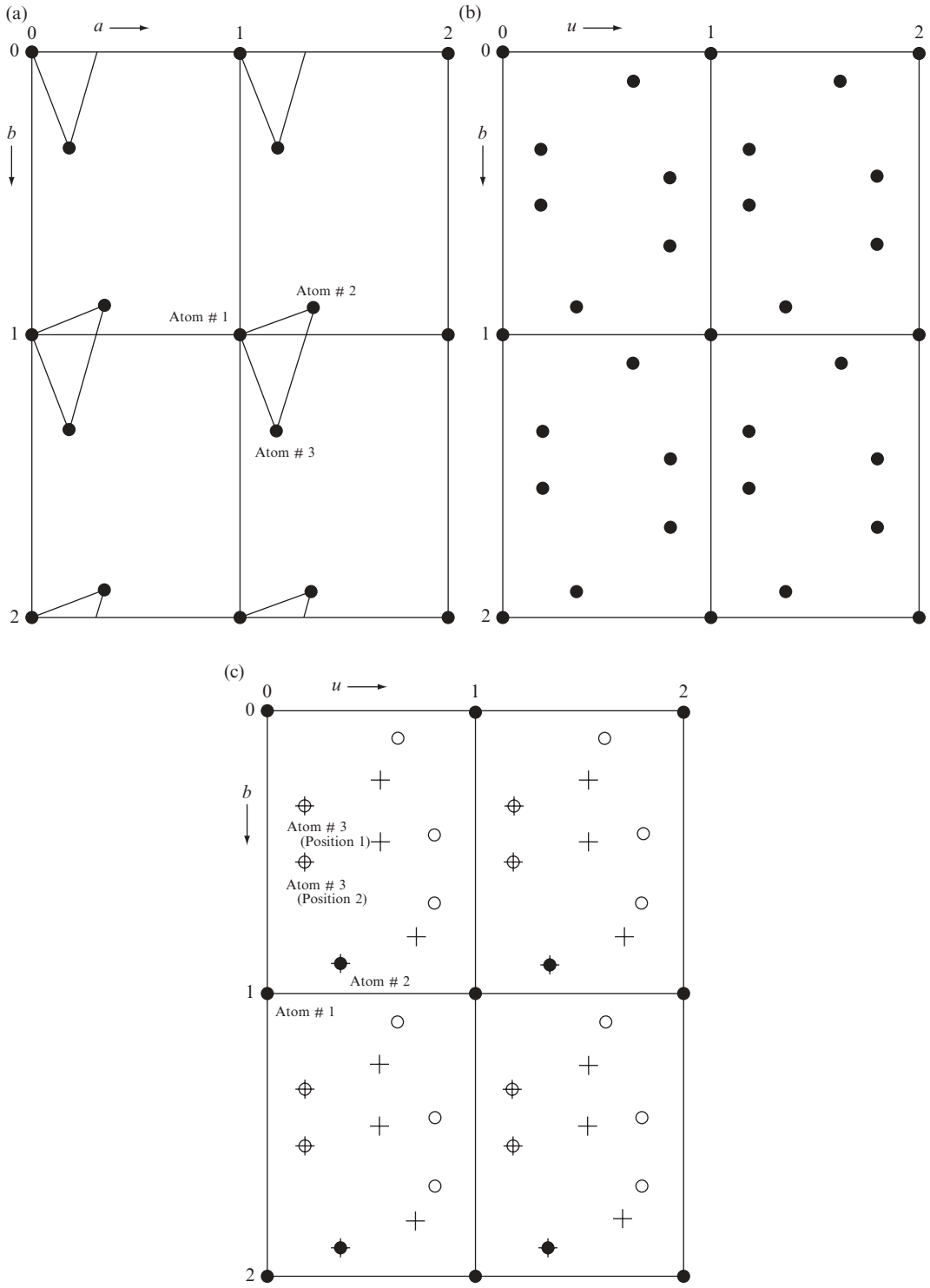


Fig. 9.4 The vector superposition method.

Patterson superposition methods

There are several methods, and many are quite powerful, for finding the structure corresponding to a Patterson map by transcribing $P(uvw)$ upon itself with different relative origins. One of the simplest methods for analyzing the Patterson map of a compound that contains an atom in a known position (such as a heavy atom that has been located in the Patterson map) is to calculate, graphically or by computer, a "vector superposition map." The origin of the Patterson map is put, in turn, always in the same orientation, at each of the symmetry-related positions of the known heavy atom, and the values of $P(uvw)$ are noted at all points in the unit cell. *The lowest value of* $P(uvw)$ *in the different superposed Patterson maps* is recorded for each point; the resulting vector superposition map is therefore also known as a *minimum function*. The principle underlying this approach is that it isolates the vectors arising from the interaction of the known heavy atom with all other atoms in the structure. A schematic example is illustrated in Figure 9.4. In some of the maps there will be other peaks at this same position, corresponding to other vectors in the structure, but the possible ambiguity that such peaks might introduce is minimized by recording the lowest value of $P(uvw)$ in any of the superposed maps. This method can be used even if no atomic positions are known, simply by moving each Patterson peak in turn onto the origin, as in the schematic example illustrated in Figure 9.4.

Rotation and translation functions

Sometimes a structure contains a complex molecule, with (necessarily) a multitude of vectors, but may include a group for which all the vectors are known (relative to one another) rather precisely—for example, a benzene ring in a phenyl derivative. The vector map of this grouping can then be calculated and the resulting vector arrangement can be compared with the arrangement of peaks around the origin of the Patterson map. There will be many more peaks in this region of the Patterson map than those arising from the known structural features alone, but, in at least one relative orientation of the two maps, all peaks in the vector map of the phenyl group will fall in positive areas

---

(a) Crystal structure.
(b) Patterson map of the structure shown in (a).
(c) Vector superposition. A search for the position of a third atom when the positions of the first two (#1 and #2, filled circles) are known. The Patterson map illustrated in (b) has been placed (i) with the origin on the position of atom #1 (to give open circles) and then, by superposition of peaks, (ii) with the origin on the position of atom #2 (to give crosses). Four unit cells are shown. It can be seen that there are four positions within each unit cell where overlap of Patterson peaks occurs (a circle and cross superposed). Two of these are, necessarily, at the positions of atom #1 (the origin) and atom #2; the other two are possible positions for atom #3; that is, there are two solutions to the vector map at this stage. In practice, this ambiguity is not found when many atoms are present, and the method will often show the structure clearly. Note that the two solutions to the structure problem are mirror images of each other.

of the Patterson map (although they will not necessarily all lie at maxima if the Patterson peaks are composite, as they usually are). This method, which involves rotation of the Patterson map, is illustrated in Figure 9.5. The fit of the calculated and observed Patterson maps can be optimized with a computer by making a "rotational search" to examine all possible orientations of one map with respect to the other and to assess the degree of overlap of vectors as a function of the angles through which the Patterson map has been rotated. The maximum overlap normally occurs (except for experimental errors) at or near the correct values of these rotation angles, thus giving the approximate orientation of the group. Then the Patterson map can be searched for vectors between groups in symmetry-related positions, and the exact position of the group in the unit cell can be found and used as part of a trial structure.

This method has been adapted to aid in the search for information on the relationship between molecules if there are more than one in the asymmetric unit. Sometimes, in crystalline proteins or other macromolecules, there is what is referred to as "noncrystallographic symmetry"; for example, a dimer of two identical subunits may be contained in one asymmetric unit. Thus, there are two identical structures with different positions and orientations within the asymmetric unit. If, however, one copy of the Patterson map of this dimer is rotated on another copy of the map, there will be an orientation of the first relative to the second that gives a high degree of peak overlap (McRee, 1993; Drenth, 2007; Sawaya, 2007). This is called a "self-rotation function." The results can be plotted in three dimensions in a map that describes the rotation angles that achieved superposition of the two maps. A large peak is expected at the rotation angles at which one subunit (or molecule) becomes aligned upon another. For example, the relative orientations of two subunits in the same asymmetric unit may be determined because the rotations required for superposition are directly related to the orientation of the noncrystallographic (local) symmetry element of the dimer, usually a two-fold axis (Rossmann and Blow, 1962). Thus a rotation function, plotted as a contour map, provides information on the results (as peaks) of the overlap of one Patterson function with the rotated version of another Patterson function. In a similar way, it may be possible to find the translational components of the noncrystallographic symmetry elements, but this is often considerably more difficult (Crowther and Blow, 1967).

This concept of a probe and the finding of its location in the unit cell by examination of the Patterson map is known as "molecular replacement," (Rossmann, 1972). For protein studies the probe may be structural data from a similar protein such as a mutant of the same protein. The method involves positioning the probe within the unit cell of the target crystal so that the calculated diffraction pattern matches that observed experimentally. The search is broken into two parts, as described above—rotation then translation, each providing three parameters. This method is very useful if atomic coordinates for

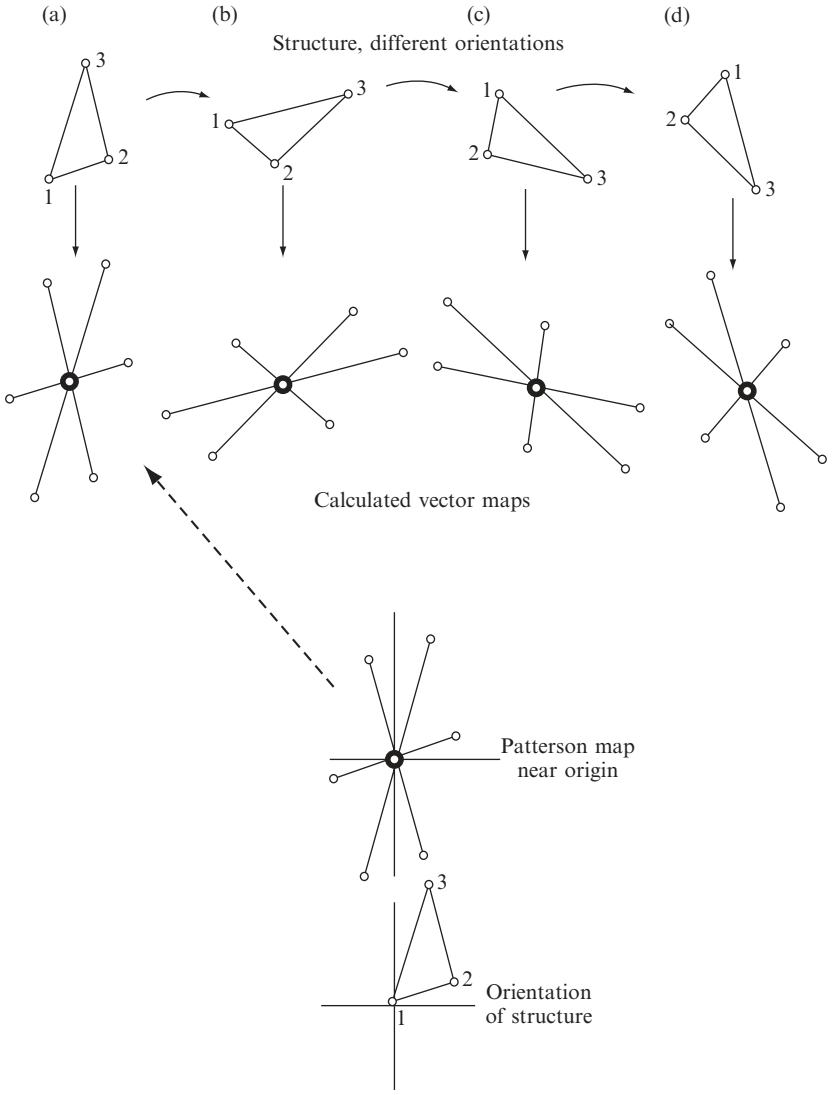


Fig. 9.5 A Patterson search by rotation.

This is a schematic example. If the dimensions of a molecule or part of a molecule in a crystal structure are known, but its orientation (and position) in the unit cell is unknown, the *orientation* may often be found by a comparison of calculated and observed vector maps around the origin. The *position* of the molecule will have to be found in some other way. A comparison of vector maps calculated from trial structures in various orientations with the Patterson map calculated from experimental data indicates that model (a) (above, left) has the trial structure in its correct orientation. The orientations in (b) to (d) are incorrect.

a similar structure, such as a similar protein from a different biological source, have already been reported.

The heavy-atom method

In the *heavy-atom method*, one or a few atoms in the structure have an atomic number Z_i considerably greater than those of the other atoms present. Figure 5.2c showed that if one atom has a much larger atomic scattering factor than the others, then the phase angle for the whole structure will seldom be far from that of the single heavy atom alone, unless, of course, many of the other atoms happen to be in phase with one another, a most improbable circumstance. Heavy atoms thus dominate the scattering of a structure, as illustrated in Figure 5.2c. If the molecule of interest does not contain such an atom, then a derivative, containing, for example, bromide or iodide, can often be prepared, with the hope that the molecular structural features of interest will not be modified in the process (Dauter et al., 2000). Heavy atoms can usually be located by analysis of a Patterson map, although this depends on how many are present and how heavy they are relative to the other atoms present. In Appendix 9 some data relevant to the Patterson function are given for an organic compound containing cobalt, a derivative of vitamin B_{12} with formula $C_{45}H_{57}O_{14}N_5CoCl \cdot C_3H_6O \cdot 3H_2O$; cobalt has an atomic number of 27 versus 6 for carbon, 7 for nitrogen, and 8 for oxygen. Therefore the scattering of cobalt, that is, $|F(hkl)|^2$, is 12–20 times greater than that of any of the three lighter atoms. The appearances of two Patterson projections for this substance are shown in Figure 9.6. In spite of the presence of many other peaks, the cobalt–cobalt peaks are heavier than most of those due to the other vectors and dominate the map. The position of the cobalt atom in the unit cell was thus found from these two Patterson projections ($P(uv)$ and $P(uw)$). In a similar way, the location of a heavy atom in a protein structure can be found. In Figure 9.7, the heavy-atom position in a protein crystal structure is found from the three Harker sections.

Once the heavy atom has been located, the assumption is then made that it dominates the diffraction pattern, and the relative phase angle for each diffracted beam for the whole structure is approximated by that for a structure containing just the heavy atom. Figure 9.8 illustrates the application of the heavy-atom method to the structure of the vitamin B_{12} derivative just mentioned, which contained one cobalt atom, one chlorine atom, and about seventy carbon, nitrogen, and oxygen atoms (the structure used for the Patterson map illustrated in Figure 9.6). The first approximation to the electron density was phased with the cobalt atom alone. Peaks in it near the metal atom that were most compatible with known features of molecular geometry were used, together with the metal atom, in a calculation of phase angles for a second approximate electron-density map. This process was continued until the entire structure had been found. The combined use of

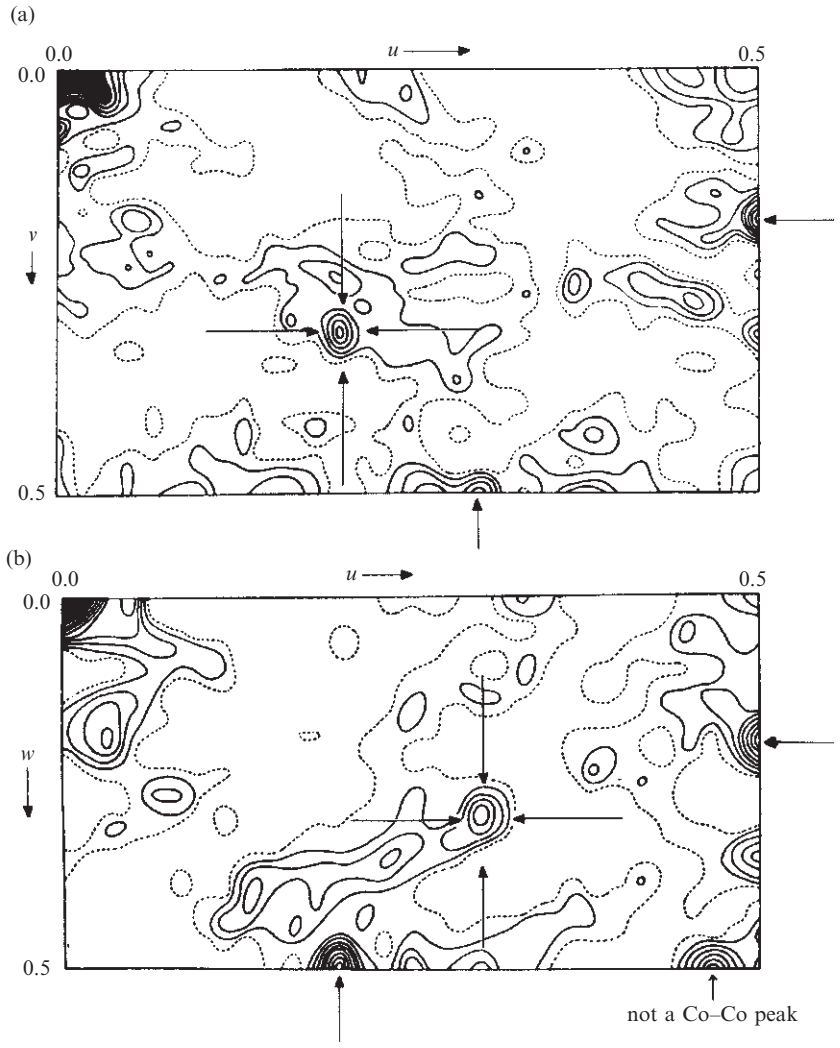


Fig. 9.6 Patterson projections for a cobalt compound in the space group $P2_12_12_1$.

Peaks identified as arising from cobalt–cobalt interactions are indicated by arrows. See Appendix 9 for an analysis of these maps.

(a) $P(u, v)$ Patterson projection down the c-axis. Co–Co peaks appear at 0.00, 0.00; 0.20, 0.32; 0.50, 0.18; 0.30, 0.50.
(b) $P(u, w)$ Patterson projection down the b-axis. Co–Co peaks appear at 0.00, 0.00; 0.30, 0.30; 0.50, 0.20; 0.20, 0.50.

Note that these particular Patterson maps are projections, not Harker sections, but that Harker peaks at half each unit-cell direction (u and v = 0.50 in (a) and u and w = 0.50 in (b)) helped solve the location of the cobalt atom.

From *Proceedings of the Royal Society* (Hodgkin et al. (1959), p. 312, Figure 3). Published with permission.

Patterson maps and heavy-atom methods made it possible for structures of moderate complexity to be solved in the 1950s and 1960s and, for a while, was the most powerful tool in the analysis of structures of moderate complexity (molecules with, say, 30 to 100 atoms). Direct methods are now more commonly used to solve such structures (small and moderate-sized), because these methods have become much more

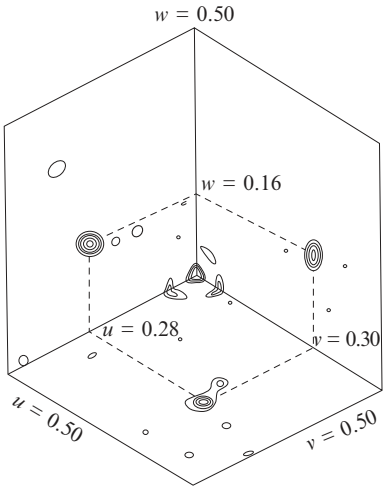


Fig. 9.7 The heavy-atom method. A difference Patterson map.

The macromolecule crystallizes in the space group $I222$. Atomic positions are (0, 0, 0 or $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$) + (x, y, z; $-x$, $-y$, z; x, $-y$, $-z$; $-x$, y, $-z$). Three Harker sections have peaks at $u = 2x$, $v = 2y$, $w = 0$, at $u = 2x$, $v = 0$, $w = 2z$, and at $u = 0$, $v = 2y$, $w = 2z$. The heavy atom is therefore found to lie at $x = 0.14$, y = 0.35, and z = 0.42.

powerful with the greatly increased availability of high-speed, high-capacity computers. One minor drawback of the heavy-atom method is that when the heavy atom has an atomic number sufficiently high to dominate the vector distribution, it will necessarily also contribute strongly to the X-ray scattering. If it is desirable to know the structure very precisely, it may be better to work on the structure of a compound that does not contain a heavy atom as a derivative. However, now, with precise low-temperature measurements and high-resolution data, it is generally possible to locate hydrogen atoms in small structures, even if a very heavy atom, such as tungsten or mercury, is present. In addition, Patterson maps can permit a search for vectors of a specific length, such as the S–S distance of a disulfide bridge or the vector between two metal ions that share a particular functional group.

The isomorphous replacement method

Isomorphous crystals are similar in shape, unit-cell dimensions, and structure. They have similar (but not identical) chemical compositions (for example, when one atom has a different atomic number in the two structures) (Mitscherlich, 1822). Ideally, the substances are so closely similar that they can generally form a continuous series of solid solutions, so that, for example, a colorless crystal of potash alum will continue crystal growth on a crystal of chrome alum. When the term "isomorphous" is used for a crystal of a biological macromolecule, it implies that the crystal, with and without a heavy-atom compound soaked into the water channels of the protein or else genetically engineered into the structure, has the same unit-cell dimensions and space group. As a result it is assumed that the macromolecules are in the same positions and orientations in the two crystals.

The high scattering power of heavy atoms has been used to help solve the structures of biological macromolecules. The isomorphous replacement method that will be discussed next has been used in large number of protein structure determinations. The Patterson map of a protein is too complex, with too many overlaps of peaks, for direct interpretation, but the location of a heavy atom, if it can be introduced into a protein crystal, can be found. Data for both the protein and its "heavy-atom derivative" are used to determine perturbations to intensities caused by the addition of heavy atoms. With multiple isomorphous replacement, the aim is to make some alteration in the crystal and examine how this change perturbs the structure factors. From the measured intensities, plus the changes on the introduction of different heavy atoms, it may be possible to obtain phases for each Bragg reflection (Bokhoven et al., 1949; Harker, 1956). For example, if a protein has a molecular weight of 24,000, it contains approximately 2000 carbon, nitrogen, and oxygen atoms. Then, at $\sin\theta = 0^\circ$, the mean value of

$$<|F_P|^2> = \sum_{j=1}^{2000} f_c^2$$

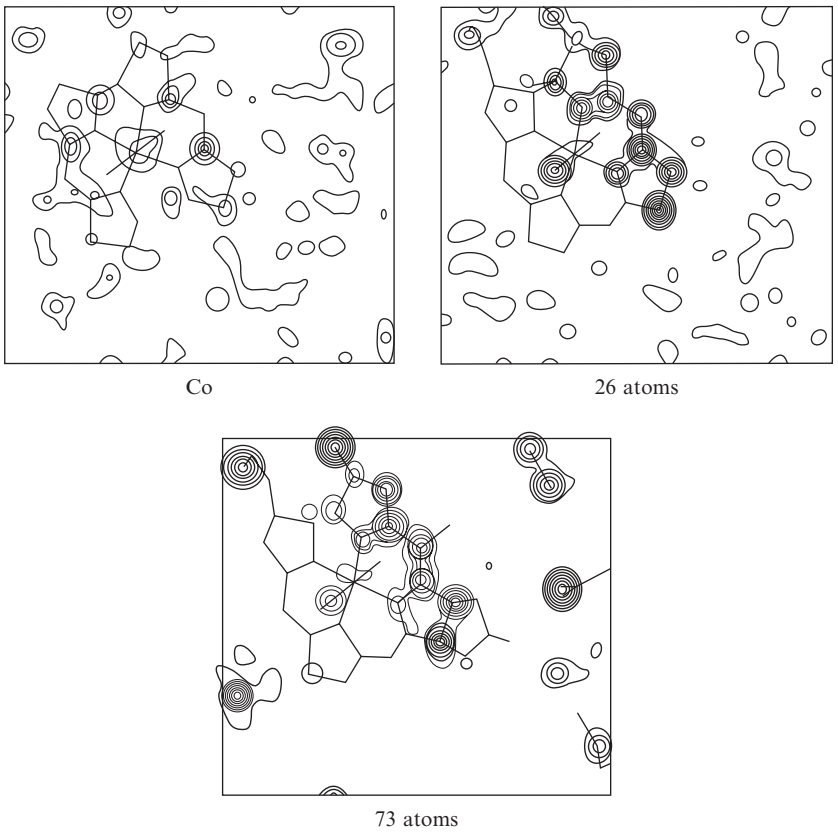


Fig. 9.8 The heavy-atom method.

One section through a three-dimensional electron-density map for a structure with 73 atoms (including various solvent molecules, but not hydrogen atoms) in the asymmetric unit is shown at three different stages of the structure analysis. In the calculation of the first map, only a cobalt atom was used to determine phases. For the second map, 26 atoms were used (one Co, 25 C and N), chosen from peaks in the first map. The third map was phased with the positions of all 73 atoms. Most of the features of the map phased with 73 atoms can be found, at least weakly, in the map phased with the heavy atom alone, although in the latter map there are many extra peaks that do not correspond to any real atoms. Note the general reduction in the background density as the correct relative phase angles are approached. Since these are sections of a three-dimensional map, some atoms that lie near but not in the plane of the section are indicated by lower peaks than would represent them if the section passed through their centers. Other atoms implied by the skeletal formula lie so far from this section that no peaks corresponding to them occur here.

From *Proceedings of the Royal Society* (Hodgkin et al. (1959), p. 328, Figure 14). Published with permission.

is about 98,000. If one uranium atom, atomic number 92, is added, this value of $<|F_P|^2>$ is increased to approximately 106,000, an 8 percent change in average intensity. Differences in intensities of the native protein and the heavy-atom derivative can be measured, many of which will significantly exceed the average. Therefore the position of the uranium atom should be obtainable from the intensity differences.

If a protein crystal is soaked in a solution of a heavy-atom compound (such as a uranyl salt), the heavy-atom compound will be distributed throughout the solvent channels in the crystal by diffusion. In some cases the heavy atom will bind to a specific group on the macromole-

cule, and this binding may occur in an ordered arrangement within the macromolecular crystal. The difference between the diffracted intensities of the "heavy-atom derivative" of the protein crystal (structure factors F_{PH}) and the diffracted intensities of the "native protein" without any added heavy atom (structure factors F_P) can be used to reveal the position of the heavy atom by a "difference Patterson map." This is done with a Patterson map that uses $||F_{PH}|^2 - |F_P|^2|$ as coefficients. In other cases, however, with this method of soaking heavy atoms into the native protein, nonspecific binding of the heavy atom may occur, since there are many binding groups on the surface of a protein. If this does happen, that particular heavy-atom derivative can probably not be used for structure determination because of the disorder of its position. To prevent random binding it is necessary to stop the soaking after an appropriate time, determined experimentally, in the hope that only specific binding will occur; the concentration of the heavy-atom salt is often critical for this. An alternative method, which involves attempting to crystallize proteins from solutions containing heavy-atom salts, has not proved very satisfactory, because the crystals so obtained are often not isomorphous with the native crystal. Crystals must be isomorphous for the use of the isomorphous replacement method that will now be described. When a pair or series of isomorphous crystals can be found, isomorphous replacement is a powerful method for the determination of *phase angles*, especially for complex structures for which purely analytical methods (see Chapter 8) are inadequate. It has provided the basis for the solution of many of the macromolecular crystal structures determined to date.

Isomorphous crystals are crystals with essentially identical cell dimensions and atomic arrangements but with a variation in the nature of one or more of the atoms present. The alums constitute probably the best-known example of a series of isomorphous crystals. "Potash alum," $KAl(SO_4)_2 \cdot 12H_2O$, grows as colorless octahedra, while "chrome alum," $KCr(SO_4)_2 \cdot 12H_2O$, forms dark lavender crystals of the same shape and structure. The Cr(III) atom in chrome alum is in the same position in the unit cell as the Al(III) atom in potash alum. A common experiment in isomorphism is to grow a crystal of chrome alum suspended from a thread and then to continue to grow it in a solution of potash alum. The result is an octahedral crystal with a dark center surrounded by colorless material (Holden and Singer, 1960). In general, however, isomorphous pairs (involving isomorphous replacement of one atom by another) are difficult to find for crystals with small unit cells, because variations of atomic size usually cause significant structural changes when substitution is tried. Even with large unit cells, patience and ingenuity are usually needed to find an isomorphous pair for a compound being studied. The rewards from this method are enormous—the entire three-dimensional molecular architecture of a protein molecule, found with minimal chemical assumptions.§ Max Perutz searched for years for ways to solve the structure of hemoglobin, and succeeded when he devised this isomorphous replacement method

§ Usually information about the sequence of the amino acids in the protein chain is needed to interpret the electron-density maps, especially in poorly resolved regions of the structure.

(Green et al., 1954). The existence of isomorphism between a protein and a heavy-atom derivative may be demonstrated by the determination that their unit-cell dimensions do not differ by more than about 0.5 percent, and that there are differences in the diffraction intensity patterns. It is hoped that there is only a change in the site of isomorphous replacement and that most of the crystal structure of the native and the heavy-atom-substituted protein is the same.

After a Patterson map has been calculated and the position of the heavy atom has been found for each derivative, it is possible to calculate the relative phases of the Bragg reflections directly by a proper consideration of the changes in intensity from one crystal to another.

The method for calculating phase angles by the isomorphous replacement method is illustrated in Appendix 8 in a numerical example involving a centrosymmetric crystal. The atoms or groups of atoms (M and M′) that are interchanged during preparation of the isomorphous pair must be located, usually from a Patterson map, as described earlier. This allows calculation of their contributions, F_{M} and $F_{\mathrm{M}'}$. If F_{M} and $F_{\mathrm{M}'}$ are positive (they necessarily have the same sign, since their only difference is in the amount of scattering power in the atom or group of atoms), then the overall F values (F_{T}) must differ in the same way that F_{M} and $F_{\mathrm{M}'}$ differ. Since the absolute magnitudes of these measured values of F are known and the difference equivalent to the change in M can be computed, it is possible to find signs for F_{T} and $F_{\mathrm{T}'}$. The solutions to the equations are in practice inexact, because of experimental errors and also because the remainder of the structure, R, may move slightly during the replacement of one ion group by another. In an interesting variation to the isomorphous replacement method, it has been found that the relationship between structures in a crystal before and after radiation damage can be used to determine phases. For the "radiation-damage-induced phasing" (RIP) method, two data sets of the same crystal are measured. Between the two data collections the crystal is exposed to a very, very large dose of X-rays. The structural changes as a result of radiation damage (by analogy with heavy-atom insertion) are enough to make it possible to determine the phasing, especially if a few somewhat heavier atoms, such as sulfur, are present in the structure (Ravelli et al., 2003).

With noncentrosymmetric structures, the situation is greatly complicated by the fact that the phase angle may have any value from 0 to 360°. This is the case for biological macromolecules. If the heavy-atom position can be found from the Patterson map, then F_H and the relative phase angle α_H can be computed for a given diffracted beam for each derivative. The construction for graphical determination of the phase angle for the protein (P) is shown in Figure 9.9. For each heavy-atom derivative (PH_1 and PH_2), two possible values[¶] for the phase angle for the protein are found; in the example in Figure 9.9, these are near 53° and 322° for PH_1 (Figure 9.9a) and near 56° and 155° for PH_2 (Figure 9.9b). The phase angle for the free protein for this particular diffracted beam must therefore be near 54°. This process of estimating the

[¶] With reference to Figure 9.9a, the law of cosines also illustrates the two-fold ambiguity in the phase angle determined from just one heavy-atom derivative:

$$\begin{aligned}\alpha_P &= \alpha_H \\ &\quad + \cos^{-1}\{(F_{PH}{}^2 - F_P{}^2 - F_H{}^2)/2F_P F_H\} \\ &= \alpha_H \pm \alpha'\end{aligned}$$

Thus two values of α_P, that is, $\alpha_H + \alpha'$ and $\alpha_H - \alpha'$, are possible and it is necessary to study several heavy-atom derivatives with substituted heavy atoms in different positions in the unit cell; the value of α_P that is common to these different studies is determined in this way.

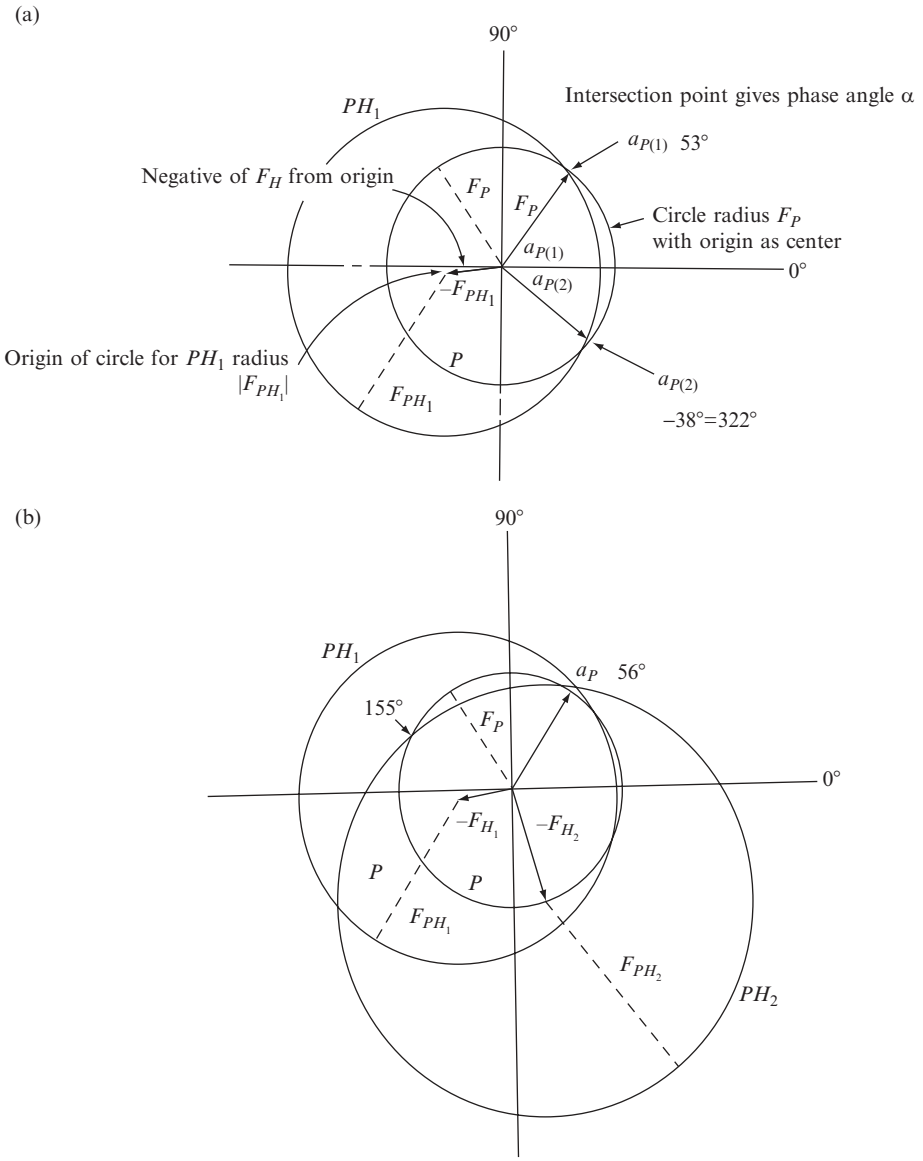


Fig. 9.9 Isomorphous replacement for a noncentrosymmetric structure.

Graphical evaluation of the relative phase, α_P, of a Bragg reflection, indices hkl, diffracted with a structure factor $|F_P|$ from a protein crystal. The diagrams illustrate the following equation: $\mathbf{F}_P = \mathbf{F}_{PH} - \mathbf{F}_H$, where P = "native protein," H = heavy atom, PH = protein heavy-atom derivative.

(a) One heavy-atom derivative is available, with a structure factor $|F_{PH1}|$ for the Bragg reflection hkl. A circle with radius $|F_P|$ is drawn about the origin. From the position of the heavy atom, determined from a difference Patterson map, it is possible to calculate both the structure amplitude and the phase of the heavy-atom contribution ($|F_{H1}|$, phase angle α_{H1}). A line of length $|F_{H1}|$ and phase angle $-\alpha_{H1}$ (i.e., $\alpha_{H1} + 180°$ to give $-\mathbf{F}_{H1}$) is drawn. With the end of this vector as center, a circle with radius $|F_{PH1}|$ is drawn. It intersects the circle with radius $|F_P|$ at two points, corresponding to two possible phase angles, $\alpha_{P(1)}$ and $\alpha_{P(2)}$, for the native protein.

(b) When two or more heavy-atom derivatives are available, then the process described in (a) is repeated and, in favorable circumstances, only one value of phase angle for the native protein is obtained. Thus, a second derivative is needed to remove the two-fold ambiguity of case (a). This method, of course, depends on accurate measurements of $|F_P|$, $|F_{PH1}|$, and $|F_{PH2}|$ and involves the assumption that no other perturbation than the addition of a heavy atom to the native protein has occurred. Additional derivatives are sometimes needed to improve the accuracy of the phase angles.

phase angle must be repeated for each diffracted beam in the diffraction pattern; usually more than two heavy-atom derivatives (in addition to the free protein) are studied, so that the phases can be more accurately determined. A measure of the error in phasing is provided by a figure of merit, m. This is the mean cosine of the error in the phase angle; it is near unity if the circles used in deriving phases (Figure 9.9) intersect in approximately the same positions. For example, if the figure of merit is 0.8, the phases are in error, on the average, by $\pm 40^\circ$, if it is 0.9, the mean error is $\pm 26^\circ$.

Thus the stages in the determination of the structure of a protein involve the crystallization of the protein, the preparation of heavy-atom derivatives, the measurement of the diffraction patterns of the native protein and its heavy-atom derivatives, the determination of the heavy-atom positions, the computation of phase angles (Figure 9.9), and the computation of an electron-density map using native-protein data and the phase angles so derived from isomorphous replacement. The map is then interpreted in terms of the known geometry of polypeptide chains so that initially this backbone of the protein is traced through the electron-density map. This was formerly done by model building (using a half-silvered mirror in a "Richards box" so that a ball-and-stick model and the electron-density map were superimposed and therefore could be visually compared). Nowadays it is more common for this interpretation of the electron-density map to be done with the help of computer graphics (as shown in Figure 9.10).

The isomorphous replacement method was one of the most used methods for determining macromolecular structure, is now being replaced by experimental methods that involve anomalous dispersion ("MAD" and "SAD" phasing). They will be described in Chapter 10, where their advantages will be described.

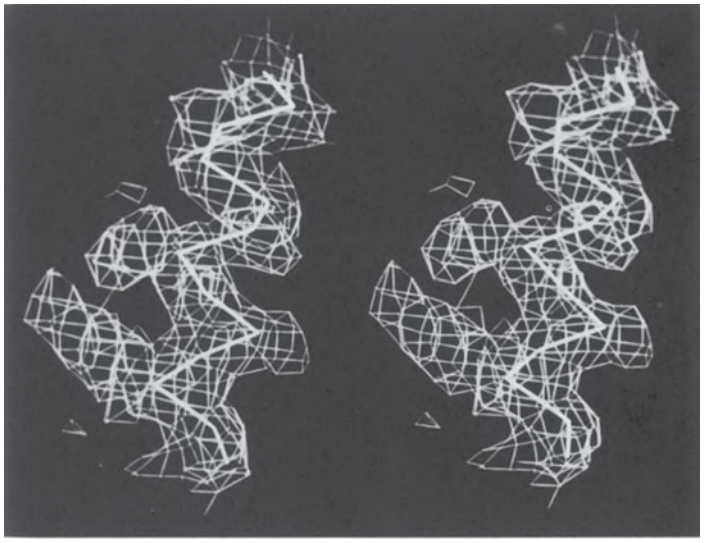


Fig. 9.10 Protein backbone fitting by computer-based interactive graphics.*

*Areas of high electron density are stored in the computer as three-dimensional information and are represented by cage-like structures on a video screen. Any desired view can be generated. The backbone of the molecule is represented as a series of vectors, each 3.8 Å in length (the distance between α-carbon atoms in a polypeptide chain). Each vector is positioned with one end on an α-carbon atom; the other end of the vector is rotated until it lies in an appropriate area of high electron density. Then coordinates of both ends of the vector are stored in the computer, the process is repeated, and the most likely location of the next α-carbon atom is sought. Such vectors are represented in this figure, a stereopair** photographed from a video screen, as heavy solid lines. In this way the "backbone," that is, the positions of the carbon atoms of the polypeptide backbone of the protein (excluding side chains), may be found.

(Photograph courtesy Dr. H. L. Carrell.)

** Such stereodiagrams can be viewed with stereoglasses, or readers can focus on the two images until an image between them begins to form, and then allow their eyes to relax until the central image becomes three-dimensional. This process requires practice and usually takes 10 seconds or more.

Multiple Bragg reflection

A very different but important approach to phase measurement involves "double reflection." This is a physical effect that occurs when the crystal is oriented so that two reciprocal lattice points, h_1, k_1, l_1 and h_2, k_2, l_2, lie simultaneously in the diffracting position; that is, both lie on the surface of the sphere of reflection (the Ewald sphere) at the same instant. The result is two beams that are diffracted in the direction normally expected for h_1, k_1, l_1, and which interfere with each other. One is the normally expected $h_1k_1l_1$ Bragg reflection (the primary beam), and the other, also in the $h_1k_1l_1$ direction, results from the $h_2k_2l_2$ Bragg reflection (the secondary beam) acting as the incident beam for the $h_1 - h_2$, $k_1 - k_2$, $l_1 - l_2$ Bragg reflection (the coupling beam). The amplitude of the resultant Bragg reflection gives information on the phase difference of these two waves. This effect is variously described as the Renninger effect (Renninger, 1937), the *Umweganregung* effect (if $I(h_1k_1l_1)$ is increased at the expense of $I(h_2k_2l_2)$), or the *Aufhellung* effect (if $I(h_1k_1l_1)$ is decreased). When a ψ-scan of the peak (through a very small angle) is done it is found that there is an asymmetry, shown in Figure 9.11, that depends on the value of the phase invariant. Therefore a direct reading of α_{sum} is obtained:

$$\alpha_{\text{sum}} = \underset{\text{primary beam}}{\alpha(-h_1, -k_1, -l_1)} + \underset{\text{secondary beam}}{\alpha(h_2, k_2, l_2)} + \underset{\text{coupling beam}}{\alpha(h_1 - h_2, k_1 - k_2, l_1 - l_2)} \quad (9.4)$$

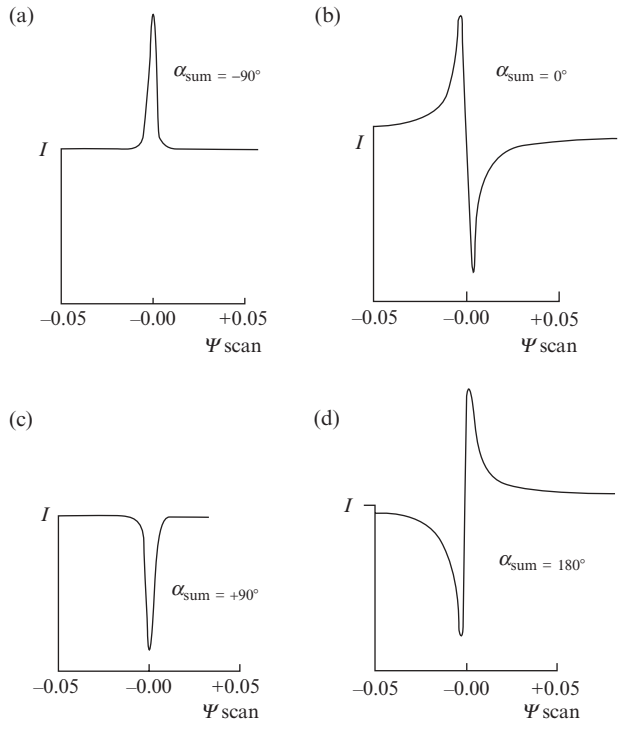


Fig. 9.11 Multiple Bragg reflections.

Some idealized ψ-scans and values of $\alpha_{\text{sum}}(hkl)$ derived from peak profiles. If the crystal is noncentrosymmetric, intermediate phase sum values will be found.

This is like a triplet phase relationship but is an equality rather than a probability. A highly precise (six-circle) diffractometer is needed for this experiment but, when available, such an instrument has provided experimental data that have been used with good success (Hümmer and Billy, 1986; Shen, 1998). Thus an experimental way of measuring origin-independent structure invariants is provided.

Summary

The Patterson map

The map computed with amplitudes $|F(hkl)|^2$, but no phase information, will give a vectorial representation of the atomic contents of the unit cell. The Patterson function, $P(uvw)$, is expressed in the coordinate system u, v, w in a cell of the same size and shape as that of the crystal. It is calculated by

$$P(uvw) = \frac{1}{V} \sum_{\text{all } h,k,l} \sum \sum \sum \left|F(hkl)\right|^2 \cos 2\pi(hu + kv + lw)$$

The peaks in this map occur at points whose distances from the origin correspond in magnitude and direction with distances between atoms in the crystal, because

$$P(uvw) = V \int\int\int \rho(x, y, z)\, \rho(x + u, y + v, z + w)\, dx\, dy\, dz$$

Ideally this map can be interpreted in terms of an atomic arrangement. In practice, however, this is only possible if there are comparatively few atoms in the structure or if some are very heavy. The map may also be "sharpened" and the high origin peak removes if values of $\{|E(hkl)^2| - 1\}$ rather than $|F(hkl)|^2$ are used. A rotation of a Patterson map can be used to identify the angle between two identical molecules when noncrystallographic symmetry is present.

The heavy-atom method

If one or a few atoms of high atomic number are present, they will dominate the scattering. These atoms can generally be located from a Patterson map and the phases of the entire structure approximated by the phases of the heavy atom(s). In the resulting electron-density map, portions or all of the remainder of the structure will usually be revealed, leading to improved phases and successively better approximations to the structure.

Isomorphous replacement method

For very large structures, such as those of proteins, the isomorphous replacement method is a good method for the experimental determination of phase angles. Two crystals are isomorphous if their space

groups are the same and their unit cells and atomic arrangements are essentially identical. Since protein crystals contain solvent channels, if heavy atoms (in solution) are soaked into them and the resulting crystals are isomorphous with the unsubstituted ("native") crystal, a comparison of the two diffraction patterns will give relative-phase information. If the positions of these added or replaced atoms can be found from Patterson maps, their contributions to the phase angle of each Bragg reflection can be calculated, and if the atoms are sufficiently heavy, differences in intensities for the two isomorphs can be used to determine the approximate phase angle for each Bragg reflection. At least two heavy-atom derivatives are necessary for noncentrosymmetric structures.

Multiple Bragg reflection

If the crystal is oriented so that two reciprocal lattice points lie simultaneously in the diffracting position, that is, both lie on the surface of the sphere of reflection at the same instant, the resulting diffracted beam contains information on a structure invariant involving three Bragg reflections—the primary beam, the secondary beam, and the coupling beam. This method requires specialized equipment.

Anomalous scattering and absolute configuration

10

The concept of the carbon atom with four bonds extending in a tetrahedral fashion was put forward by van't Hoff and Le Bel in 1874. It coincided with the realization that such an arrangement could be asymmetric if the four substituents were different, as shown in Figure 10.1a (van't Hoff, 1874; Le Bel, 1874). Thus, for any compound containing one such asymmetric carbon atom, there are two isomers of opposite chirality (individually called enantiomers), for which three-dimensional representations of their structural formulas are related by a mirror plane. Aqueous solutions of these enantiomers rotate the plane of polarized light in opposite directions. As discussed in Chapter 7, Pasteur showed that crystals of sodium ammonium tartrate had small asymmetrically located faces and that crystals with these so-called "hemihedral faces" rotated the plane of polarization of light clockwise, while crystals with similar faces in mirror-image positions rotated this plane of polarization counterclockwise. Thus the external form (that is, the morphology) of the crystals illustrated in Figure 10.1b was used to separate enantiomers (see Patterson and Buchanan, 1945). Pure enantiomers can only crystallize in noncentrosymmetric space groups unless both isomers are present.

But even if the chemical formula and the three-dimensional structure of a molecule such as tartaric acid have been determined by standard X-ray diffraction methods, there is an ambiguity about the absolute configuration. Information about the absolute configuration is not contained in the diffraction pattern of the crystal as it is normally measured. Thus, although the substituents on the asymmetric carbon atoms have been identified, and even the detailed three-dimensional geometry of the molecule has been determined, it is not known which of the two enantiomers (mirror-image forms, analogous to those shown in Figure 10.1a) represents the three-dimensional structure of a particular individual molecule that has some distinguishing chiral property, such as the ability to rotate the plane of polarized light to the right. In other words, what is the absolute structure of the dextrorotatory form of the compound under study?

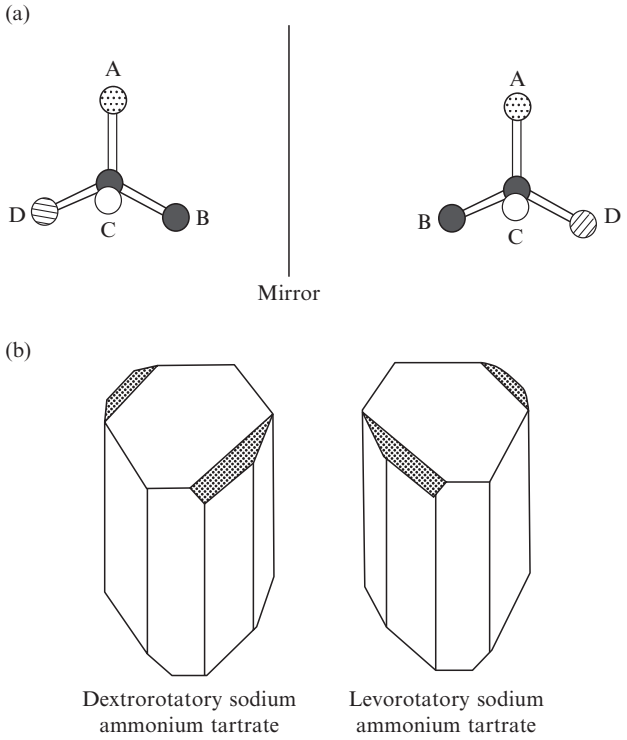


Fig. 10.1 Absolute configurations.

(a) The asymmetric carbon atom. If A, B, C, and D attached to the tetrahedral carbon atom are all different, there are two chiral isomers related to each other by a mirror plane. In a similar way, the entire structure of a crystal may be chiral.
(b) Hemihedral faces (shaded) on sodium ammonium tartrate crystals (used by Pasteur to differentiate dextrorotatory from levorotatory forms).

A means of determining the absolute configurations of molecules was, however, provided by X-ray crystallographic studies. It was made possible by the observation that the absorption coefficient of an atom for X-rays shows discontinuities when plotted as a function of the wavelength of the incident X-radiation. These discontinuities, shown in Figure 10.2, are graphically described as "absorption edges." At wavelengths at the absorption edge of an atom, the energy (inversely proportional to wavelength) of the incident X rays is sufficient either to excite an electron in the strongly absorbing atom to a higher quantum state or to eject the electron completely from the atom. This has an effect on the phase change of the X rays on scattering. The scattering factor for the atom becomes "complex," and the factor f is replaced by

$$f = f_i + f_i' + \mathrm{i} f_i'' \tag{10.1}$$

where f' and f'' vary with the wavelength of the incident radiation. While f' causes no change in phase (it remains at 180°), f'' causes a phase change of 90°, which is the reason that the Friedel-pair symmetry breaks down. The value of f'' is largest when the wavelength is near

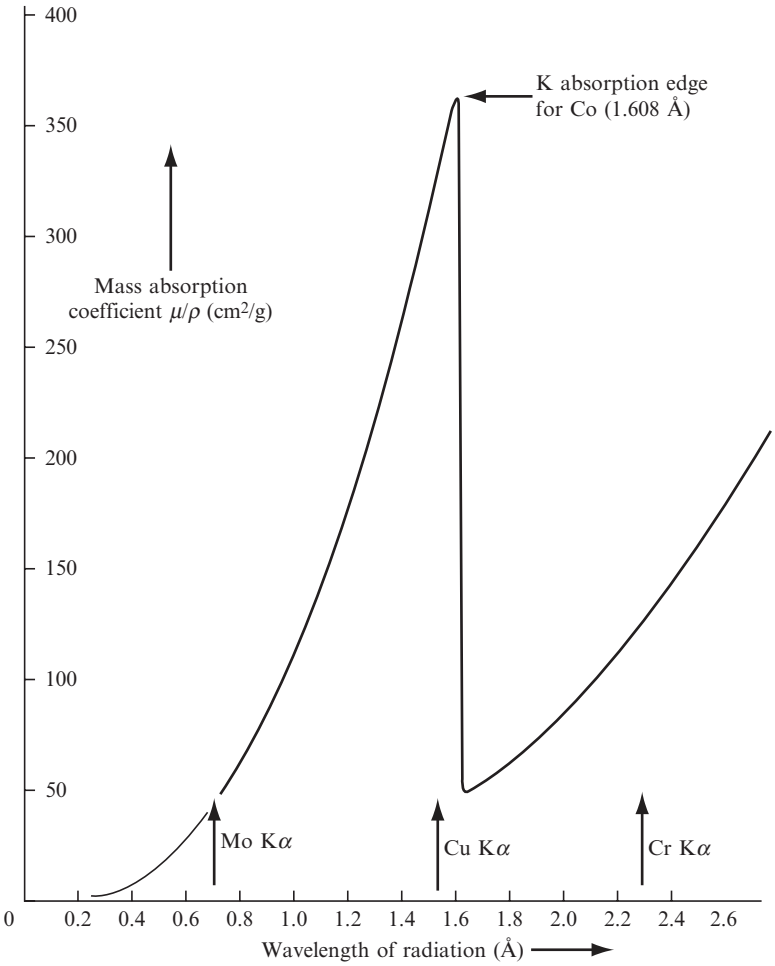


Fig. 10.2 Absorption of X rays of various wavelengths by a cobalt atom.

The mass absorption coefficient for cobalt as a function of wavelength. Note the discontinuity near the absorption edge (1.608 Å); beyond it, there is a gradual increase in the coefficient as the wavelength of the radiation increases.

the absorption edge (as for cobalt in a structure studied with copper Kα radiation; see Figure 10.2).

When visible light passes through transparent matter, such as a glass prism or a colorless crystal, its speed is decreased from the value it had in a vacuum. This decreased speed depends on the wavelength of the light. The refractive index of a material is the ratio of the velocity of light *in vacuo* to its velocity when it passes through this medium. Since white light consists of rays with a variety of wavelengths (from red to violet), rays with different wavelengths will be refracted at slightly different angles when they enter a material at an angle. This separation of light so that the individual colors of the component waves become visible is called "dispersion." The violet and blue rays (of shorter wavelengths) are slowed more and therefore are bent to a greater extent

than are the red and orange rays, with longer wavelengths; rays with shorter wavelengths have a larger refractive index. The result of such dispersion is a beautiful rainbow-like display of colors, such as that seen when sunlight passes through and exits a glass prism. This dispersion becomes "anomalous" when an energy absorption band is encountered and a discontinuity occurs; the dispersion is normal on either side of the absorption band, but at the absorption edge the refractive index is larger for longer wavelengths, rather than shorter as is normal. In this area near an absorption edge this plot of wavelength versus refractive index shows an increase of refractive index with wavelength (so that blue light is less refracted than red), the opposite of normal expectation. This is called "anomalous dispersion" or "anomalous scattering," meaning that one is studying the area of a spectrum near an absorption edge.

All atoms scatter anomalously to some extent, but at wavelengths near the absorption edge of a scattering atom, anomalous scattering will be especially noticeable. If an atom in the structure absorbs, at least moderately, the X rays being used, then this absorption will result in a *phase change for the X rays scattered by that atom, relative to the phase of the X rays scattered by the other atoms* of the structure, the equivalent of advancing or delaying the radiation for a short time as shown in Figure 10.3 [that is, equivalent to a hesitation ("gulp") at the time of scattering].

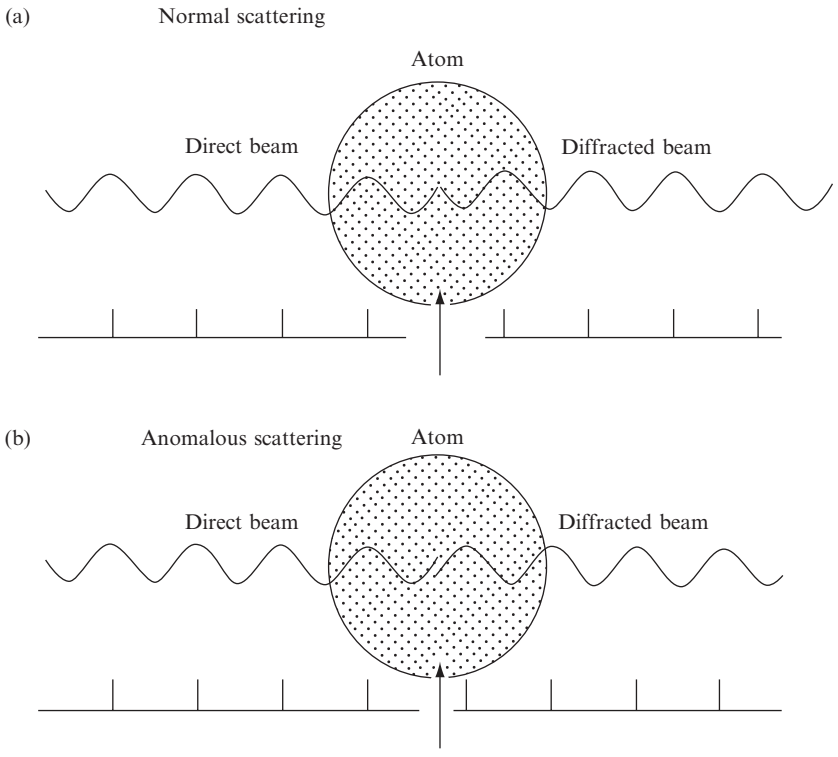


Fig. 10.3 Phase change on anomalous scattering.
(a) Normal scattering with a phase change of 180°. (b) Anomalous scattering with a different phase change.

This implies that in order to demonstrate anomalous scattering, the crystal must contain at least two different types of atoms. The phase change caused by f'' changes the path length of the scattered radiation, as illustrated schematically in Figure 10.3, and the result is an effect on the intensities of the diffracted beams. When there is none of this so-called "anomalous scattering," the intensities of the Bragg reflections with indices h, k, l and $\bar{h}, \bar{k}, \bar{l}$ are the same (Friedel's Law). When there is anomalous scattering, the intensities of these two Bragg reflections may be different because of changes in effective path differences between scattered waves arising from the phase change on absorption by the anomalously scattering atom.

The difference in intensities may alternatively be thought of as a result of the complex nature of the scattering factor, f_i [see Eqn. (10.1)], so that the absolute value of $F(hkl)$ is different from that of $F(-h, -k, -l)$, as illustrated in Figure 10.4. We showed in Eqn. (10.1) that if there is an anomalous scatterer in the crystal, f is replaced by $f + f' + \mathrm{i}f''$. Let $A' = G(f + f') + A$ and $B' = H(f + f') + B$, where A and B refer to the rest of the structure and G and H to the anomalous scatterer. Remember that f and f' scatter with a phase change of 180°, while f'' scatters with a phase change of 90°. As a result, since

$$F(hkl) = (A' + \mathrm{i}Gf'') + \mathrm{i}(B' + \mathrm{i}Hf'') = (A' - Hf'') + \mathrm{i}(B' + Gf'') \quad (10.2)$$

$$|F(hkl)|^2 = (A' - Hf'')^2 + (B' + Gf'')^2 \quad (10.3)$$

and similarly

$$F(\bar{h}\bar{k}\bar{l}) = (A' + \mathrm{i}Gf'') - \mathrm{i}(B' + \mathrm{i}Hf'') = (A' + Hf'') - \mathrm{i}(B' - Gf'') \quad (10.4)$$

$$|F(\bar{h}\bar{k}\bar{l})|^2 = (A' + Hf'')^2 + (B' - Gf'')^2 \quad (10.5)$$

it then follows that

$$|F(hkl)|^2 - |F(\bar{h}\bar{k}\bar{l})|^2 = 4f''(B'G - A'H) \quad (10.6)$$

Thus, when the incident X-radiation is of a wavelength near that of the absorption edge of an atom in a noncentrasymmetric structure, $|F(hkl)|$ does not equal $|F(\bar{h}\bar{k}\bar{l})|$. Under normal conditions the wavelength of the X-radiation used for a diffraction experiment is far from any absorption edge and these two quantities, $|F(hkl)|$ and $|F(\bar{h}\bar{k}\bar{l})|$, are equal. If anomalous scattering occurs, the magnitude of the difference between $|F(hkl)|^2$ and $|F(\bar{h}\bar{k}\bar{l})|^2$ for the two Bragg reflections (called "Friedel pairs" or "Bijvoet pairs") is a function of f'' (which depends on how near the incident radiation is to the absorption edge) and the positional parameters of both the anomalous scatterer and the rest of the structure.

It is possible from Eqn. (10.6) to calculate the expected differences between $|F(hkl)|^2$ and $|F(\bar{h}\bar{k}\bar{l})|^2$ for a given enantiomorph (with a specific absolute configuration). In practice, the indices of the Bragg reflections are assigned so that h, k, and l are in a right-handed system. Therefore the axes of x, y, z (that is, **a**, **b**, and **c**) must also be in a right-handed system. Values of $|F|^2$ for pairs of Bragg reflections hkl and $\bar{h}\bar{k}\bar{l}$ are measured and the magnitude and sign of their difference are

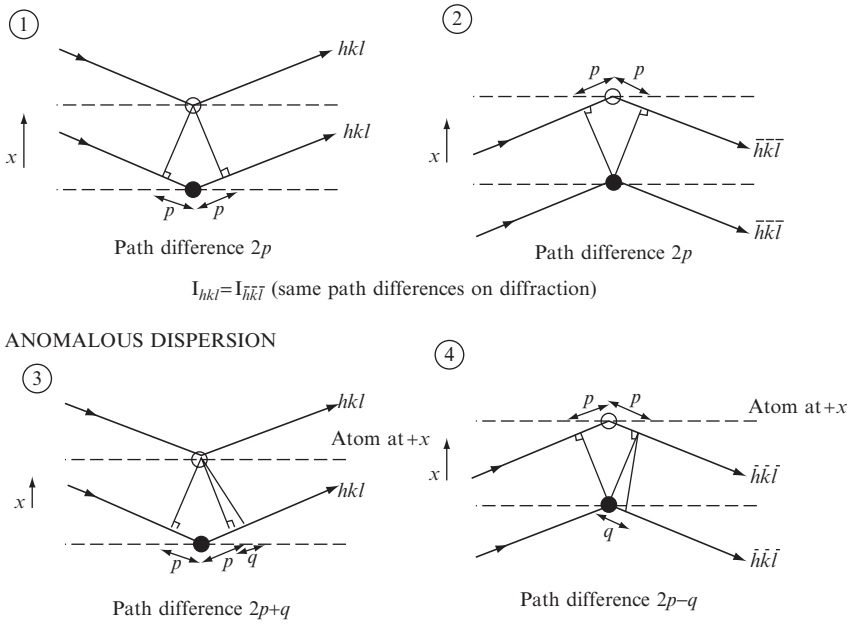


Fig. 10.4 Path differences on anomalous scattering in a noncentrosymmetric structure.

Effect of anomalous scattering on the path lengths of diffracted X-ray beams. Suppose that for a particular reflection the anomalous scatterer (black circles) causes in effect a path difference, q, in addition to the usual difference of $2p$ between the radiation scattered by a normal scatterer at this position and by a normal scatterer at some other position (open circles). As shown, the path difference for the hkl reflection with anomalous scattering is $2p + q$ and that for the $\bar{h}\bar{k}\bar{l}$ reflection is $2p - q$. If no anomalous scattering had occurred, these would be the same—namely, $2p$. Since the intensity of a diffracted beam depends on the path differences between waves scattered by the various atoms in the unit cell, the result of anomalous scattering is an intensity difference between hkl and $\bar{h}\bar{k}\bar{l}$. It is possible to compute values of $|F(hkl)|$ and $|F(\bar{h}\bar{k}\bar{l})|$ and see which should be the larger. If for many reflections the relations of the calculated values to the experimentally measured values are the same as those calculated for the model, then the model has the correct handedness (configuration); if not, the configuration of the model must be changed. That is, if $|F_o(hkl)| > |F_o(\bar{h}\bar{k}\bar{l})|$, then we must necessarily have $|F_c(hkl)| > |F_c(\bar{h}\bar{k}\bar{l})|$. See Appendix 11.

compared with the calculated value of $4f''(B'G - A'H)$ [see Eqn. (10.6)]. G and A' are cosine terms and do not change sign if the "handedness" of the system in which the model is calculated is changed. However, B' and H are sine terms, and if the signs of x, y, and z for all the atoms in the model are reversed, then B' and H change sign. Therefore, if $(|F(hkl)|^2 - |F(\bar{h}\bar{k}\bar{l})|^2$ and $(B'G - A'H)$ have opposite signs, the values of x, y, and z in the model must be replaced by $-x, -y, -z$ to give the correct model. An example is given in Appendix 11. The result of *maintaining the same handedness for the axes in real and reciprocal space* is a three-dimensional representation of the molecule from which the absolute configuration can be seen directly.

In order to establish the absolute configuration of a crystal structure it is necessary (if anomalous scattering has taken place) to compare $I(hkl)$ and $I(\bar{h}\bar{k}\bar{l})$, note which is larger, and compare this information with the result of a structure factor calculation done with a model of the structure. If there is not agreement between the signs of these observed and calculated intensity differences, the handedness of the model should be reversed. The signs of the differences should be correct in all cases where they are large (keeping in mind the standard uncertainties of their measurements). Alternatively, a Flack parameter, x, can be calculated. This is obtained by the equation

$$I(hkl) = (1 - x)|F(hkl)|^2 + x|F(\bar{h}\bar{k}\bar{l})|^2 \qquad (10.7)$$

and is often part of the least-squares refinement (Flack, 1983). The value found for x for all data generally lies between 0 and 1. If x is near 0 with a small standard uncertainty, the absolute structure that has been obtained is probably correct. If x is near 1, then the signs of all x, y, and z in the structure must be reversed. If x is near 0.5, the crystal may be racemic or twinned, and further investigation is necessary.

In 1930 Coster, Knol, and Prins were able to determine the absolute configuration of a zinc blende (ZnS) crystal (Coster et al., 1930).* This contains, in one direction (a polar axis) through the crystal (the one perpendicular to the 111 face), pairs of layers of zinc and sulfur atoms separated by a quarter of the spacing in that direction and then another pair one cell translation away, and so on (Figure 10.5). The sense or polarity of that arrangement was determined by the use of radiation (gold, AuLα_1, λ = 1.276 Å, AuLα_2, λ = 1.288 Å) near the K-absorption edge of zinc (1.283 Å). The AuLα_1 radiation caused anomalous scattering by the zinc atoms, but the AuLα_2 radiation did not. As a result it was shown that the shiny $(\bar{1}\bar{1}\bar{1})$ faces have layers of sulfur atoms on their surfaces and the dull (111) faces have layers of zinc atoms on their surfaces (see Figure 10.5).

This method was extended, as described above and in Appendix 11, by Bijvoet, Peerdeman, and van Bommel in 1951 to establish the absolute configuration of (+)-tartaric acid in crystals of its sodium rubidium double salt using zirconium radiation, which is scattered anomalously by rubidium atoms and ions (Bijvoet et al., 1951). The result is shown in Figure 10.6a. The absolute configuration was unknown until that time;

* Zinc blende, ZnS, crystallizes in a cubic unit cell, a = 5.42 Å, space group $F\bar{4}3m$. The structure contains Zn at (0,0,0), (0, ½, ½), (½, 0, ½), and (½, ½, 0) and sulfur at (¼, ¼, ¼), (¼, ¾, ¾), (¾, ¼, ¾), and (¾, ¾, ¼). The shiny, well-developed faces have sulfur atoms on their surfaces, while the rougher, matte faces have zinc on their surfaces. When pressure is applied perpendicular to the 111 face, the shiny faces become, by the piezoelectric effect, positively charged and the matte faces become negatively charged.

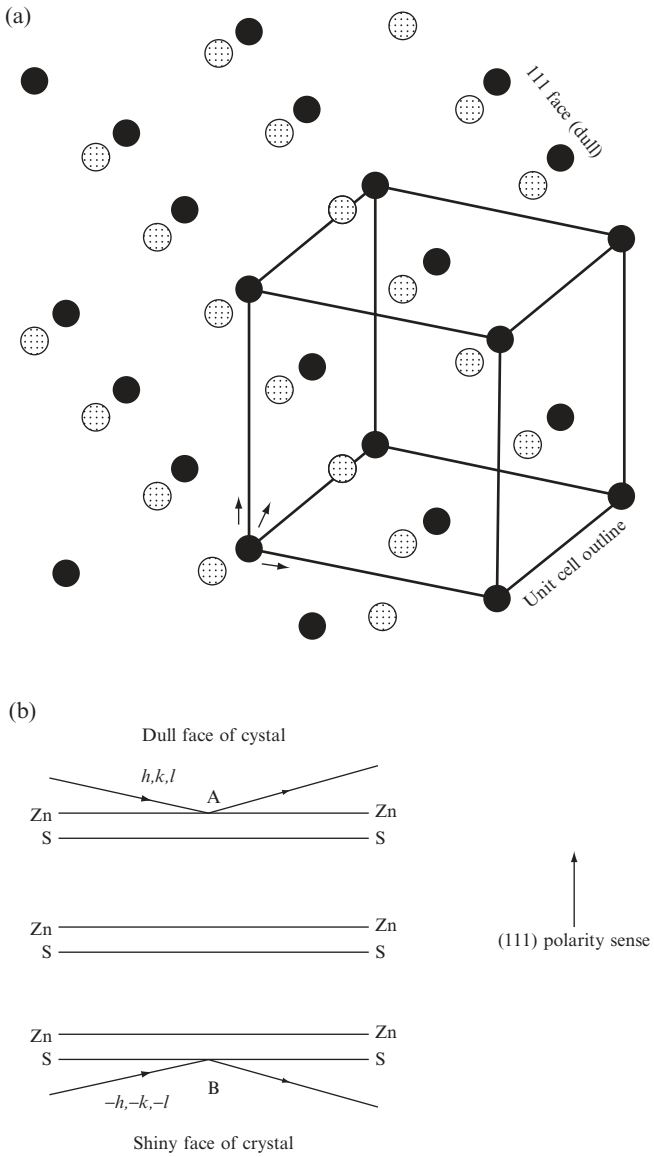


Fig. 10.5 Polarity sense of zinc blende.

(a) The structure of zinc blende, showing the arrangement of zinc and sulfur atoms. Two views are shown, one down an axis and the other to show the planes of atoms in the [111] direction. Zinc blende is often called sphalerite. Zn black, S stippled.

(b) Reflections from the two faces of zinc blende (dull and shiny) will have different relative path differences for the zinc and the sulfur atoms (compare with Figure 10.4). If the radiation is near the absorption edge of zinc, the two types of reflections will have different intensities, allowing one to determine (as did Coster, Knol, and Prins in 1930) that the dull face has zinc atoms on the surface and the shiny face has sulfur atoms on the surface.

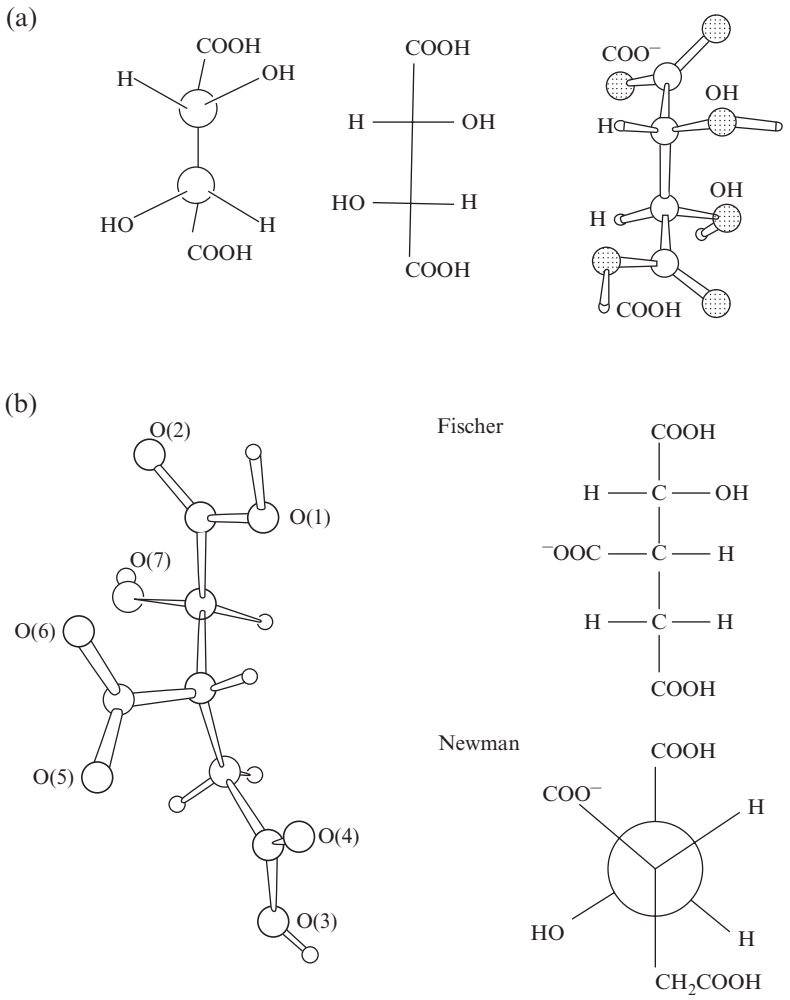


Fig. 10.6 Absolute configurations of biological molecules.

(a) Absolute configuration of (+)-tartaric acid (dextrorotatory tartaric acid). Note that in the actual structure (right) the chain of four carbon atoms has effectively a planar zigzag arrangement. In the formula on the left, and by convention in all "Fischer formulas," vertical carbon chains are represented as planar but with successive bonds always directed into the page. Thus, in the formulas in the center and left here, the lower half of the molecule has been rotated 180° relative to the upper half as compared with the actual structure. This affects the conformation but not the absolute configuration of the molecule. The conformation of tartaric acid illustrated on the left is a possible one for this molecule, but it is of higher energy (because bonds are eclipsed) than that shown on the right, the conformation observed in the crystals studied by Bijvoet et al. (1951). Still other conformers may exist in solution or in other crystals.

(b) Absolute configuration of the potassium salt of (+)-isocitric acid (isolated from the plant *Bryophyllum calycinum*). Fischer and Newman formulas are shown. The correct designation of this enantiomer is 1*R*:2*S*-1-hydroxy-1,2,3-propanetricarboxylate. The torsion angles are shown in Figure 12.5.

From *Acta Crystallographica* **B24** (1968), p. 585, Figure 4.

fortunately that which was found was the one arbitrarily chosen from the two possibilities half a century earlier by Fischer (Fischer, 1890, 1894), so the current organic chemistry textbooks did not have to be changed. The absolute configurations of many other molecules have been determined either by X-ray crystallographic methods or by chemical correlation with those compounds for which the absolute configuration had already been established (see Figure 10.6b). Values of anomalous scattering factors, especially those near the absorption edge, have been measured in detail with synchrotron radiation (see, for example, Templeton et al., 1980).

But how can absolute configuration be represented? The R/S system of doing this involved assigning a priority number to the atoms around an asymmetric (carbon) atom so that atoms with greater atomic number have the higher priority (Cahn et al., 1966). If two atoms have the same priority, *their* substituents are considered until differentiation of priorities can be established (otherwise, of course, the central atom is not asymmetric). Then the structure is viewed with the atom of lowest priority directly behind the central (carbon) atom and the other substituents are examined. Then if the order of the substituents going from highest to lowest priority is clockwise, the central atom is designated R (Latin *rectus*, right). If it is anticlockwise, the central atom is designated S (Latin *sinister*, left). As a result, once the absolute configuration is established and each asymmetric tetrahedral atom has an R or S designation, sufficient information is provided from these designations to make it possible to build a model with this correct absolute configuration.

The effect of anomalous scattering was used to solve the structure of a small protein, crambin, containing 45 amino acid residues (and which crystallized with 72 water and 4 ethanol molecules per protein molecule) (Hendrickson and Teeter, 1981). The nearest absorption edge of sulfur is at 5.02 Å, but for CuKα radiation, wavelength 1.5418 Å, the values of f' and f'' are 0.3 and 0.557, respectively, for sulfur. Pairs of reflections [$|F(hkl)|$ and $|F(\bar{h}\bar{k}\bar{l})|$] were measured to 1.5 Å resolution (the crystals scatter to 0.88 Å resolution); sulfur atom positions were calculated from Patterson maps with $|\Delta F|^2$. While it was necessary to take into account possible errors in such measurements of the differences of two large numbers, it was, in fact, possible to determine the positions of the three disulfide links (six sulfur atoms). The structure was then determined from an analysis of the Fourier map calculated on the heavy-atom parameters of the sulfur atoms together with a partial knowledge of the amino acid sequence.

SAD and MAD phasing

The use of anomalous scattering in structural work has increased recently since the advent of "tunable" synchrotron radiation—that is, X rays whose wavelength may, within certain limits, be selected at

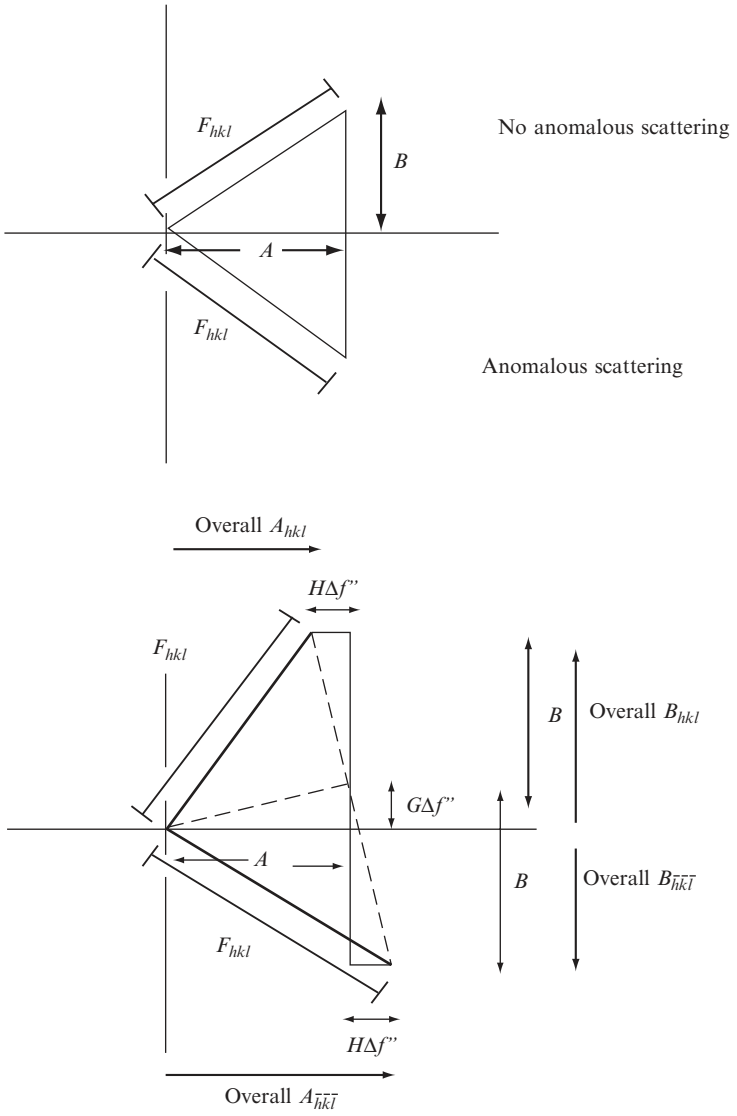


Fig. 10.7 Effects of anomalous scattering on F values.

The top diagram shows the structure factor vectors for $\mathbf{F}(hkl)$ and $\mathbf{F}(-h, -k, -l)$ in the absence of anomalous scattering and the lower diagram shows the effect of anomalous scattering on such a diagram. When there is anomalous scattering, $A(hkl) \neq A(-h, -k, -l)$. The same occurs for B values, as shown (see Chapter 5 and especially Figure 5.1 for the fundamentals of such diagrams).

will. As a result it is possible to measure the diffraction pattern of a macromolecular crystal with X-radiation of wavelengths near and also far from the absorption edges of any anomalous scatterers present. Two data sets can be measured, one near and one far from any absorption edge of atoms in the crystal. The integrity of the crystal during so much radiation exposure is maintained by flash freezing. For example, the

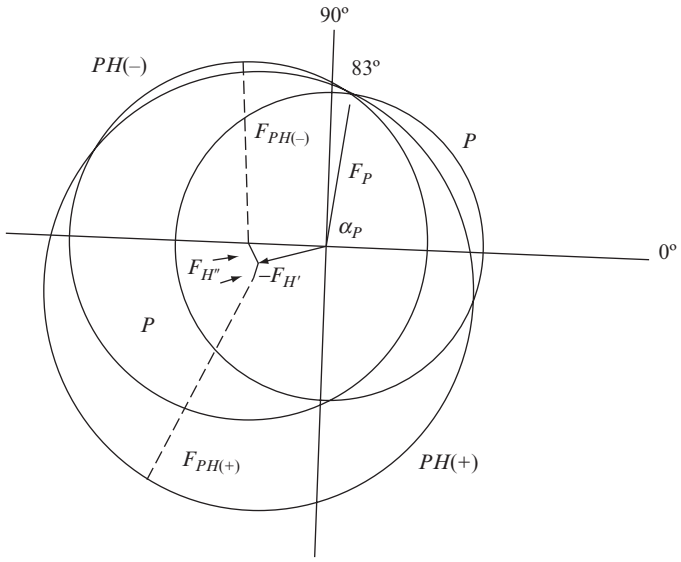


Fig. 10.8 Isomorphous replacement plus anomalous scattering (noncentrosymmetric).

The effect of anomalous scattering by an atom M, introduced to replace another atom, may be used to resolve the ambiguity in phase-angle determination by the isomorphous replacement method. The effect of anomalous scattering (Appendix 11) is to introduce a phase shift, which means in effect, to change the atomic scattering factor of atom M from a "real" quantity, f, to a "complex" one, $(f + f') + \mathrm{i}f''$. Suppose A and B refer to that part of the structure that does not scatter anomalously, and A' and B' to the total structure without any f'' component; then $A' = G(f + f') + A$ and $B' = H(f + f') + B$, where G and H refer to geometric components for the anomalous scatterer, M. A'' and B'' are components of the structure when anomalous scattering is present, and A''_{M} and B''_{M} are components for the anomalous scatterer M alone. Then

$$A''_{\mathrm{M}} = G(f + f' + \mathrm{i}f'') = A_{\mathrm{M}} + Gf' + \mathrm{i}Gf''$$

$$B''_{\mathrm{M}} = H(f + f' + \mathrm{i}f'') = B_{\mathrm{M}} + Hf' + \mathrm{i}Hf''$$

Then, for the entire structure, including anomalous-scattering effects, we have

$$A'' = A + Gf + Gf' + \mathrm{i}Gf'' = A' + \mathrm{i}Gf''$$

$$B'' = B + Hf + Hf' + \mathrm{i}Hf'' = B' + \mathrm{i}Hf''$$

As shown in Appendix 11, we have for the entire structure with anomalous scattering (by separating and squaring the "real" and "imaginary" components)

$$|F(hkl)| = \sqrt{(A' - Hf'')^2 + (B' + Gf'')^2}$$

$$|F(\bar{h}\bar{k}\bar{l})| = \sqrt{(A' + Hf'')^2 + (B' - Gf'')^2}$$

(see Figure 10.4). Therefore, $|F(hkl)|$ and $|F(\bar{h}\bar{k}\bar{l})|$ are different; the intensity of each reflection is measured to see which is the greater, as shown in Appendix 11.

A diagram to illustrate the determination of a phase angle for a macromolecule (that is, α_P) by the combination of isomorphous replacement and anomalous scattering is shown. This diagram is constructed in a similar way to Figure 9.9. Circles of radii $|F_{PH(+)}|$ and $|F_{PH(-)}|$ (for reflections of the heavy-atom derivative with indices h, k, l and $-h$, $-k$, $-l$, respectively) are drawn with centers at $-(\mathbf{F}_{H'} + \mathbf{F}_{H''})$ and $-(\mathbf{F}_{H'} - \mathbf{F}_{H''})$, respectively. There are now three circles, radii $|F_P|$, $|F_{PH(+)}|$, and $|F_{PH(-)}|$, and these intersect at a phase angle of $\alpha_P(83^\circ)$. This is probably the phase angle of this reflection, h, k, l, for the native protein.

method of anomalous scattering may be combined with isomorphous replacement in protein structure determination. Three data sets are needed for this. One involves the protein crystal, and one a heavy-atom derivative of this protein. A third data set is measured with X rays of a wavelength that will cause maximal anomalous scattering by the heavy atom. The heavy-atom position is located from the first two data sets, and phase information is aided by the nonequivalent Friedel pairs of Bragg reflections; these remove ambiguities in phase determination (see Figures 10.7 and 10.8). This makes it possible to obtain approximate phases from just one heavy-atom derivative.

The multiwavelength anomalous dispersion (MAD) method, suggested by Wayne Hendrickson, is now a method of choice for phase determination (Hendrickson, 1991). Generally proteins that are used have been biologically expressed in a medium that contains only selenomethionine. As a result the protein contains selenomethionine in its sequence where methionine would normally be expected. Therefore the strong anomalous signal of selenium can be used to derive phases. X-ray diffraction data are measured near the absorption edge (where f'' has a maximum value, 1.15 electrons), and also at one or two wavelengths remote from any absorption edge. Only one crystal is needed, and the data are generally measured at a synchrotron source.

In the single-wavelength anomalous dispersion (SAD) method, diffraction data for one wavelength of radiation are measured on a heavy-atom-containing protein, not necessarily near an absorption edge. Since when heavy atoms are soaked into a crystal they may attach to various side chains in a disordered manner, the strategy has been to generate a protein with a heavy atom, such as that in iodophenylalanine, chemically incorporated into one amino acid (Dauter, 2004). The value of f'' for iodine is 6.91 electrons for CuKα radiation. Alternatively, chromium Kα radiation ($\lambda = 2.2909$ Å) may be used to locate sulfur, which has an anomalous signal (f'') of 1.14 electrons for CrKα radiation, twice that for CuKα radiation (Yang et al., 2003). This means that naturally occurring protein side chains such as those of methionine or cysteine are sufficient to provide phasing with CrKα radiation. Also, the data collection can be done in a laboratory and no tuning of radiation wavelength is needed; a simple X-ray tube can be used. The phase ambiguity that comes from measuring just one data set can be aided by direct methods or by density modification. Obviously crystallographers are now experimenting with different wavelengths of radiation and different possible anomalous scatterers, and the trend is to study a crystal that contains only the molecule under study (and not variations such as heavy-atom derivatives).

Sine-Patterson map

A modified Patterson map can be used to determine the absolute configuration of a structure provided Bijvoet pairs of reflections have been

measured and correctly indexed. The map is calculated with a function with $\{|F(hkl)|^2 + |F(\bar{h}\bar{k}\bar{l})|^2\}$ as coefficients and a cosine term; this gives peaks corresponding to Eqn. (9.1), that is, vectors between atoms. Another function, with $\{|F(hkl)|^2 - |F(\bar{h}\bar{k}\bar{l})|^2\}$ as coefficients and a sine term, known as the sine-Patterson map,

$$P_s(u, v, w) = \frac{1}{V_c} \sum \sum_{\text{all } hkl} \sum \left\{ \left|F(hkl)\right|^2 - \left|F(\bar{h}\bar{k}\bar{l})\right|^2 \right\} \sin 2\pi(hu + kv + lw) \tag{10.8}$$

will have only vectors between anomalously and nonanomalously scattering atoms, and these peaks are positive if the vector is from an anomalously scattering atom to a normal atom, and negative if the vector is in the other direction. This map is asymmetric. Thus the absolute configuration of the structure may be determined from such a map (Okaya et al., 1955).

What effect does anomalous scattering have on the calculated electron density, since a term in the scattering factor now has an "imaginary" component? The answer is that the calculated electron density must be real, and, to obtain this, any effect of anomalous scattering (which involves a complex scattering factor) must be removed (as described in Appendix 11) (Patterson, 1963). This, of course, also removes any means of distinguishing one enantiomorph from the other; such information is contained only in the anomalous-scattering data.

Summary

If an atom in the crystal appreciably absorbs the X rays used, there will be a phase change for the X rays scattered by that atom relative to the phase of the X rays scattered by a nonabsorbing atom at the same site. This phase change on absorption leads to anomalous scattering and, for a noncentrosymmetric structure, results in differences in values of $|F(hkl)|^2$ and $|F(\bar{h}\bar{k}\bar{l})|^2$ that are not found in the absence of anomalous scattering. If the structure contains only one enantiomorph of a molecule, its absolute configuration may be determined by a comparison of the signs of the observed and calculated values of $(|F(hkl)|^2 - |F(\bar{h}\bar{k}\bar{l})|^2)$.