

[2] Introduction to Crystallography

By ROBERT M. SWEET

Introduction

The modern language of enzymology and molecular biology owes much of its sophistication to the success of X-ray diffraction. By 1965, clever chemists had learned a lot about how enzymes worked. They had found residues that lay close to one another in active sites and had proposed mechanisms of action that were tested by kinetic and model studies. But no one really began to know how it all worked until three-dimensional structures were known for several of those biochemical factories. Today, students of biochemistry take more as dogma than as experimental findings the knowledge of molecular structure that is their heritage. This knowledge has its value, of course. With rare exceptions (Mozart, Einstein) the imagination of man is limited by what he already *knows*. Therefore, we may presume that modern students of enzymology will be able to ask questions we might not have thought of 20 years ago, when we were ignorant of structure.

The winning of the ability to determine these structures was very difficult. The workers who produced most of the results have stood firmly on the shoulders of the scientists whose labors are described in the previous chapter. Figuring out how to determine protein structure required some 40 years. Producing the structural information we have now has been accomplished in 20. The purpose of this chapter is to introduce the general scientific reader to some of the principles that are used in the determination of molecular structure by X-ray diffraction techniques. Our goal is to present the physical and mathematical basis of these techniques and to provide an intuitive approach to understanding them. We refer you to the comprehensive textbooks in the field for a more thorough treatment of this background information. Titles which may be of use are *Protein Crystallography* by Blundell and Johnson and Sherwood's *Crystals, X-Rays and Proteins*. Brief summaries of the field can be found in *Crystal Structure Analysis: A Primer* by Glusker and Trueblood, and in a chapter from Vol. 13 of *Methods of Biochemical Analysis* by Holmes and Blow entitled "The Use of X-Ray Diffraction in the Study of Protein and Nucleic Acid Structure."

We shall introduce you to the field in two steps. The first step is to teach the fundamental principles of diffraction, and we do this in two

different ways, in parallel. The first will show how waves constructively and destructively interfere after they are scattered from atoms. The second way will show that X-ray diffraction is mathematically equivalent to the taking of the Fourier transform of the scattering object. The second step is to describe the experimental procedures that a crystallographer must perform in the determination of macromolecular structure. This will be in the form of an approximate chronology of work that might be done in the crystallographic laboratory.

Diffraction

Scattering of X Rays by Atoms. The X rays commonly used for diffraction studies of biological molecules have a wavelength of 0.154 nm and an energy of 8 keV. They are produced when a beam of electrons driven by a potential of roughly 40,000 V strikes a copper anode. These high-energy electrons ionize the copper atoms, removing an inner shell electron. X rays are emitted when a higher energy electron falls to fill the void. The 0.154-nm radiation results when an L shell electron fills a hole in the K shell of a copper atom. This choice of radiation is a compromise: longer wavelength X rays allow investigation of larger molecules, damage the specimen less, and scatter more strongly; shorter wavelength X rays are absorbed less by the specimen and can allow solution of a structure to higher resolution. In some large laboratories, electron synchrotrons or storage rings are used to produce X rays for diffraction studies. In these accelerators electrons travel at nearly the speed of light, their orbits being bent to a circular path by powerful magnets. The radiation is produced in the direction of travel of the electrons, essentially because of the work done on them by the bending magnets. Synchrotron radiation is polychromatic; therefore one may choose precisely the wavelength for use by diffraction of the beam from a monochromator crystal. A major advantage of the synchrotron sources is that they are two to three orders of magnitude more intense than conventional ones.

X rays are electromagnetic radiation. When they pass by the electrons in an atom, the oscillating electric field of the X-ray photon or wave causes the electrons on the atom to oscillate, much like the sloshing of coffee in a cup. These oscillating electrons serve as a new source of X rays which are emitted in almost all directions. (A little thought will show the reader that there are two directions in which no radiation can be emitted when the incident radiation is polarized.) It happens that the scattered radiation is precisely out of phase with the incident radiation, that is, it is phase-shifted by π .

Diffraction of X Rays by Atoms in a Lattice. A crystal has some simple arrangement, or motif, of atoms repeated by reiterated translations in one, two, or three directions. The lattice that describes the translational repetition can be defined by single points chosen at an equivalent point in each motif. The simplest possible crystal is made of two atoms, and its lattice consists of two points at the atomic centers.

When a single X-ray photon passes through a crystal, it diffracts from the entire crystal as if it were a plane wave. This is an example of the wave-particle duality explained by quantum mechanics. Thus, when a photon strikes a lattice of two atoms, as in Fig. 1, it is scattered by both atoms. Because of the constructive and destructive interference of waves, there are special directions in which radiation scattered from these two atoms will form a diffracted beam. One can see in Fig. 1 that the two diffracted waves are in phase because the lower of the two waves travels exactly one wavelength farther than the upper. Notice the phase shift of π in the scattered waves. In general this path difference must be an integral number of wavelengths for diffraction maxima to occur:

$$a \sin \alpha = n\lambda, \quad n = 1, 2, \dots \quad (1)$$

The pattern of intensity which would be observed by an X-ray detector is a series of ripples, starting at the direct beam and going off in both directions.

Diffraction and the Fourier Transform. If the elder Bragg was the father of crystallography, then Jean Baptiste Joseph Fourier was surely its godfather. Fourier was a bureaucrat in the government of Napoleon Bonaparte and, among other services to his government, accompanied Napoleon on his visit to Egypt and was an able administrator after their return to France. He was an accomplished mathematician and made contributions that were of great value to crystallographers. He showed that

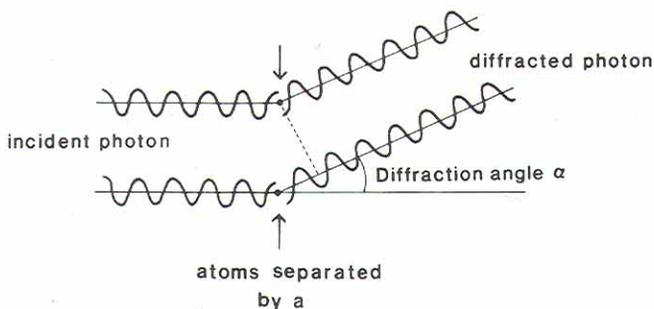


FIG. 1. Diagram of photon striking a lattice of two atoms.

any periodic function can be approximated by sums of sine and cosine functions whose wavelengths were integral fractions of this periodicity. He also devised the technique we now know as the Fourier transform, which again involves sums of trigonometric functions, and which transforms functions between coordinate systems with different dimensionality. The Fourier transform is a precise mathematical description of the physical phenomenon of diffraction, and therefore it is tremendously useful to us.

The one-dimensional Fourier transform of some function $f(x)$ is

$$F(h) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x)e^{ihx} dx \quad (2)$$

If $f(x)$ exists in ordinary space, defined by x , then $F(h)$ exists in reciprocal space, defined by the variable h . Because the exponent must be dimensionless, h must have dimensions of reciprocal distance.

The beauty of Fourier's transform is that having gotten there, one can get back again with

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(h)e^{-ihx} dh \quad (3)$$

Notice several features of these two transformations. First, both contain a complex exponential that yields the trigonometrics mentioned above. Second, the principal difference between them is the sign of the exponent. And third, readers who still remember college calculus can prove that the second follows from the first.

We may now use this transform to calculate an expression for diffraction from a two-atom lattice. (I refer you to Sherwood's book for a thorough treatment.) If we represent each atom by an appropriate δ function, that is, we make the atom a point scatterer of X rays, we get that

$$F(h) = 2 \cos 2ha \quad (4)$$

$F(h)$ is the amplitude (square root of the intensity) of diffraction and a is the spacing between scatterers. A definition of h awaits our discussion of the work of the Braggs. Notice, however, that this function is like the array of ripples that we expected.

Diffraction from a Two-Dimensional Lattice. In our discussion of diffraction from a one-dimensional lattice, we arranged for the incident beam to be perpendicular to the line of atoms. Now we shall show how to calculate conditions for diffraction for any beam that is incident on a two-dimensional lattice of atoms. The situation we shall consider is shown in Fig. 2.

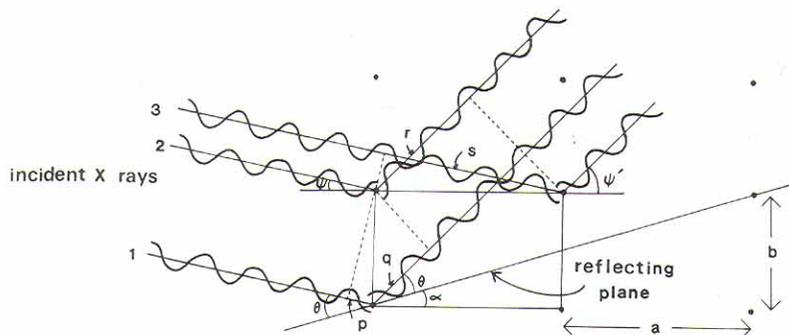


FIG. 2. Diffraction from a two-dimensional lattice.

For simplicity we have chosen a rectangular lattice. The lattice spacing is a in the horizontal and b in the vertical directions. Parallel beams 1, 2, and 3 approach the crystal at angle ψ . We have drawn a number of wave fronts as dotted lines at points where one or another of the waves strikes an atom. We place the same condition on the beams here as in the two-atom case: for diffraction to occur, the path length difference between any two beams must be an integral number of wavelengths. In studying this drawing, note that it has been carefully made, including the phase shift of π on scattering, to show the path length differences. These differences between beams 1 and 2 and between beams 2 and 3 are, respectively,

$$\begin{aligned} p + q &= b \sin \psi + b \sin \psi' = n\lambda \\ s - r &= a \cos \psi - a \cos \psi' = m\lambda \end{aligned} \quad (5)$$

Now we construct a "reflecting" plane which makes angle θ with both incident and diffracted beams. This plane lies at angle α from the horizontal axis. We can now redefine ψ and ψ' in terms of θ and α , as follows:

$$\psi = \theta - \alpha, \quad \psi' = \theta + \alpha \quad (6)$$

Finally, if we substitute these values into the equations above and rearrange, we are left with a remarkably simple equation which constitutes an important condition for diffraction:

$$\tan \alpha = mb/na \quad (7)$$

In this example $m = 1$ and $n = 2$ and we have drawn the reflecting plane to pass through the appropriate two atoms in the lattice so that this expression holds true. Planes of this sort, that pass through two lattice points (or three in a three-dimensional lattice), or are parallel to planes that do, are known as General Lattice Planes. These planes have useful algebraic properties which we shall discuss later.

But first, we shall place yet another set of constraints on diffraction from a lattice. We have shown that diffraction can be treated as reflection from general lattice planes, but have placed no constraints on the reflection angle θ . The following derivation is similar to that which sent the Braggs on their way and led to the founding of the science of X-ray diffraction crystallography.

Consider two general lattice planes, as in Fig. 3, with interplanar spacing d , and incident and reflected beams of X rays that make angle θ with these planes. As before, we require that the path length difference between the two rays be an integral number of wavelengths, and we arrive at Braggs' Law:

$$n\lambda = 2d \sin \theta \quad (8)$$

Braggs' law does not tell us everything we need to know about diffraction. Coupled with the idea of general lattice planes, it tells us only about the geometry of diffraction, but it says nothing about the relation between diffraction and the way the atoms of the crystal are arranged within its repeating units. The Fourier transform does this. But before we see how the Fourier transform of the crystal is related to Braggs' law, we must first learn a bit more about general lattice planes and the ways they can be represented.

Representation of Lattice Planes. General lattice planes are described by sets of *general indices*. The smallest unit of a crystal that is repeated by translation alone is called the *unit cell*. We often draw unit cells with lattice points at their corners. When general lattice planes are constructed in a unit cell, the indices which describe those planes are the number of segments into which the planes cut each of the unit cell edges. An equivalent definition is that the indices are reciprocals of the fractional lengths into which unit cell edges are cut by the planes. Examples appear in Fig. 4. The sign conventions are as follows. When the plane nearest a lattice point cuts two axes that go both in the positive or both in the negative direction, the two corresponding indices have the same sign. When the

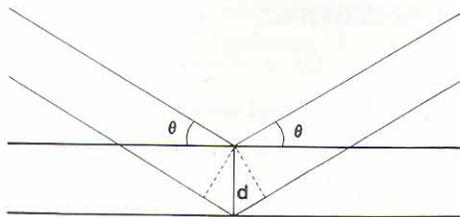


FIG. 3. Diagram of general lattice planes showing interplanar spacing and angles of reflectance and incidence.

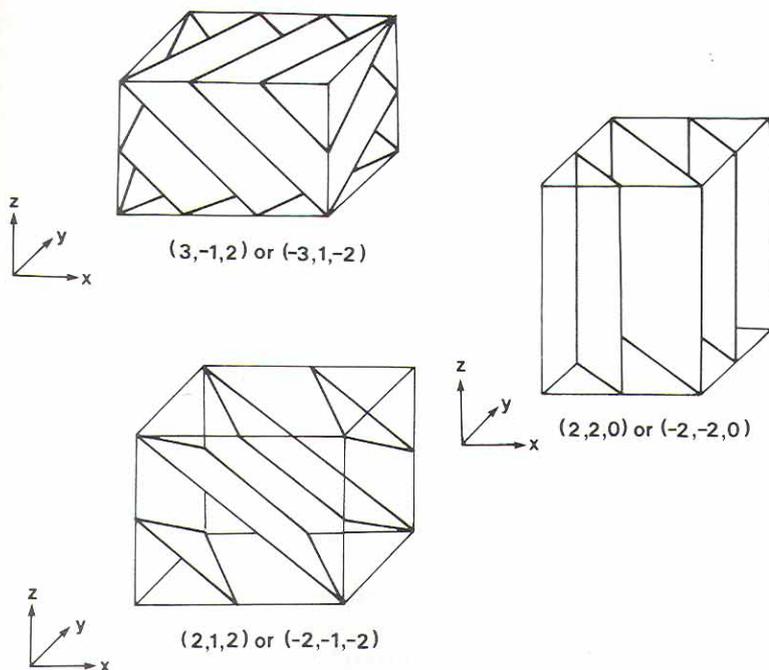


FIG. 4. Representation of lattice planes.

plane cuts axes that go in directions with opposite signs, the indices have opposite signs. A set of indices can be multiplied by -1 and still represent the same set of planes.

The Sphere of Reflection and Reciprocal Space. P. P. Ewald noticed a geometrical simplification of Bragg's law that links this law to the Fourier transform and provides a simplified way of looking at general lattice planes.

Notice first, as in Fig. 5, that if the Bragg reflection angle is θ then the total deflection angle for diffracted X rays must be 2θ . Let us then place this figure, an intersection of beams with a plane, at the center of a sphere, known as the Ewald sphere, that has as its radius the reciprocal of the wavelength of the radiation. Then one can make the construction shown in Fig. 6. One can see immediately that this figure follows the rule

$$\frac{1}{2}OA/(1/\lambda) = \sin \theta \quad \text{or} \quad \lambda = (2 \sin \theta)/OA \quad (9)$$

If we substitute $OA = 1/d$, we obtain

$$\lambda = 2d \sin \theta \quad (10)$$

which is Bragg's law with $n = 1$.

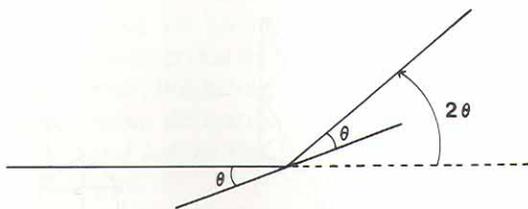


FIG. 5. Diagram of reflection and deflection angles.

Significant features of this construction are first that the chord OA is perpendicular to the reflecting plane, and second that when we define its length to be the reciprocal of the distance between planes which will cause diffraction at angle θ , Bragg's law is obeyed. This suggests that in a crystal each set of general lattice planes (hkl), with spacing d_{hkl} might be represented by a vector \mathbf{s}_{hkl} that is perpendicular to the planes and has length which is $|\mathbf{s}_{hkl}| = 1/d_{hkl}$. The vectors \mathbf{s}_{hkl} define the mathematical space with dimensions of reciprocal distance that we know as *reciprocal space*.

Reciprocal space, the reciprocal lattice (that is, the set of points defined by all of the vectors \mathbf{s}_{hkl}), and the Ewald sphere are a remarkably useful heuristic tool for the crystallographer. He can think about diffraction in simple geometric terms as involving the intersection of a point with a sphere rather than having to think about reflecting planes, their spacings, and the angles they make with incident and diffracted rays. In Fig. 6, when point A , which represents vector \mathbf{s}_{hkl} for the plane we have chosen,

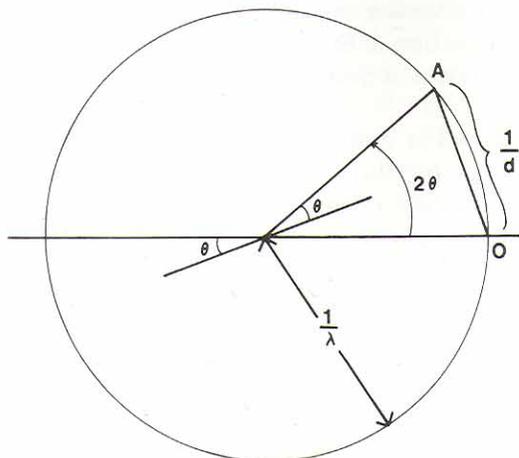


FIG. 6. Ewald sphere.

touches the Ewald sphere, Bragg's law is obeyed and diffraction occurs. Rotating the crystal, so that the angle with the incident beam is no longer θ , will also move A away from the the surface of the sphere since \mathbf{s}_{hkl} is perpendicular to the plane (hkl) . Therefore, Bragg's law will not be obeyed and no diffraction will be observed.

It is easy to define the reciprocal lattice vector \mathbf{s}_{hkl} to be perpendicular to the planes (hkl) and to have length $s_{hkl} = 1/d_{hkl}$:

$$\mathbf{s}_{hkl} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \quad (11)$$

The principal reciprocal space vectors \mathbf{a}^* , \mathbf{b}^* , and \mathbf{c}^* are defined, in terms of the "real space" vectors or unit cell principal axes, as

$$\mathbf{a}^* = \frac{\mathbf{b} \times \mathbf{c}}{\mathbf{a} \times \mathbf{b} \cdot \mathbf{c}}, \quad \mathbf{b}^* = \frac{\mathbf{c} \times \mathbf{a}}{\mathbf{a} \times \mathbf{b} \cdot \mathbf{c}}, \quad \mathbf{c}^* = \frac{\mathbf{a} \times \mathbf{b}}{\mathbf{a} \times \mathbf{b} \cdot \mathbf{c}} \quad (12)$$

One can use the reciprocal lattice vector \mathbf{s}_{hkl} to calculate useful parameters for the crystal. For example, one can easily calculate the spacings for a particular set of lattice planes. For the case where the angle between unit cell edges \mathbf{a} and \mathbf{c} is unconstrained (call it β) but both \mathbf{a} and \mathbf{c} are perpendicular to \mathbf{b} , we can readily derive an expression for d_{hkl} :

$$d_{hkl} = (\mathbf{s}_{hkl} \cdot \mathbf{s}_{hkl})^{-1/2} = (h^2 a^{*2} + k^2 b^{*2} + l^2 c^{*2} + hla^*c^* \cos \beta^*)^{-1/2}$$

$$(\beta^* = \pi - \beta)$$

$$a^* = \frac{bc}{abc \cos(\beta - 90^\circ)} = \frac{1}{a \sin \beta}, \quad b^* = \frac{1}{b}, \quad c^* = \frac{1}{c \sin \beta} \quad (13)$$

$$d_{hkl} = \left(\frac{h^2}{a^2 \sin^2 \beta} + \frac{k^2}{b^2} + \frac{l^2}{c^2 \sin^2 \beta} - \frac{hl \cos \beta}{ac \sin^2 \beta} \right)^{-1/2}$$

Therefore, the concept of the reciprocal lattice and Ewald's sphere of reflection are sufficient to provide us with a thorough geometrical description of X-ray diffraction. For a crystal with any particular unit cell dimensions one can define the set of vectors \mathbf{s}_{hkl} that will determine the reciprocal lattice for that crystal. The conditions for diffraction from any particular set of planes are only that the crystal be oriented so that the reciprocal lattice point corresponding to those planes touches the Ewald sphere.

After brief digression to show how the reciprocal lattice and the Ewald sphere play a role in diffraction experiments, we shall return to show how Bragg's law and the diffraction pattern are related to the Fourier transform of a crystal.

The Nature of Diffraction from Crystals. In principle we understand how diffraction might actually occur from a crystal placed in a beam of monochromatic X rays. How does it really work and how is it used?

Notice first that if the crystallographer is to sample diffraction from all possible sets of lattice planes he must bring the reciprocal lattice points for all these planes into contact with the Ewald sphere. To do this, he must *move* the crystal in the beam. Martin Buerger devised an elegant technique for the photography of a diffraction pattern. He developed a camera that moves the crystal in the beam in a precessional motion. This motion rocks single planes of points in the reciprocal lattice through the Ewald sphere. If a metal screen with an annular slit is placed so that it precesses with the crystal and if the film is made to precess about its center point as well, an undistorted image of a single plane of reciprocal space can be recorded on the X-ray film.

In Fig. 7 you can see a diffraction photograph taken from a crystal of the protein phycocyanin with the use of a Buerger precession camera. This is the *image* of a single plane in the reciprocal lattice for this crystal, displayed on a film. Each reciprocal lattice point is represented by a spot of blackened silver grains. The darkness of the spot is proportional to the intensity of the X rays reflected from the set of planes described by that particular reciprocal lattice point.

There are several features one should notice about this photograph. First, as advertised, the Buerger camera produces reflections or diffrac-

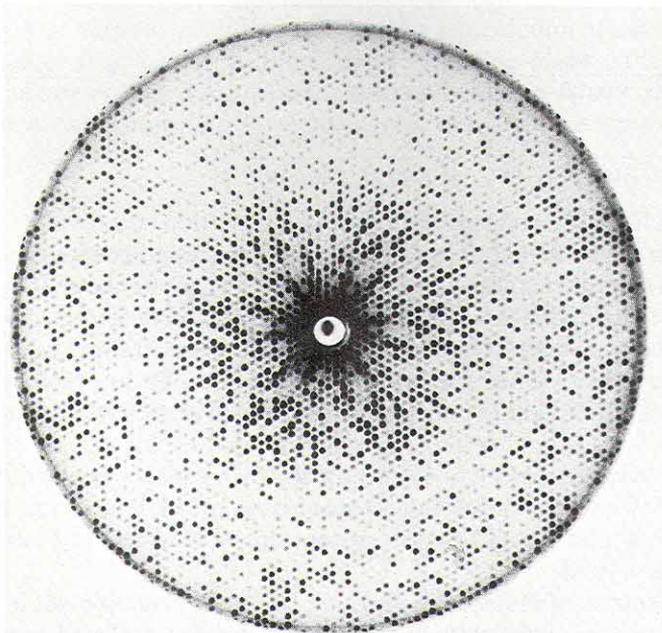


FIG. 7. Diffraction photograph of phycocyanin.

tion maxima that lie on the film along the straight lines of a lattice. The distances between the spots on the lines are proportional to the reciprocal lattice spacings a^* , etc., and these important crystal parameters can be measured directly from a precession film such as this. Second, there is a wide range of intensities in the diffraction maxima that have been recorded. This is because these intensities are determined by the Fourier transform of the contents of the phycocyanin unit cell, sampled at the points that are shown. More directly, while the *arrangement* of spots on the film tells us about the size and shape of the crystal unit cell, only their *intensities* can tell us about the arrangement of atoms in the crystal. The effort to discover how this arrangement of atoms is embedded in these intensities occupies much of the lives of crystallographers and is the subject of several of the chapters that follow. Third, one readily notices that this photograph has striking symmetry. In particular, the pattern of intensities can be put back on itself by a rotation of the photograph by one-sixth of a rotation or by 60° .

The symmetry of this photograph shows much about the total symmetry of intensities in the entire reciprocal lattice for this crystal. One can readily interpret the symmetry of the reciprocal lattice to deduce the symmetry of the molecular arrangement inside the unit cell. This volume will not deal at all with this aspect of crystallography; however the reader can find a comprehensive and comprehensible discussion of crystal symmetry in Martin Buerger's *Elementary Crystallography*.

There are numerous other techniques for the measurement of diffraction intensities, that is, for the sampling of reciprocal space. One of the simplest to understand is use of the single-crystal diffractometer (see Wyckoff [24], this volume, for an exhaustive discussion). On this instrument, a diagram of which appears in Fig. 8, a crystal can be manipulated so that any reciprocal lattice point one chooses can be made to touch the Ewald sphere in the horizontal plane. When the detector is placed at the proper diffraction angle 2θ , it can measure the intensity of diffracted rays. By this technique one can often measure the intensity of one reflection per minute on a computer-controlled instrument. A difficulty with this device is that it measures reflections only one at a time, and geometric constraints make it difficult to use with very finely sampled reciprocal lattices, that is, with very large unit cells.

A technique which suffers neither of these problems is rotation photography, described in some detail in *The Rotation Method* by Arndt and Wonacott and [19], this volume. Here we use the simplest possible geometry: a crystal is mounted in the X-ray beam and a flat piece of X-ray film in a light-tight cassette is placed a short distance away, perpendicular to the beam. The crystal is rotated through an axis perpendicular to the

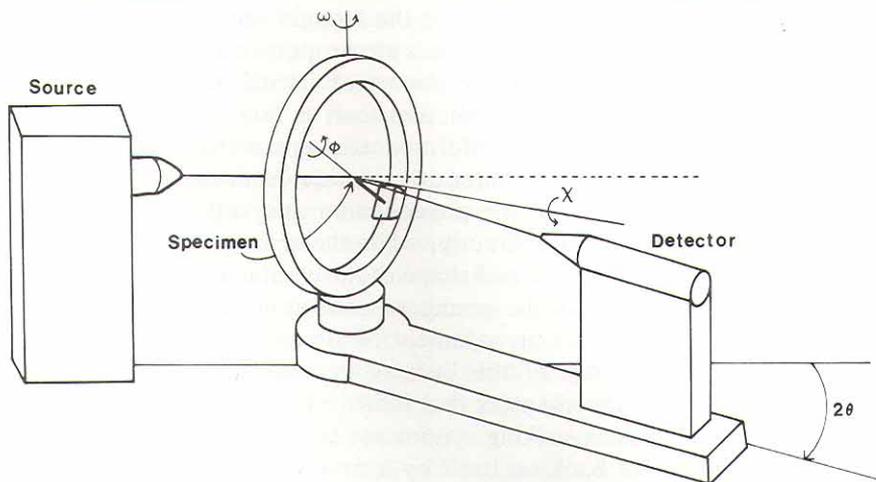


FIG. 8. Diagram of single-crystal diffractometer.

beam through only a small angle, say $1-5^\circ$, so that reflections do not superimpose one another on the film. In Fig. 9 appears a "rotation photograph" of another crystal of phycocyanin. Here the crystal was rotated through 3.5° so that regions of several reciprocal lattice planes were swept a short way through the Ewald sphere. The hexagonal arrangement of spots can be seen on the several "lunes" that appear on the photograph. Each of these continuous lunes corresponds to part of a single plane in the reciprocal lattice.

The Sampled Fourier Transform. We now need to understand how the arrangement of molecules in a crystal is manifested in a diffraction pattern, such as those in Figs. 7 and 9. To do this, we return to the idea of the Fourier transform. For simplicity we choose a simple object to represent the pattern of molecules in a crystal's unit cell. We shall use the "top hat" function (Fig. 10), which, also for simplicity, we define in one-dimensional space. The Fourier transform of this function is easily calculated as being

$$F(h) = (2b/h) \sin ah \quad (14)$$

a function that can be plotted as in Fig. 11.

Returning to our analogy of molecular crystals, Fig. 11 shows the transform of the contents of a single unit cell from a crystal. Now we place the object into a crystal, that is, we repeat it many times in Fig. 12, with each repetition being equally spaced from the last. Again we can calculate the transform of this repetitive pattern, and we find that it is a

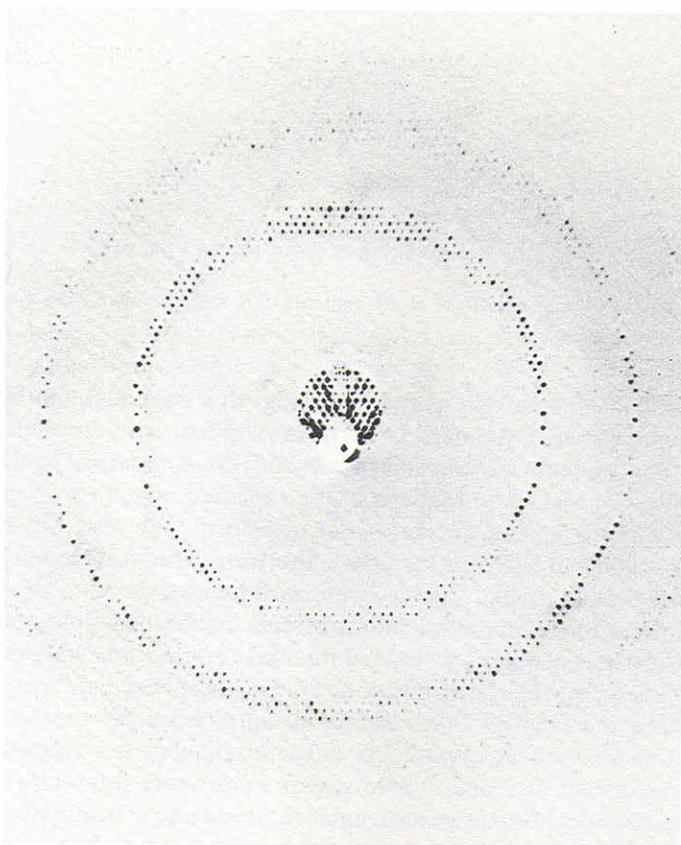


FIG. 9. Rotation photograph of phycocyanin.

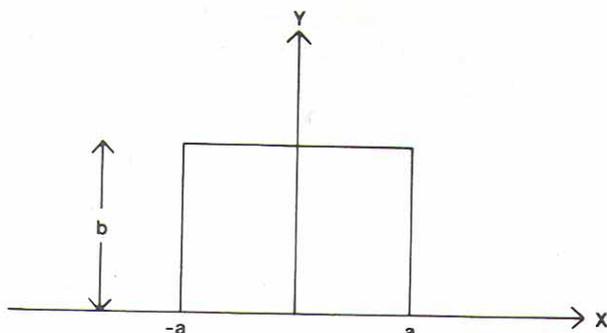


FIG. 10. "Top hat" function.

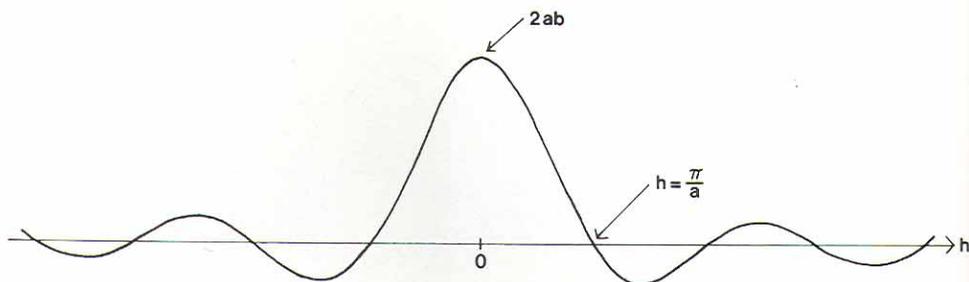


FIG. 11. Plot of Fourier transform of the contents of a single unit cell from a crystal.

series of equally spaced spikes, with spacings that depend upon the reciprocal spacing between the objects in the crystal and with amplitudes that trace out the original transform of the object. This transform is plotted in Fig. 13. We say that the transform of the repeated object (the top hat) is "sampled" at the spikes of the crystal transform. We also can see that when the objects are placed far apart, the transform will be finely sampled; when they are close, the sampling will be coarse.

Notice that the sampling of the transform occurs at points not unlike those in the reciprocal lattice. Indeed it is this observation that completes the connection from Bragg diffraction, the Ewald sphere, and the reciprocal lattice to the idea that the diffraction pattern simply represents the Fourier transform of the crystal. As h , the argument of $F(h)$, has dimensions of reciprocal distance, it also serves to measure the reciprocal lattice. The reciprocal lattice points, each representing a Bragg plane from which reflection occurs, are the same points that have nonzero values for the transform of a repetitive crystal.

Now let us take the next logical step and see how, if we know the structure of the molecules in the unit cell, we can calculate the value of the transform as it is sampled at each reciprocal lattice point.

The Structure Factor and Its Phase. How does one use mathematical notation to represent electromagnetic radiation? We say that the oscillating electric field that accompanies an X-ray photon has a wavelength, an

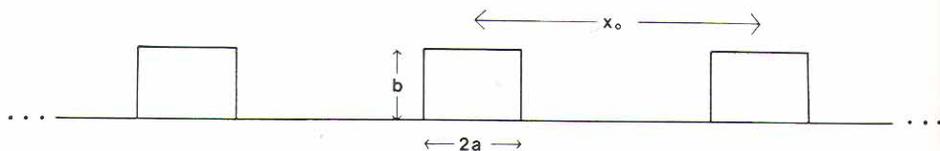


FIG. 12. Repetition of object within a crystal.

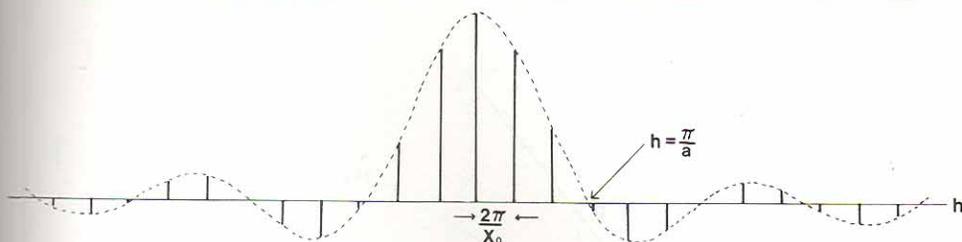


FIG. 13. Plot of Fourier transform of repetitive pattern.

amplitude, and a phase relative to other waves. At any instant the electric field E varies sinusoidally with distance as in Fig. 14. We call the argument of this sine function the *phase angle*. The origin from which this phase is calculated is really arbitrary, but it must be the same for all waves considered together. This function has the form

$$A(x) = A_0 \cos(2\pi x/\lambda) \quad (15)$$

The phase of this wave, relative to that at the origin, is $2\pi x/\lambda$. How might we represent a wave's amplitude and phase? One could simply use two numbers, the amplitude A_0 and the phase angle ϕ . These two numbers are easily graphed in polar coordinates such as in Fig. 15. Notice that the wave is then easily represented as a point on the graph. Another way to represent a point on a graph is in terms of its horizontal and vertical components A_r and A_i . Yet another is to represent it as a single complex number, having as its real component the horizontal distance to the point and as its imaginary component the vertical distance. This ability to use a single complex number to represent a wave's amplitude and phase is especially useful because there are several interchangeable ways to represent the complex number. The following are all equivalent, with ε being complex and the A values and ϕ being real.

$$\varepsilon = A_r + iA_i = A_0 \exp(i\phi) \quad (16)$$

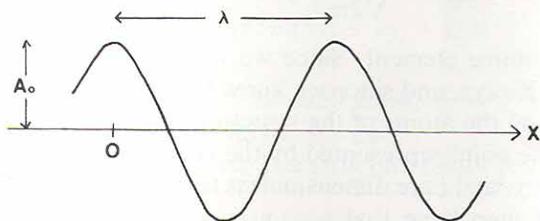


FIG. 14. Cosine function of electric field versus distance.

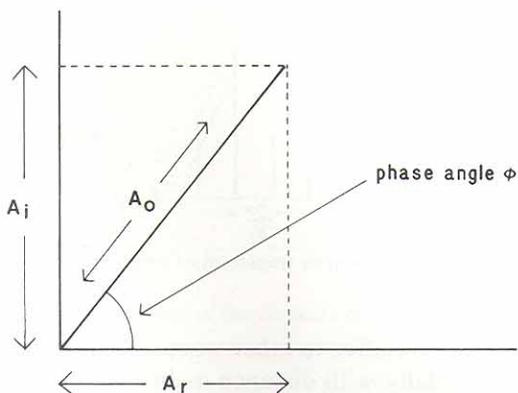


FIG. 15. Plot of amplitude (A) versus phase angle (ϕ).

Another value of this notation is that we can represent the interference among waves as the sum of the complex numbers which describe those waves. As an example, let us evaluate the result of interference between three waves with arbitrary amplitudes and phases. We can see in Fig. 16 that whether we sum the three waves, the vectors, or the complex numbers we arrive at the same result.

We now want to calculate an expression, which we shall call the structure factor, that represents the wave reflected from a single set of Bragg planes in a crystal. We know from the discussion above that this structure factor is the value of the Fourier transform of a single unit cell, evaluated at one of the sampling points that arise from the crystal repetition. These sampling points are reciprocal lattice points.

How do we evaluate the structure factor? Let us extend to three dimensions the expression we defined earlier for the one-dimensional Fourier transform, and at the same time introduce some crystallographic notation. The Fourier transform of a single unit cell is

$$F(\mathbf{s}_{hkl}) = \frac{1}{\sqrt{2\pi}} \int_v \rho(\mathbf{r}) \exp(2\pi i \mathbf{s}_{hkl} \cdot \mathbf{r}) d\mathbf{r} \quad (17)$$

where $d\mathbf{r}$ is volume element. Since we want this transform to represent diffraction of X rays, and since we know that the X rays are scattered by the electrons on the atoms of the structure, we let $\rho(\mathbf{r})$ be the density of electrons at the point represented by the vector \mathbf{r} . We define \mathbf{r} as $x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$, where x , y , and z are dimensionless fractional coordinates within the unit cell. We then have that $\mathbf{s}_{hkl} \cdot \mathbf{r} = hx + ky + lz$, a result of the properties of the real and reciprocal lattice vectors. Finally we notice that

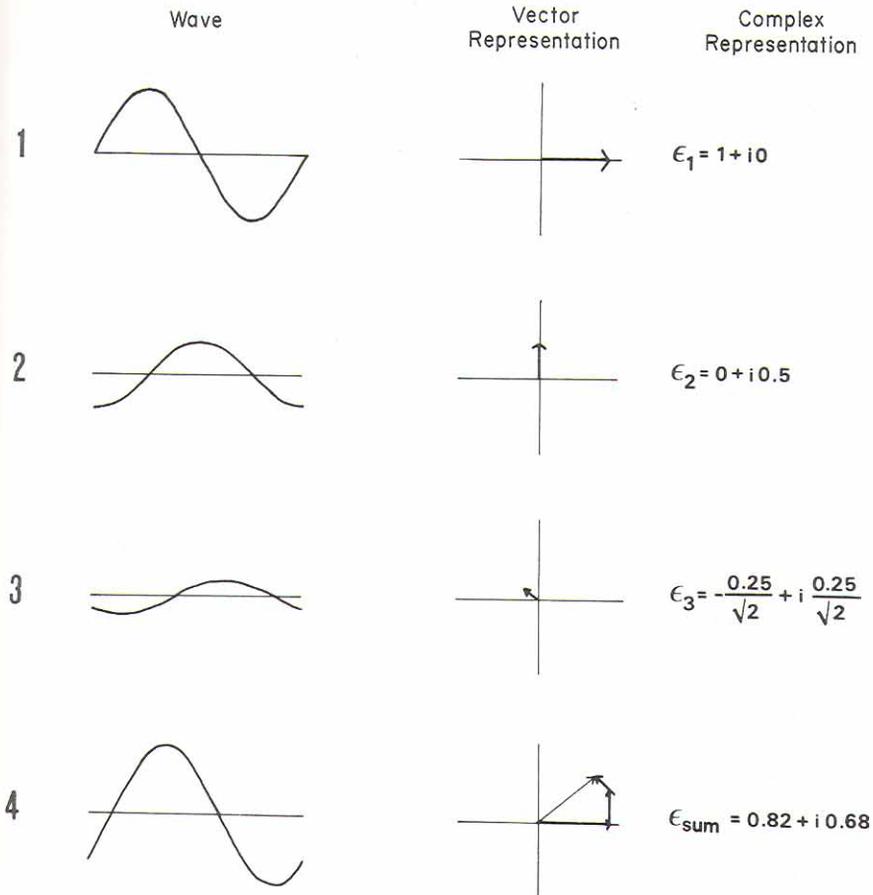


FIG. 16. Representation of result of interference between waves.

we can obtain the integral over the entire unit cell by simply performing a summation over the atoms in that cell.

This gives us the common expression for the structure factor:

$$F_{hkl} = \sum_{\text{atoms}} f_j \exp[2\pi i(hx_j + ky_j + lz_j)] \quad (18)$$

where f_j is the "scattering power" of each atom, x_j , y_j , z_j are its coordinates, and we have dropped the factor $1/\sqrt{2\pi}$.

We can show that this is correct. The structure factor for a single atom is

$$f_{hkl} = f_j \exp[2\pi i(hx_j + ky_j + lz_j)]$$

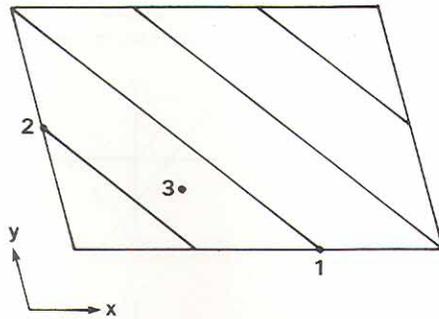


FIG. 17. Two-dimensional unit cell.

where the phase angle is $2\pi(hx_j + ky_j + lz_j)$. Does this make sense? We can place atoms in a unit cell with the lattice planes (hkl) drawn in and see if the phases calculated by this expression are correct. In Fig. 17 we have a two-dimensional unit cell with the $(3,2)$ planes in place and several atomic positions marked. Bragg's law requires that for diffraction to occur, the phases of the waves scattered from any two atoms must be equal when both lie on any one of the lattice planes in question. For example, we calculate that the phase of scattering from atom 1 at $(2/3, 0)$ is $2\pi(3 \cdot 2/3 + 2 \cdot 0) = 4\pi = 0$, and that for atom 2 at $(0, 1/2)$ is $2\pi(3 \cdot 0 + 2 \cdot 1/2) = 2\pi = 0$. We can see that the phase of scattering from atom 3, which lies midway between two planes at $(1/3, 1/4)$, should be π . This turns out to be the case since $2\pi(3 \cdot 1/3 + 2 \cdot 1/4) = 3\pi = \pi$.

Coming to Focus; Regenerating the Image. Recall what we have learned. Starting with very simple physical ideas, the scattering of X rays by atoms and the interference of scattered rays, we have built up a rather complete picture of the way X rays are diffracted from crystals. We understand the geometry of diffraction and can even calculate the amplitude and phase of the diffracted rays. What remains is to see how this information might be used to reconstruct the structure of the crystal.

Let us take our cue from Fourier. He showed that any periodic function can be approximated by a sum of trigonometric functions. For example, Fig. 18 shows a function that is periodic with a repeat distance of a . The "Fourier sum" which could approximate this function is

$$g(x) = \sum_{j=0}^n \left(A_j \cos 2\pi j \frac{x}{a} + B_j \sin 2\pi j \frac{y}{a} \right) \quad (19)$$

The coefficients A_j and B_j are real. This function $g(x)$ will come closer and closer to the true shape of $f(x)$ as n , the number of terms included in the

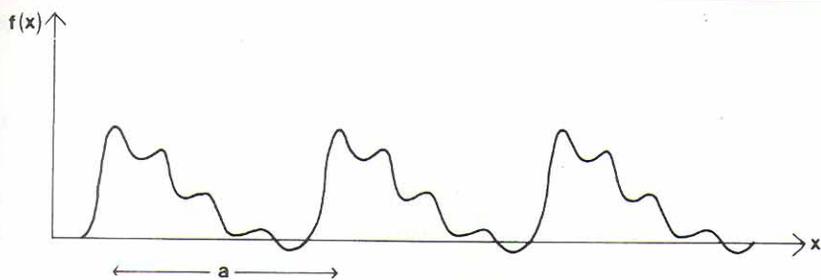


FIG. 18. Periodic function.

sum, gets larger. In Fig. 19 you can see that this is so. The waves represented on the left are shown summed along the right. The more waves used, the more nearly the sum approaches the shape of the periodic function in Fig. 18.

We have a periodic function to represent: the electron density in a crystal. It is periodic in three dimensions over distances that are the lengths of the principal axes of the crystal unit cell. Let us write a Fourier sum that could approximate it. Here we shall use a slightly different form of the Fourier sum. We shall write the function in three dimensions, use a complex exponential to represent the trigonometrics, allow the coefficients

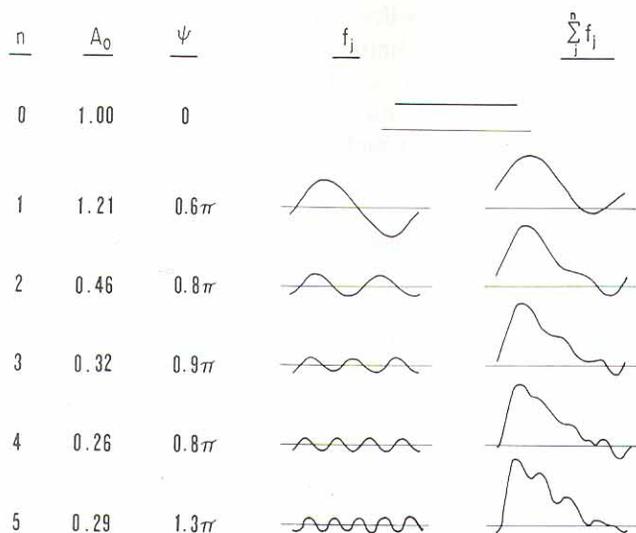


FIG. 19. Summation of waves approaches the shape of the periodic function.

to be complex, and finally sum over all negative and positive values of the indices. The function we want is

$$\rho(x,y,z) = \sum_{h'=-\infty}^{\infty} \sum_{k'} \sum_{l'} C_{h'k'l'} \exp[-2\pi i(h'x + k'y + l'z)] \quad (20)$$

We can evaluate the complex coefficients $C_{h'k'l'}$ by use of a standard mathematical trick. We start by rewriting the structure factor, in a slightly different but recognizable way:

$$F_{hkl} = \int_V \rho(x,y,z) \exp[+2\pi i(hx + ky + lz)] dV \quad (21)$$

Now we substitute Eq. (20), the Fourier sum for electron density $\rho(x,y,z)$, into this expression and simplify the result. What we find is that the complex coefficients $C_{hkl} = F_{hkl}/(\text{volume of unit cell})$. This then gives us the correspondence between the complex structure factor and the real electron density in the crystal, the Fourier electron density function:

$$\rho(x,y,z) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_k \sum_l F_{hkl} \exp[-2\pi i(hx + ky + lz)] \quad (22)$$

One can compare these Eqs. (21) and (22) to Eqs. (2) and (3), written during our first mention of Fourier, and see that the diffracted waves and the structure are Fourier transforms of one another.

Calculation of electron density is a simple computational chore; mathematicians have devised a technique called the "fast" Fourier transform in which the summation above is factored in a way that decreases substantially the number of calculations to be made. Something missing from our discussion, however, is a method for determining the phase of the complex structure factor F_{hkl} .

Phase Calculation: The Isomorphous Replacement Method. One of the most important contributions Perutz and his co-workers made to protein structure determination was to develop this method. It depends upon the slight perturbation that is caused by a few very heavy atoms being bound to the protein. When the binding of the heavy atoms causes no substantial changes to the crystal structure, the structure is said to be *isomorphous*, and the crystal with heavy atom bound is an *isomorphous heavy-atom derivative* of the native crystal.

Let us preview the way the method works before we examine the details. First one measures diffraction data for the native protein and for one or more isomorphous heavy-atom derivatives. Then one determines positions of the heavy atoms in the crystal. Doing this depends upon the fact that a hypothetical diffraction pattern from the structure that contains

the heavy atoms alone is very similar to the *differences* between the diffraction patterns from a native protein crystal and its heavy-atom derivative. As a result, many of the same techniques that are used to solve structures with only a few atoms can be used to locate the heavy atoms in a protein structure; only here the structure factors used are the differences between the two sets of measured structure factors. These techniques usually include the use of the Patterson function, a calculation that results in a knowledge of the *vectors* between atoms in the structure. Determination of the positions of several atoms from knowledge only of the vectors between them is often a problem of exquisite complexity. Other techniques that are sometimes used are the "direct methods" of structure determination. Here the statistical relationships among the amplitudes of structure factors can be used to place constraints on their phases. When the methods are applied to the differences between diffraction patterns, the result is the structure of the constellation of heavy atoms.

After the positions of the heavy atoms have been found, these positions are used in the calculation of structure factors for the heavy atoms alone. As you will see, these structure factors place constraints on the possible values of the phases for the native structure factors.

Recall that the structure factor is a sum of terms, one for each atom in the structure. If the protein atoms in the crystal are not perturbed by the binding of heavy atoms, the structure factor for a heavy-atom derivative of a protein crystal is simply that of the native protein with that for the heavy atoms added on:

$$F_{PH} = F_P + f_H \quad (23)$$

Here F_{PH} is the structure factor for the heavy-atom derivative of the parent protein, F_P is that for the parent, and f_H is that for the heavy atoms alone. This equation involving complex numbers represents a triangle in the complex plane (see Fig. 20). Of course, one does not know the phase of F_P or F_{PH} , but only the amplitudes $|F_P|$ and $|F_{PH}|$ and the phase and amplitude of f_H . As a result, this triangle can be drawn in two possible ways consistent with the data, as in Fig. 21, leaving one with a 2-fold ambiguity for the phase of F_P .

One normally resolves this ambiguity by performing the whole set of measurements and calculations again for at least one more heavy-atom derivative. One hopes that a second derivative will have phase indications close to only one of those from the first. Statistical methods, worked out by Blow and Crick, are used to compare the various indications for the phase and to choose one which will minimize the errors in the final electron density map.

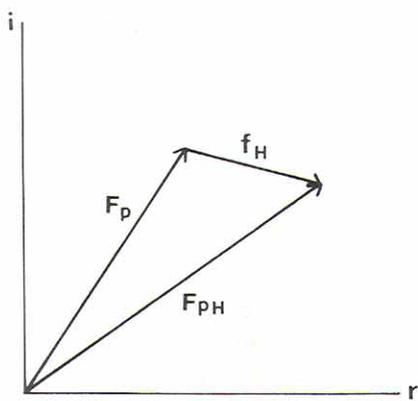
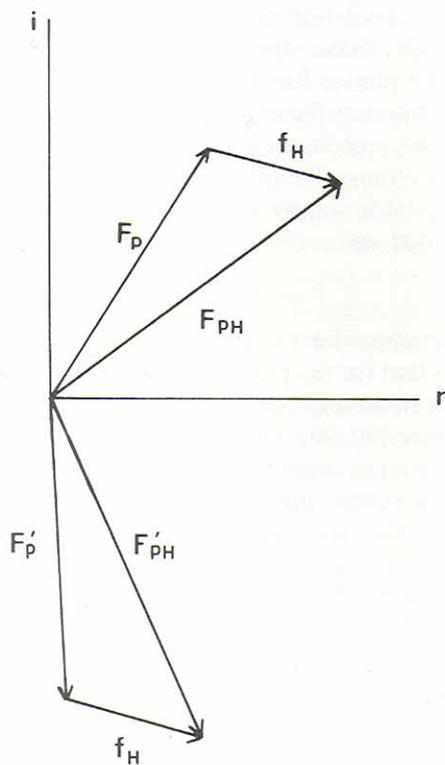


FIG. 20. Plot of the structure factor.

FIG. 21. Plot of 2-fold ambiguity for the phase of F_p .

Practicing Crystallography

From the biochemist's perspective, the crystallographer's chore must seem a long and complicated process. Acknowledging that this is true (which I do) does not make it simpler. It will pay, however, to show that a complete project is a serial sequence of steps whose only dependence upon each other is that they be performed carefully and in the proper order. Very briefly, these steps are

1. Growth of large, perfect crystals and preliminary characterization of the crystals and their diffraction pattern.
2. Preparation of heavy-atom derivatives.
3. Measurement and processing of diffraction data.
4. Calculation of phases for Fourier coefficients.
5. Interpretation of electron density maps and refinement of molecular models.
6. Analysis of the structure.

These items all are treated in the chapters that follow in this volume. Let us preview briefly what we shall find.

Crystal Growth and Characterization. In modern times one has access to a sophisticated arsenal of protein purification techniques, such as ion-exchange and gel-exclusion chromatography plus preparative-scale gel electrophoresis and isoelectric focusing. One forgets that a major purification tool of the early protein chemists, many of whom started life as organic chemists, was crystallization. Even now, some primeval instinct provides a thrill of satisfaction to the modern biochemist when he sees the opalescent sheen of microscopic crystals in a swirled flask of precipitated protein.

The conditions for growth of large crystals for X-ray diffraction are little different from those that produce these "biochemist's" crystals. One need only provide a pure protein and add to it, under the proper conditions, a suitable precipitant and a pinch of patience, and crystals may form. Much of the crystallographer's burden is to find these proper conditions and the proper precipitant. One must choose the correct pH, temperature, and protein concentration. One may select precipitants such as salts, which cause protein to crystallize because of hydrophobic interactions, or organic liquids, such as alcohols, which strengthen electrostatic interactions between protein molecules. In addition, one may need to worry about special conditions, such as the presence of allosteric effectors, that will favor one particular conformation for the molecule.

Section II of this volume discusses many of these considerations. As a testament to the parsimony of many crystallographers, who would rather

spend their time thinking about molecular structure than preparing protein, many of the chapters in that section concern methods to grow crystals with the smallest possible quantity of material.

Once crystals have been grown, they require a rather thorough analysis before one can embark upon a full structure determination. Several things must be learned.

1. What is the size of the crystallographic unit cell? What is its symmetry? Both of these are determined by scrutiny and measurement of the diffraction pattern.

2. What form does the molecule adopt in the crystal? How many molecules occupy each unit cell? A number of factors will help to answer these questions. Among them are knowledge of the crystal density, determination of which is discussed in Section III, and knowledge of the size and symmetry of the unit cell and of possible subunit structure in the molecule.

3. How accurately may one hope to learn the structure? Is this a tractable problem? These questions depend upon the quality of the diffraction pattern and upon one's ability to measure the diffraction data in practice. They also depend upon the stability of the crystals in the X-ray beam.

Preparation of Heavy-Atom Derivatives. If crystal growth is an art, this task involves alchemy. Although there are inspired exceptions, variations are usually played on only a few themes. Heavy metals may be bound to the protein before crystals are grown, but more often the heavy-metal compounds are allowed to diffuse into the crystals from solutions in which the crystals are soaked. A number of compounds, for example many containing Pb or Hg, bind tightly to the free thiol on cysteine. Iodine will react with tyrosine. Beyond this, one can find a sizable grab-bag of coordination compounds of Pt, Au, U, etc., that fix themselves to polar or ionic sites on proteins, and even a few, such as dimethyl mercury, that bind in hydrophobic pockets. The process is very tedious, and occasionally a little art, say in the design of a heavy-atom-labeled mimic of a substrate, will save a lot of alchemy.

Measurement of Diffraction Data. As we have mentioned in the first half of this chapter, one may measure diffraction intensities by photographic means, a process that has changed little over decades, or by use of electronic detectors, a field that promises to grow rapidly in the future.

As with all physical measurements, great care must be taken in the collection of X-ray data. One must minimize systematic and random error in the measurements. One must think hard about sources of systematic

error—nonuniformity in the detector or X-ray source, absorption of X rays by the specimen, or decay of the specimen in the X-ray beam—and must take steps to eliminate them. Replicate measurements must be made, both to monitor systematic error and to decrease random error. Systematic errors must be corrected if they can be measured. Often this can be accomplished for absorption of X rays and for crystal decay.

The chapters in Section III concern themselves with many aspects of data collection. In them we see both the norm of modern practice and the state of the art as it will stand in the future.

Calculation of Phases for Fourier Coefficients. Before the electron density in a crystal can be calculated, phases must be assigned to the structure factor amplitudes that have been measured. As we mentioned earlier, the most generally successful method available for doing this is *multiple isomorphous replacement*. Here a small number of heavy atoms are bound to each protein molecule in the crystal, perturbing the intensities and phases of the diffracted waves. If the positions of the heavy atoms are known, so that the heavy-atom contribution to the overall diffraction can be calculated, information about the phase can be determined.

Although the method of isomorphous replacement often serves for the initial calculation of phases, one sometimes can use other methods, such as those discussed in Part B, Vol. 115, to improve the initial estimates. This is possible when extra information is available about the structure of the protein in the crystal.

One example of “extra information” is when the molecule possesses symmetry that is not part of the crystal symmetry, that is, when the asymmetric unit of the crystal contains more than one identical piece of protein. In this case, identical but independent portions of the structure can be averaged to produce a new and more accurate electron density map. This is the basis for a powerful technique for phase improvement known as the *molecular replacement method*, described in a collection of papers by that name assembled by Michael Rossmann. It involves several separate steps. First the electron density map is averaged according to the highest possible molecular symmetry. Next, the envelope defining the surface of the molecule is determined by inspection. Third, the regions of density outside this envelope are all set to some average value to represent the interstitial liquid. Finally, this averaged electron density map with smoothed solvent regions is used for calculation of structure factors and a new set of phases is compiled, based on a comparison of the calculated phase and the one from isomorphous replacement. These new phases are used to calculate a new electron density map, and the process is repeated iteratively. The technique is especially powerful when the number of symmetry-related pieces in the asymmetric unit is large, and it

has been used with great success in work on the icosahedral viruses and on the coat protein of tobacco mosaic virus.

A variation on this theme is a technique known as *density modification*. Here, in the absence of multiple subunits in the asymmetric unit, one makes whatever improvements seem justified on the electron density and uses this modified density to calculate new phases. The modifications are based on reasonable assumptions about protein crystals. Two assumptions that can be made are first that the electron density in solvent regions of the crystal is fairly smooth, and second that there is an absolute minimum below which the electron density may not go. After the electron density is modified to meet these conditions, structure factors are calculated, calculated phases are combined with those from isomorphous replacement, a new map is calculated, and the process is repeated. Especially in cases where the fraction of the crystal volume occupied by solvent is very high, say with tRNA, the process provides a marked improvement in the structure.

Production of Molecular Models. The result of the diffraction experiment is the three-dimensional map of electron density in the crystal. This, map, however, serves merely as *data* to be used to interpret the molecular structure. One must build an atomic model to fit the electron density. Although efforts are being made to automate this process, it often depends greatly upon the chemical intuition of the person doing the work. A fundamental principle of the model building is that one really knows quite a lot about the structure of a protein, particularly if the primary structure is known. In particular, there are fairly strict constraints placed on covalent bond lengths and angles. In addition, although the constraints are not so strong, simple stereochemical arguments make some torsional configurations more likely than others.

One makes use of this extra knowledge as much as possible. Kendrew and Watson invented rigid brass-wire models, which accounted for as many of these covalent constraints as possible, for the model-building work on myoglobin. Modern crystallographers usually use computer graphics to assist in the model-building chore, but again, the known covalent dimensions of amino acid residues are built into the programs that are used.

Because the initial calculation of structure factor phases is inaccurate, the electron density map, hence the model that can be built from it, is also inaccurate. It is always true that more information exists in the original diffraction data than can be represented by a molecular model that fits an electron density map phased by multiple isomorphous replacement. A variety of methods exists to improve this model. These include (1) classical "difference Fourier" techniques, (2) optimization of the model by a least-

squares minimization of the difference between observations and the structure factors calculated from the model, and (3) a minimization of the energy of noncovalent interactions in the molecule, a procedure that has nothing to do with the X-ray data. In each case, the only truly unbiased test of the quality of the model is the accuracy with which structure factors calculated from it match the observations.

Analysis of the Structure. There are several senses in which one might "analyze" the structure of a protein molecule. We shall discuss only two.

Crystal structure work on an enzyme is initiated to learn about the structural basis of its biological activity. Knowing the three-dimensional structure of an enzyme is not unlike seeing a lathe or a mill all cleaned up and sitting, silent, on the shop room floor. With a little imagination, one can make pretty good presumptions about how it works. It is not the same, however, as seeing it in action. It is also not the same as seeing the same machine, again silent, with the work and tools in place. To retrieve this metaphor, the crystallographer sees an enzyme in stasis; the active machine is not accessible, although recent work in low-temperature crystal structure analysis shows how the machines can be slowed down a lot. However, it is often possible to catch the enzyme at one end of a process. Specifically, the biochemist can often react the protein in the crystal with cofactors, competitive inhibitors, substrate mimics that bind to an active site, etc. When this has been done, a procedure known as difference electron density synthesis can be used to learn about the complex. Diffraction data are measured from the derivatized crystals, and amplitudes that are the signed difference between the derivative structure factors and the native protein structure factors are used with the native crystal phases to calculate an electron density map. This map closely approximates the difference between the electron density in the derivative and the native crystals. It can often be used to determine the structure of the pseudosubstrate or effector as they are bound to the enzyme and this knowledge often can be used to learn much about the action of the enzyme.

A second way in which one may analyze the structure of a protein molecule is that one can compare it, in a topological sense, with the known structures of other proteins. This is something that has only begun to be possible in the last half-dozen years, after a large number of structures had become known. Some of the ideas that have evolved during this time are reviewed in the following companion volume.

A number of systematic features of protein structure have been noticed. The most often recurring of these is that a rather small number of topologies of folding of the β -pleated sheet structure are observed at a surprisingly high frequency. Another is that at least one rather specific arrangement of peptide chains, the nucleotide-binding fold found in sev-

eral dehydrogenases, has been found on several occasions to perform a similar function in different proteins.

In the final analysis, the structures themselves, not the X-ray data, are grist for the enzymologist's mill. From these structures we are beginning to understand the way in which enzymes catalyze and control reactions, and to glimpse some of the principles upon which all of molecular, hence cellular, structures are based. A natural consequence of this new knowledge is that our questions become ever more sophisticated and our curiosity about larger and more complicated structures continues to grow. The message in this for the crystallographer is that the success of his methods leads to new demands. Whatever he can discover now will make crucial his ability to discover much more in the future. Although he may not rest on his laurels, he can labor in secure self confidence that the explosion in his capabilities which is occurring now, and which is chronicled in these two volumes, not only will persist, but will continue to be of tremendous value as the future unfolds.