

## CHAPTER 1

---

# AN OVERVIEW OF MACROMOLECULAR CRYSTALLOGRAPHY


---

The only technique that allows direct visualization of protein structure at the atomic, or near atomic level is X-ray diffraction analysis as applied to single crystals of pure proteins. The technique has been applied to conventional small molecules now for over 80 years with extraordinary success, and very few chemical structures of less than a hundred atoms, obtainable in the crystalline state, have proven refractory. The successful application of X-ray diffraction to protein structure is relatively new, the first protein structure, that of myoglobin, having only been solved in 1960 (Dickerson, Kendrew, and Strandberg, 1961; Dickerson, 1991). Since that time nearly 50,000 additional protein structures have been added to our data base (Berman et. al., 2000), but even this collection represents only a very small fraction of the hundreds of thousands of different protein molecules that play some role in living processes. Thus the determination of protein structure by X-ray crystallography occupies the energy of several hundred laboratories in the world, and this number is ever growing as the need for more and increasingly precise structural information expands in step with the molecular biological revolution.

### WHAT DO WE MEAN BY THE STRUCTURE OF SOMETHING?

In common language when we ask, “what is its structure?” we mean by that, how are the various components or elements that make up the object disposed, or placed, with respect to one another in three-dimensional space. More simply, “what does the arrangement look like?”

While seeming a straightforward question, it is one that has perplexed scientists, philosophers, and poets for centuries. Answers have been formulated, for example, by homology—

for example, “a rock, a craig, nay a peninsula<sup>1</sup>”—by describing the physical qualities of the object—for example, “it was a one eyed, one horned, flying, purple people eater<sup>2</sup>”—by analytical expressions<sup>3</sup>—for example,  $r = a\theta^2$ —by visual illustration<sup>4</sup>—for example,  and undoubtedly by other means as well.

However, in proper scientific terms there is only one way to precisely describe the structure of an object, be it simple, or intricate and complex. That is by specifying, as in Figure 1.1, the coordinates in three-dimensional space of each point within the object, each with respect to some defined and agreed-upon system of axes in space, namely a coordinate system. Generally, the system is chosen to be an orthogonal, Cartesian coordinate system, but it need not be. It may be nonorthogonal, cylindrical, spherical, or any number of other systems.

The object’s inherent structure, being fixed, remains the same no matter how the spatial coordinate system is chosen, or where its origin is taken to be. Because the structure is invariant, even if its constituent points are transformed from one coordinate system to another, the relative positions of the points within the object remain the same. The structure is not dependent on the coordinate system we choose. Thus, if the structure of a molecule is defined by specifying the coordinates in space  $x_j, y_j, z_j$  of each atom  $j$  in the molecule, atom 5 (or 7 or 18, or whatever) maintains the same relationship in space to atom 14 (or 3, or whatever) no matter what the coordinate system.

To define the structure of a molecule in a precise manner then, we must create a list (the order is not important) of atomic coordinates  $x_j, y_j, z_j$  (and here the order is important). A molecular structure becomes a set of ordered triples  $x_j, y_j, z_j$ , one for every atom. This is imminently suitable not only for translation into a visual representation, but for manipulation and analysis in a computer, and presentation on the screen of a computer graphics workstation in any number of manifestations.

When we solve the structure of a molecule, any kind of molecule, including proteins, nucleic acids, or even large assemblies such as viruses or ribosomes, in the end we seek to identify and specify the  $x_j, y_j, z_j$  coordinates of every atom in the molecule. The form, the shape, the image, must always first be defined in these simple numerical terms, as ordered triplets. Only from these can we faithfully reproduce the precise structure of the molecule in more familiar visual terms, as pictures or images.

## AN ANALOGY

Let us assume that we want to determine the structure, as defined above, of some object that is invisible. It has the supernatural property that it is nonresponsive to any electromagnetic radiation such as light. But, to make the example more concrete, let’s assume it is an invisible, yellow, 1963 Volkswagen beetle, like that shown in Figure 1.2. If we have never seen such a glorious object before, how can we learn of its structure? How can we visualize it?

One way we might approach this problem is to take advantage of the fact that although the Volkswagen does not reflect light, it retains all of its other physical properties. We might, for example, take a basketball and throw it at the invisible object from some direction  $\vec{k}_0$

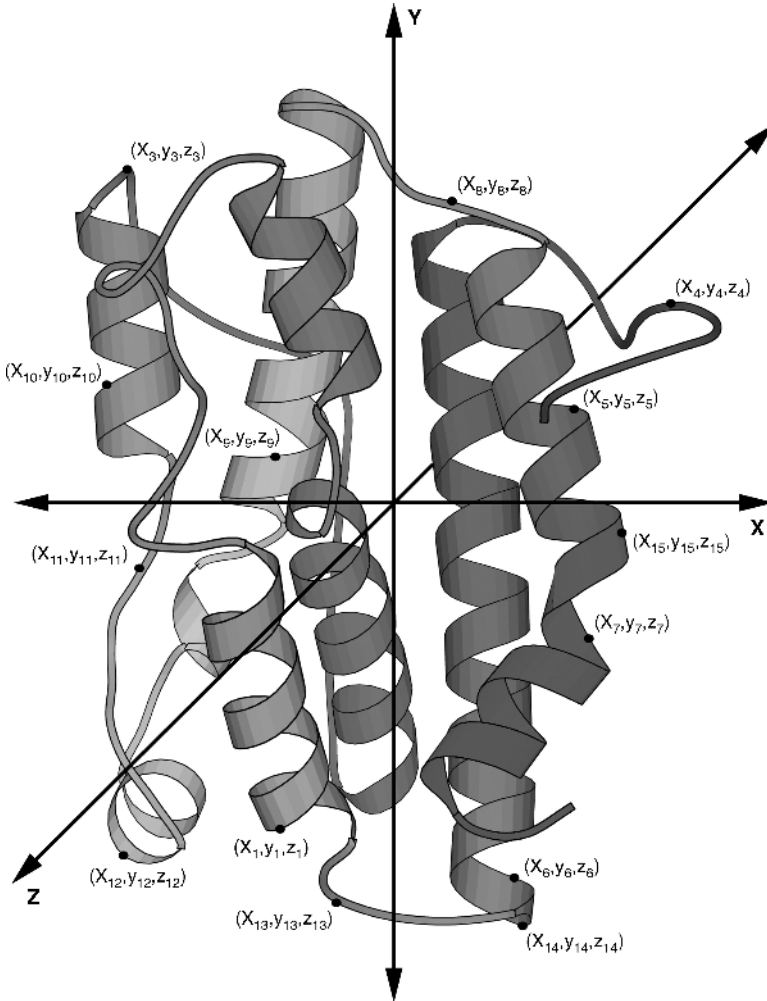
<sup>1</sup> Cyrano de Bergerac describing his own nose.

<sup>2</sup> From the song *The Purple People Eater* by Sheb Wooley.

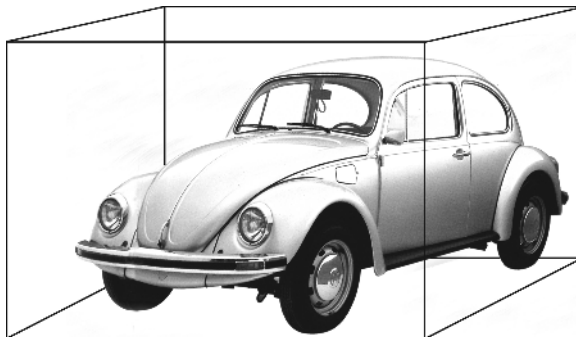
<sup>3</sup> The equation for the logarithmic spiral of a seashell.

<sup>4</sup> University of California, Irvine Anteater (official mascot).

(a vector, which has direction, is defined by some character or symbol with a line over the top), and note which direction  $\vec{k}$  it bounces off the object. More informatively, we might throw a hundred, or a thousand balls at the invisible Volkswagen and note how many balls bounce in all directions  $\vec{k}$ . Some will hit the fender, others the hood, others the windshield, and so forth, and, depending on the orientation of the car with respect to the balls thrown along  $\vec{k}_0$ , some directions  $\vec{k}$  for the reflected balls will be much favored over others. If the



**FIGURE 1.1** The structure of an object, such as the protein molecule shown here in caricature form, can be described in precise, quantitative terms by specifying every point in the object by a set of three coordinates  $x$ ,  $y$ ,  $z$ . The origin and orientation of the particular coordinate system is arbitrary, and is usually chosen for convenience or economy. Coordinates specified according to one linear coordinate system can always be transformed into another without alteration of the relative positions of the points making up the object. The entire structure of the object is a list of ordered triples  $x$ ,  $y$ ,  $z$ . For a molecule, the  $x$ ,  $y$ ,  $z$  coordinates are assigned to the atoms making up the structure. In X-ray crystallography, when we determine the structure of a molecule, we mean that we are determining the numerical values of the individual  $x$ ,  $y$ ,  $z$  coordinates that comprise the list of atoms.



**FIGURE 1.2** A “continuous object,” here an imagined 1963 Volkswagen Beetle regally aligned in a defined parking location in three-dimensional space, with the supple lines amenable to description in terms of bouncing balls.

direction  $\vec{k}_0$  corresponds to one aiming directly at the front of the car, for example, balls bouncing off the hood and windshield will be strongly favored.

Lets assume, however, that we can walk around the invisible Volkswagen and throw the basketballs from many, in fact all possible directions  $\vec{k}_0$ . Each time we note carefully how many balls bounce in which direction. Ultimately, we will know for every direction of our incoming beam of basketballs  $\vec{k}_0$  how many are reflected in every direction  $\vec{k}$ .

This ensemble of observations  $\vec{k}_0, \vec{k}$  and the number, or intensity, of balls  $I$  in the direction  $\vec{k}$  contains information about the structure of the invisible object, the orientations of its various external planes (doors, windows, hoods, fenders, etc.) from which the balls bounce. Now the question is, can we, from the observations, synthesize the shape of the object that gave rise to the pattern of reflected basketballs? The answer is, of course, yes. Mathematical procedures do indeed exist for extracting the shape of a 1963 Volkswagen beetle (see Figure 1.2) from a scattering pattern of basketballs. We might even invent some analogue device that we could place in a manner that it could accumulate automatically the reflected balls and somehow translate the pattern into an image of the object. We would call such a device a lens.

Now basketballs are rather large objects (probes), and when they bounce from a surface plane, they are rather insensitive to its finer details such as windshield wipers, door handles, and bolt heads. We could, however, make our investigation more sensitive by using, instead of basketballs, tennis balls, and even more sensitive still by using ping-pong balls, or even marbles (no, lets not use marbles as that would damage the paint job). Then the directions in which our probes bounce would more closely reflect the undulations of the hood, and the presence of door handles. That is, we would obtain a more refined, higher resolution image.

The approach illustrated here is not exactly what is done in X-ray diffraction, but it is similar. For example, we don't learn anything about the shape of the engine because our various balls cannot penetrate the interior of the car, whereas X radiation can penetrate and reflect from the internal atoms of molecules. But in many other ways the approach is the same.

Let us alter our analogy a bit and now assert that the reason our 1963 Volkswagen beetle is invisible, is because it is too small to see. It is smaller than the wavelength of visible light. We could, in principle, carry out the same experiment of walking around the nanoscale Volkswagen and directing a probe at it from all directions  $\vec{k}_0$ , then noting in every case

what intensity of reflected probes  $I$  was observed for all directions  $\vec{k}$ . If the probe we used sometimes penetrated into the interior parts of the object (e.g., and struck the transmission) so much the better. Although our pattern of diffracted probes would be considerably more complex, we would then learn about the structures of things inside the car as well. As long as the size of the probe is comparable to the sizes of the molecular features we wish to see (a very important point), we could do as well as we did with basketballs and ping-pong balls.

In organic molecules, the distances between bonded atoms are usually 1 to 2 Å; hence the size of our probe must be comparable. The wavelength,  $\lambda$ , of X rays used in diffraction experiments are usually between 1 and 2 Å.  $\text{CuK}_\alpha$  radiation produced by most conventional laboratory sources, for example, is 1.54 Å wavelength.  $\lambda$  is exactly analogous to the probe size. The shorter  $\lambda$  is, the smaller the diameter of the ball we are using, and the greater is the detail we can resolve.

Now a single Volkswagen of molecular size, impressive though it might be, would be so very small that, in practice, it would be impossible to hit it with enough balls (or probes, or waves) from a particular direction  $\vec{k}_0$  and to measure the intensity of reflected waves,  $I$ , in a particular direction,  $\vec{k}$ . How could we amplify the effect so that we could measure it?

Consider an enormous parking lot full of identical 1963 Volkswagen beetles, all perfectly parked by their drivers so that every car is identically oriented and placed in exact order. That is, they form a vast periodic array of cars. If we now direct millions of basketballs at this Volkswagen array from the same direction  $\vec{k}_0$ , then every car, having identically the same disposition, would reflect the balls exactly the same way. The signal, or reflected pattern of probes, would be amplified by the number of cars in the parking lot, and the end result, which we call the signal, would be far more easily detected and measured because of its strength. In our diffraction experiment, which is what we are really doing here, the Volkswagen beetles are the molecules, the basketballs analogous to X radiation, and the numbers of basketballs scattered in each direction are the intensities of the diffracted waves. Instead of an automobile parking lot, we have a molecular parking lot, a crystal.

A single voice, in a coliseum, though shouting, cannot be heard at a distance. Even a stadium full of voices cannot be heard far away if each individual is shouting a different cheer at random times. But if every voice in the stadium (or at least those favoring one particular team) are united in time in a single mighty cheer, the sound echoes far and wide.

*Then from five thousand throats and more there rose a lusty yell;  
it rumbled through the valley, it rattled in the dell;  
it pounded through on the mountain and recoiled upon the flat;  
for Casey, mighty Casey, was advancing to the bat.<sup>5</sup>*

It is this cooperative effort of many individuals united in space and time, as occurs in a crystal scattering X rays, that makes a molecular diffraction experiment possible.

Clearly, by this analogy, we have simplified things to get at the essentials, but the details will come later. It is important to keep in mind that in carrying out our experiments on vast, ordered arrays of structurally unknown objects, we sacrifice information that might have been obtained from a single individual (i.e., there is no free lunch). In addition because our probes in X-ray crystallography are not particles, or balls, but are waves, an additional complication is introduced. This is because waves add together, or interfere with one another, in a manner unlike that of single particle probes. Thus we ultimately must consider

<sup>5</sup> From "Casey at the Bat," by Earnest Lawrence Thayer.

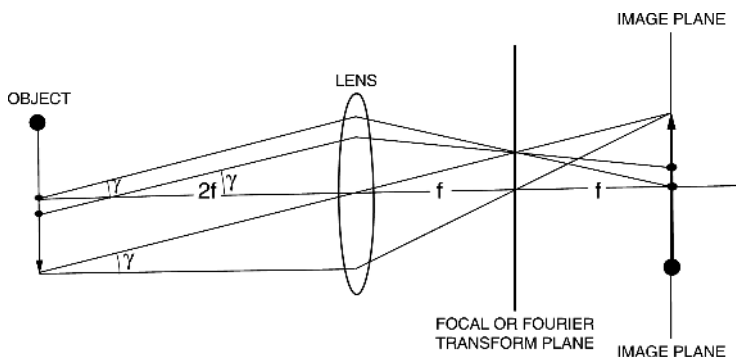
the diffraction pattern from our molecular array, or crystal, as sums of waves. Later this sacrifice of information will emerge as what is known as “the phase problem” in X-ray crystallography. That problem we will address in due course.

## A LENS AND OPTICAL DIFFRACTION PATTERNS

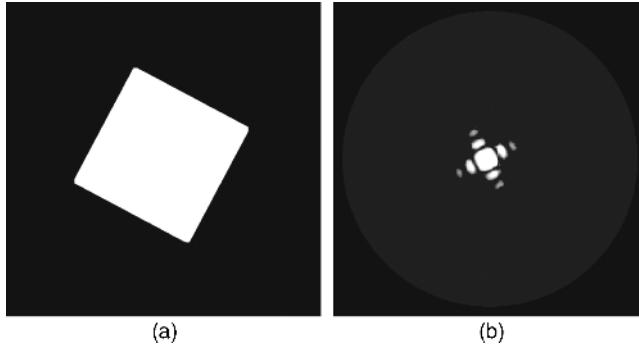
In our common experience, we rarely even think of waves and how they add together, even though we depend on light (wavelength  $\lambda = 3500 \text{ \AA}$  to  $6000 \text{ \AA}$ ) for visualizing nearly everything. We use our eyes, microscopes, telescopes, cameras, and other optical devices that depend on waves of light, yet we never, it seems, have to deal directly with waves or how they interact with one another. The reason is that we have lenses that gather together the light waves scattered by objects, and focus them into an image of the original object. The lenses of our eyes focuses light waves scattered by an object at a distance into an image of that object on our retinas. The lens of a microscope focuses the light scattered by a minute object in the path of a light beam into an image of the object, and magnifies it for us at the same time.

Figure 1.3 illustrates the essential features of image formation by a lens using a simple ray diagram. There are two unique planes where the rays emitted by the light-scattering object intersect after passage through the lens. One plane is twice the focal length of the lens ( $2f$ ). There, an inverted image of the object is formed by the summation of rays from discrete points on the object converging at corresponding points on the plane. The rays converge in a different manner, however, on a second plane at a distance  $f$  between the lens and the image plane. In that plane, rays intersect that do not originate at the same point on the scattering object, but that have the same direction (defined by the angle  $\gamma$ ) in leaving the object. The convergence of the various sets of rays, each having a different direction parameter  $\gamma$ , forms in this plane a second kind of image, which is called the diffraction pattern of the object. The diffraction pattern is known mathematically as the Fourier transform of the object.

What does the diffraction pattern of an object look like? We can visualize the diffraction pattern, the Fourier transform, of an object by making a mask about the object and then passing a collimated beam of light through the mask and onto a lens. The lens, as in



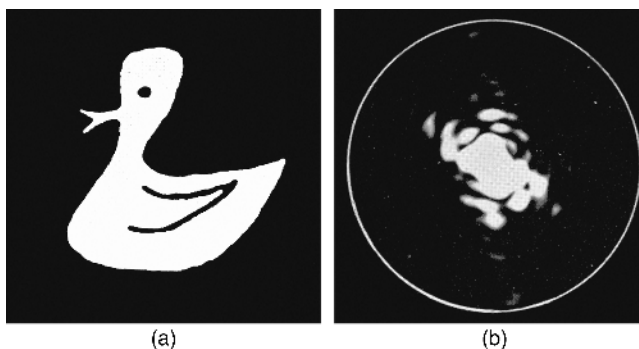
**FIGURE 1.3** Formation of the Fourier transform (or diffraction pattern) of an object by a lens having focal length  $f$ . The rays leaving the object are caused, by the refractive properties of the lens, to converge at both the image plane ( $2f$ ) and at a second “focal” plane  $f$ . The rays converging at each point on this transform plane at  $f$  are those that form a common angle with the plane of the scattering object, denoted here by  $\gamma$ ; that is, they have a common scattering direction.



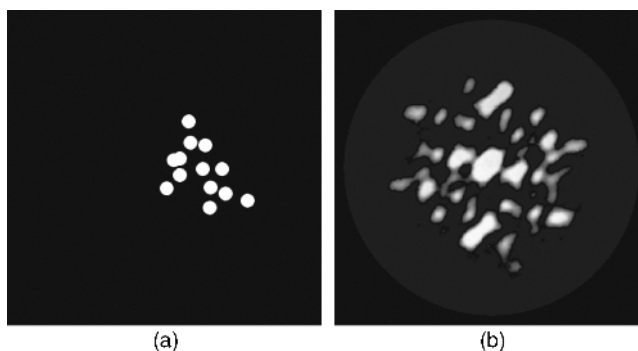
**FIGURE 1.4** If a mask containing the square in (a) has a parallel beam of light directed through it and onto a lens, as in Figure 1.3, then if a screen were placed at one focal length  $f$  behind the lens, the diffraction pattern in (b) would be observed. At twice the focal length behind the lens, an image of the original square would appear. The pattern of light and dark seen in (b) is both the optical diffraction pattern of the square, and, in mathematical terms, it is the Fourier transform of the square.

Figure 1.3, then creates the diffraction pattern at a distance  $f$ , which we can view on a screen, or record on a film.

Figure 1.4 is a simple example. The object is the square shown on the mask in (a). If we look at a distance  $f$  behind the lens, then we see the diffraction pattern of the square in (b). A second example is shown in Figure 1.5. With the possible exception of the diffraction pattern of DNA, this is probably the most reproduced diffraction pattern in the history of X-ray crystallography (see Taylor and Lipson, 1964, for its origin). Here, the object in (a) is a duck (probably rubber), and in (b) we see the duck's Fourier transform, its diffraction pattern. Significantly, if we were to place the diffraction pattern in (b) in the place of the object in (a), then at distance  $f$  behind the lens we would now see the duck. In other words, (b) is the Fourier transform of (a), but (a) is also the Fourier transform of (b). The transform is symmetrical, and it tells us that either side of the Fourier transform contains all the information necessary to recreate the other side.



**FIGURE 1.5** If the object mask contains the image of a duck, as in (a), its optical diffraction pattern, or Fourier transform, seen at  $f$ , is the seemingly meaningless pattern of light and dark in (b). It is important to note that the object, the duck, is a continuous object of basically arbitrary placed points (in a mathematical, not a biological sense), so the diffraction pattern is likewise a continuous function of intensity consisting of patches and islands of light. Any continuous object, such as the duck seen here, might be expected to yield a diffraction pattern having these characteristics.



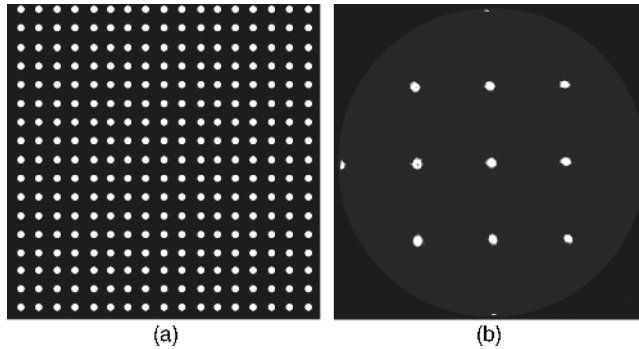
**FIGURE 1.6** In (a) the object for the optical transform is not a continuous object, but is a set of points arbitrarily distributed in space, as we might expect to find in a molecule made up of discrete atoms. That is, they bear no fixed mathematical relationship to one another. In (b) the optical diffraction pattern of the set of points is again a continuum consisting of islands of light and dark. Such a transform is typical of one we might expect from any conventional organic molecule. The locations of the light and dark areas in the transform are dependent only on the  $x$ ,  $y$  positions of the individual points in the object. If a point in the object were moved, the transform would change. If the entire set of points were rotated in the plane, the transform would undergo a corresponding rotation.

In the examples above, (a) was what we call a continuous object in that it was composed of a continuum of points covering a defined area, namely a square or the surface of a duck. The diffraction patterns were similarly continuous. Molecules, however, are not really continuous; they are composed of atoms, which serve as discrete scattering points. In Figure 1.6, for example, we have an arbitrary distribution of scattering points, like atoms in a molecule, and in (b) we see the diffraction pattern of the atom set. Note that even though the object is composed of unique scattering points, the diffraction pattern is still more or less continuous. Thus we should expect the diffraction pattern of a single molecule to be continuous, even if the molecule itself is not.

A final example, but of a different kind of object, is shown in Figure 1.7. This object is a discrete set of points distributed over the surface of a mask in a periodic (uniformly repetitive) array. We call such a periodic point array in space a lattice. In (b) is the diffraction pattern of the lattice in (a), and vice versa. The diffraction pattern in (b) is also a lattice composed of discrete points (it is what we call a discrete transform), but the spacings between the points are quite different than for the lattice in (a). We will see later that the distances between lattice points in (a) and (b) are reciprocals of one another. The Fourier transform of a lattice then is a reciprocal lattice.

We will further see in later chapters that it is possible to combine the two kinds of transforms illustrated here, the continuous transform of a molecule with the periodic, discrete transform of a lattice. In so doing, we will create the Fourier transform, the diffraction pattern of a crystal composed of individual molecules (sets of atoms) repeated in three-dimensional space according to a precise and periodic point lattice.

We can get some idea as to what to expect by again using optical diffraction. In Figure 1.8a is a pattern of scattering points having no internal symmetry or periodicity. It might well represent the set of atoms in a molecule. Its diffraction pattern is seen in Figure 1.8b. If the molecular motif in Figure 1.8a is repeated in a periodic manner in two dimensions, that is, according to a point lattice, then we can generate the array seen in Figure 1.8c. And what



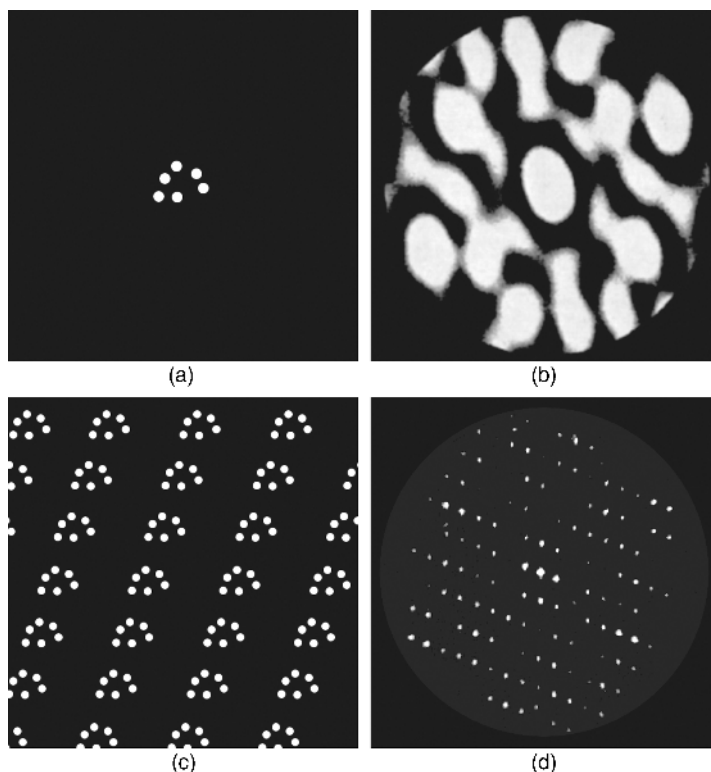
**FIGURE 1.7** In (a) the object, again exposed to a parallel beam of light, is not a continuous object or an arbitrary set of points in space, but is a two-dimensional periodic array of points. That is, the relative  $x$ ,  $y$  positions of the points are not arbitrary; they bear the same fixed, repetitive relationship to all others. One need only define a starting point and two translation vectors along the horizontal and vertical directions to generate the entire array. We call such an array a lattice. The periodicity of the points in the lattice is its crucial property, and as a consequence of the periodicity, its transform, or diffraction pattern in (b) is also a periodic array of discrete points (i.e., a lattice). Notice, however, that the spacings between the spots, or intensities, in the diffraction pattern are different than in the object. We will see that there is a reciprocal relationship between distances in object space (which we also call real space), and in diffraction space (which we also call Fourier space, or sometimes, reciprocal space).

is the diffraction pattern, the Fourier transform of the periodic distribution in Figure 1.8c? It is shown in Figure 1.8d.

The diffraction pattern of the molecular array in Figure 1.8d is also periodic, but the spacings between the diffraction intensities are reciprocals of the molecular point lattice. The intensities in Figure 1.8d vary from point to point as well, unlike the example shown in Figure 1.7. If the diffraction pattern in Figure 1.8d were superimposed on that in Figure 1.8b, we would find that the intensities at the discrete points in Figure 1.8d are identical to the intensities of the corresponding points in the continuous transform in Figure 1.8b, which they overlay. The discrete lattice according to which the molecules are periodically arrayed in Figure 1.8a has the effect of allowing us to see, or sample, the transform (diffraction pattern) of an individual molecule at periodic, specific points. The lattice appearing in diffraction space, having reciprocal spacings between points, is the corresponding reciprocal lattice.

To discriminate, or resolve, individual points in an object, as we saw in the Volkswagen parable, one must utilize a radiation of wavelength comparable to the distances between the scattering points. Thus we can use microscopy with light to resolve detail within an object that is on the order of a few thousand angstroms. We can use radio waves, as in radar, to resolve details measured in meters. If the objective is to produce an image of a macromolecule composed of atoms separated by an average bond length of about  $1.5 \text{ \AA}$ , then one is obligated to use a radiation of comparable wavelength. Conveniently, the characteristic X radiation produced by the collision of high-energy electrons with a number of different metal targets is of the range  $1$  to  $3 \text{ \AA}$ , precisely what is required. Less convenient is the unfortunate reality that nature has not provided us with any known lens or mechanism for the focusing of scattered X rays.

Unlike light, which, because of its refractive properties, can be focused by a properly ground glass lens, and unlike electrons, which, because of their charge, can be focused by



**FIGURE 1.8** In (a) is an arbitrary set of points that might represent the atoms in a molecule, and in (b) is the optical diffraction pattern of that set of points. It is a continuum of light and dark over the whole surface of the screen. The mask (object) in the optical diffraction experiment in (c) is the periodic arrangement of the fundamental set of points in (a) in two dimensions (i.e., the repetition of the object according to the instruction of a lattice). The diffraction pattern of (c) is shown in (d). We would find that if we superimpose the point array in (d) upon the continuous transform in (b), the intensity at each lattice point in (d) corresponds to the value of the continuous transform beneath. That is, the diffraction pattern in (d) samples the continuous transform in (b) at specific points determined by the periodic lattice of (c).

electromagnetic fields, X rays have no properties that permit an analogous process. Thus X radiation can be scattered from the electrons of an object, just as light or electrons are scattered by a specimen as they pass through it. But contrary to the situation we enjoy with a microscope, no lens can be interposed between specimen and observer to gather the scattered radiation and focus it into a meaningful image.

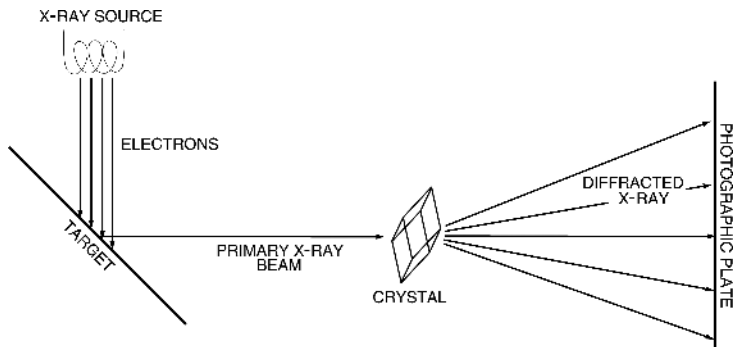
The crystal lattice, however, plays a second role. It not only amplifies the diffraction signal from individual molecules, it also serves as half a lens. The X rays scattered by the atoms in a crystal combine together, by virtue of the periodic distribution of their atomic sources, so that their final form is precisely the Fourier transform, that is, the diffraction pattern that we would ordinarily observe at  $f$  if we did in fact have an X-ray lens. Thus the situation is not intractable, only difficult. We find in X-ray crystallography that while we cannot record the image plane, we can record what appears at the diffraction plane. It is then up to us to figure out what is on the image plane from what we see on the diffraction plane.

## HOW X-RAY DIFFRACTION WORKS

If a collimated beam of monochromatic X rays is directed through an object, such as a macromolecule, the rays are scattered in all directions by the electrons of every atom in the object with a magnitude proportional to the size of its electron complement. This is the fundamental experiment that we perform in a diffraction experiment, and it is illustrated in Figure 1.9. If the object were composed of more or less arbitrarily placed atoms, as they are in a single macromolecule, then at any point in space about the isolated object a measurable amount of scattered radiation would be expected to be recorded by an observer; that is, the distribution of scattered rays would be a continuum of varying intensity. This was shown in Figures 1.4 through 1.6 and in Figure 1.8. The variability of intensity throughout this continuous scattering distribution, which is again the Fourier transform, or diffraction pattern of the macromolecule, would depend on the relative positional coordinates and atomic numbers of the atoms in the object, and would ostensibly be independent of any other property of the object. We will show this in Chapter 5.

If a number of identical objects were arranged in three-dimensional space in such a way that they form a periodically repeating array, the scattering distribution, or diffraction pattern from the collection of objects would tend to be less continuous, taking on observable values at some points and approaching zero elsewhere. This was illustrated in Figure 1.8. When the number of objects in the array becomes very large, as it does in a crystal, the scattering distribution, the diffraction pattern, becomes absolutely discrete.

Now, as discussed already, the scattering of X rays from a single protein or nucleic acid molecule would be immeasurably small. Due to its size alone, such an object could not

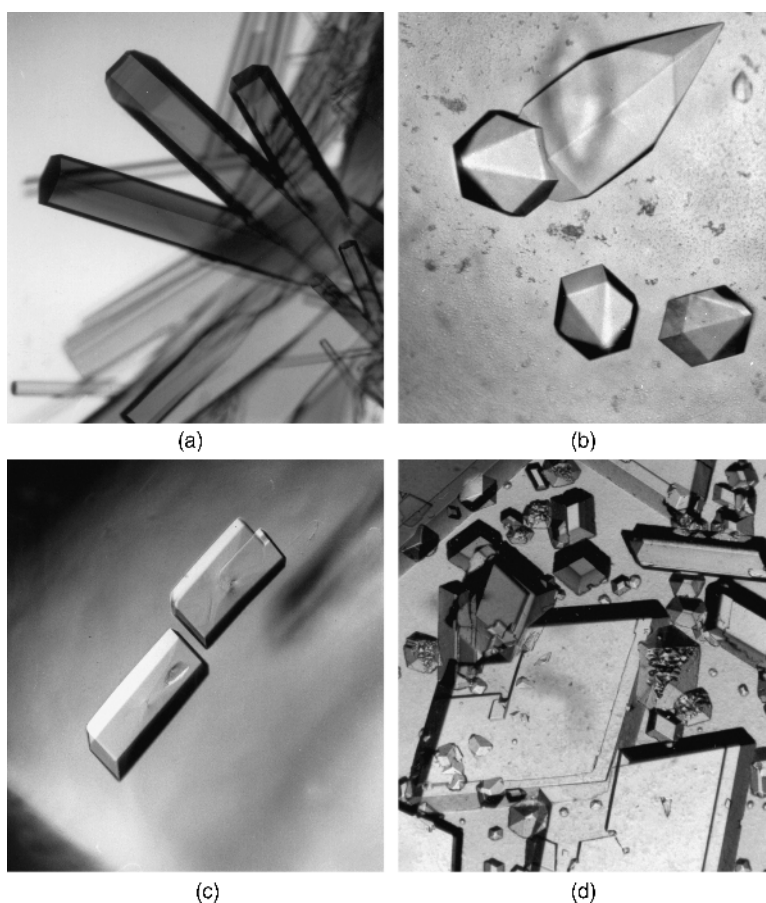


**FIGURE 1.9** The basic X-ray diffraction experiment is shown here schematically. X rays, produced by the impact of high-velocity electrons on a target of some pure metal, such as copper, are collimated so that a parallel beam is directed on a crystal. The electrons surrounding the nuclei of the atoms in the crystal scatter the X rays, which subsequently combine (interfere) with one another to produce the diffraction pattern on the film, or electronic detector face. Each atom in the crystal serves as a center for scattering of the waves, which then form the diffraction pattern. The magnitudes and phases of the waves contributed by each atom to the interference pattern (the diffraction pattern) is strictly a function of each atom's atomic number and its position  $x$ ,  $y$ ,  $z$  relative to all other atoms. Because atomic positions  $x$ ,  $y$ ,  $z$  determine the properties of the diffraction pattern, or Fourier transform, the diffraction pattern, conversely, must contain information specific to the relative atomic positions. The objective of an X-ray diffraction analysis is to extract that information and determine the relative atomic positions.

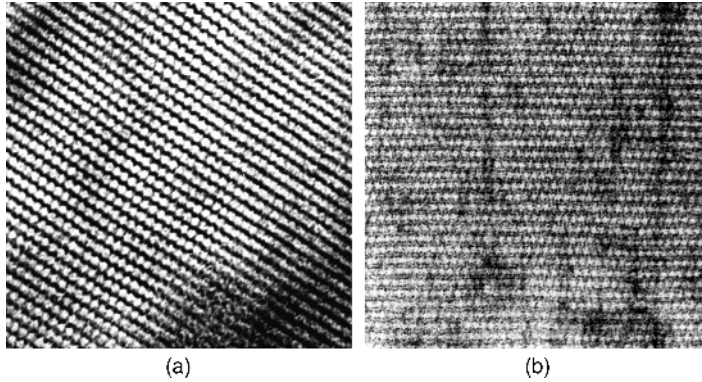
be directly imaged. If, however, a vast number of such molecules are organized into an array so that their scattering contributions are cooperative, then the resultant radiation can be observed and quantitated as a function of direction in space. This is precisely what is provided by macromolecular crystals, or in fact any crystals.

Some examples of typical macromolecular crystals (see McPherson, 1999, and Chapter 2) are shown in Figure 1.10. While objects of beauty, their regular features only begin to suggest the degree of their internal order. In Figure 1.11 we see with electron microscopy, and in Figure 1.12, with atomic force microscopy evidence of their exquisite, periodic nature. In Figures 1.11 and 1.12 the individual molecules that comprise the crystals are aligned in rows and columns, indeed in all three dimensions, in perfect register, every molecule disposed, every molecule in precisely the same environment as any other. It is the molecular equivalent of our parking lot for Volkswagen beetles, but in three dimensions.

The resultant radiation scattered, or diffracted, in specific directions create the intensities we see, precisely arranged, on an X-ray diffraction photograph. Because of the uniformity

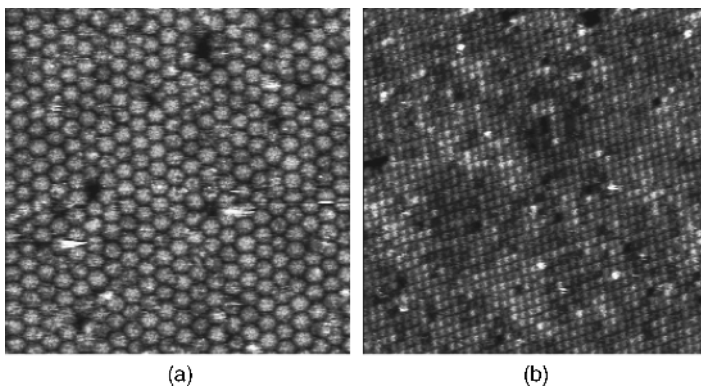


**FIGURE 1.10** Crystals of a variety of proteins. In (a) hexagonal prisms of beef liver catalase. In (b) crystals of  $\alpha_1$  acid glycoprotein, in (c) Fab fragments of a murine immunoglobulin, and in (d) rhombohedral crystals of the seed storage protein canavalin.

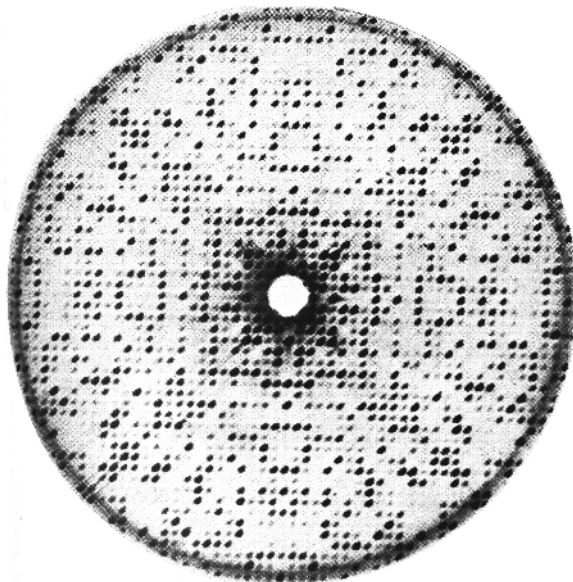


**FIGURE 1.11** Electron micrographs of negatively stained crystals of (a) pig pancreatic  $\alpha$  amylase, and (b) beef liver catalase. The dark areas represent solvent filled areas in the crystal, which are replaced by dense heavy metal stain; the light areas correspond to protein molecules where the stain is excluded. The underlying periodicity of the crystals is evident here, even after dehydration and staining.

of orientation and periodic position imposed on the molecules by the crystal lattice, the scattered X radiation, being waves, constructively interferes in unique directions dictated by the parameters that define the periodicity of the crystal lattice. It destructively interferes and sums to zero in all other directions. Hence we observe that the diffraction patterns from the ordered arrays that exist in crystals are absolutely discrete and that the observable diffraction pattern is an array of intensities that falls on a regular net or lattice (a reciprocal lattice). The spacings between the intensities, or reflections, and the symmetry properties that govern their distribution are manifestations of the periodic and symmetric disposition of the molecules in the crystal. Because the physical relationship between the diffracted rays and the crystal lattice is well understood, mathematical expressions, such as Bragg's law, can be written that describe the correspondence.



**FIGURE 1.12** Atomic force microscopy (AFM) also reveals the fundamental periodicity of macromolecular crystals. In (a) is the surface layer of a crystal of bromo mosaic virus, a particle having a diameter of about  $280 \text{ \AA}$ . In (b) is an AFM image of a monoclinic crystal of duodecahedral complexes of intact immunoglobulins, which have a diameter of about  $230 \text{ \AA}$ .

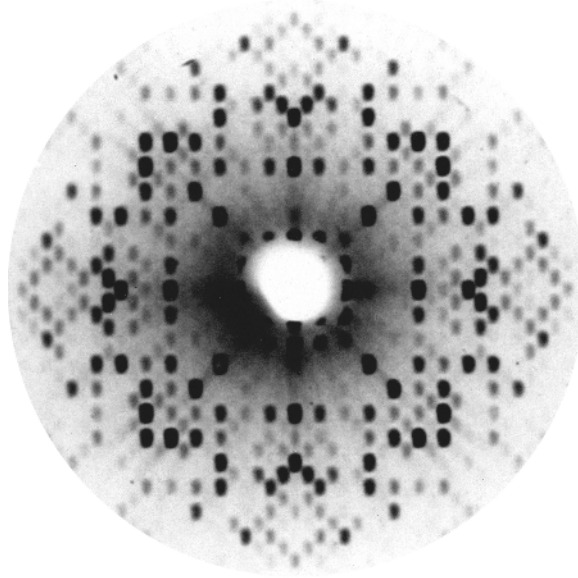


**FIGURE 1.13** The  $hk0$  zone diffraction image from a tetragonal crystal of lysozyme, an enzyme from hen egg white. Here, the fourfold symmetry of the pattern is striking, and it reflects the fourfold symmetry of the arrangement of the protein molecules in the unit cells of the crystal. Again, the intensities fall on a very regular, periodic net, or reciprocal lattice. The net is based on a tetragonal axis system (see Chapter 3).

It is clear from looking at diffraction patterns obtained from real crystals, such as those in Figures 1.13 and 1.14, that all of the reflections are not equal. They span a broad range of intensity values from very strong to completely absent. It will be shown in Chapter 5 (and was already demonstrated by Figure 1.8) that the variation in intensity from reflection to reflection is a direct function of the atomic structure of the macromolecules that comprise the crystal and occupy its lattice points. That is, the relative intensities of the reflections that make up the three-dimensional diffraction pattern, or Fourier transform, of a crystal are directly related to the relative  $x_j, y_j, z_j$  coordinates of all of the atoms  $j$  that define an individual molecule, and to the relative strength,  $Z_j$ , with which the different atoms scatter X rays.  $Z_j$  is the electron complement of each atom and is, therefore, its atomic number.

The complete diffraction pattern from a protein crystal is not limited to a single planar array of intensities like those seen in Figures 1.13 and 1.14. These images represent, in each case only a small part of the complete diffraction pattern. Each photo corresponds to only a limited set of orientations of the crystal with respect to the X-ray beam. In order to record the entire three-dimensional X-ray diffraction pattern, a crystal must be aligned with respect to the X-ray beam in all orientations, and the resultant patterns recorded for each. From many two-dimensional arrays of reflections, corresponding to cross sections through diffraction space, the entire three-dimensional diffraction pattern composed of ten to hundreds of thousands of reflections is compiled.

Because the diffraction pattern from a macromolecular crystal is the Fourier transform of the crystal, a precise mathematical expression can be set down that relates the diffracted



**FIGURE 1.14** Seen here is the  $hk0$  zone diffraction pattern from a crystal of M4 dogfish lactate dehydrogenase obtained using a precession camera. It is based on a tetragonal crystal system and, therefore, exhibits a fourfold axis of symmetry. The hole at center represents the point where the primary X-ray beam would strike the film (but is blocked by a circular beamstop). Note the very predictable positions of the diffraction intensities. All the intensities, or reflections, fall at regular intervals on an orthogonal net, or lattice. This lattice in diffraction space is called the reciprocal lattice.

waves to the distribution of atoms in the crystal. This expression, called the electron density equation, is a three-dimensional spatial transform that we refer to as a Fourier synthesis. It is sufficient here to simply understand that it is a summation of terms, one for each reflection observed in the diffraction pattern, and that the relative intensity of each reflection is the absolute magnitude of one of the terms in the series.

## THE PHASE PROBLEM

The situation, in truth, is somewhat more involved than this explanation would suggest. The individual reflections of the diffraction pattern are the interference sum of the waves scattered by all of the atoms in the crystal in a particular direction and, therefore, are themselves waves. Being waves they have not only an amplitude, but also a unique phase angle associated with each of them. This too depends on the distribution of the atoms, their  $x_j, y_j, z_j$ . The phase angle is independent of the amplitude of the reflection, but most important, it is an essential part of the individual terms that contribute to the Fourier synthesis, the electron density equation. Unfortunately, the phase angle of a reflection cannot be recorded, as we record the intensity. In fact we have no practical way (and rather few impractical ways either) to directly measure it at all. But, without the phase information, no Fourier summation can be computed. In the 1950s, however, it became possible, with persistence, skill, and patience (and luck), to recover this elusive phase information for

protein crystals, thus permitting the calculation of Fourier summations and hence images of macromolecules. The technique, which is known as multiple isomorphous replacement, is based on the chemical derivatization of protein crystals with heavy metal atoms such as mercury. Its development (Boyes-Watson et al., 1947; Bragg and Perutz, 1954; Blow and Crick, 1959) was the major breakthrough in modern crystallography that ultimately made possible the determination of macromolecular structures.

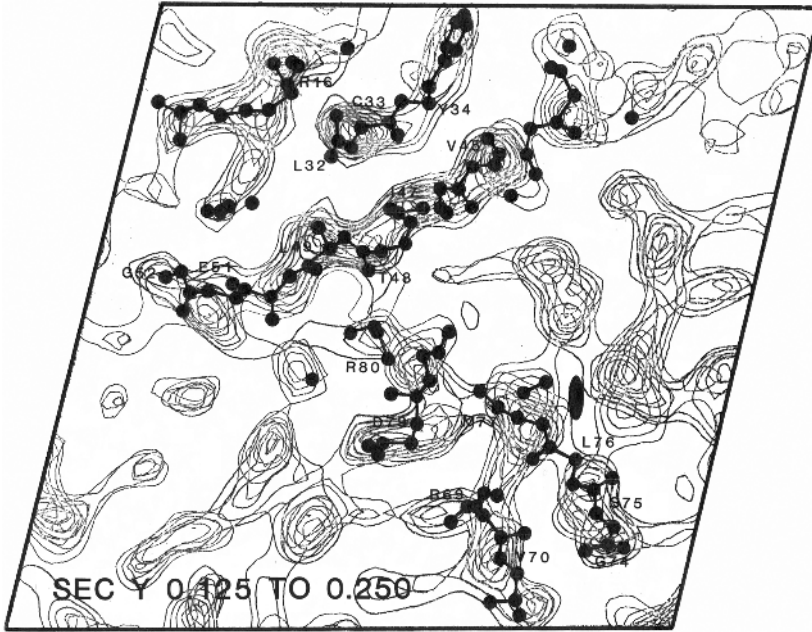
In this technique, described in more detail in Chapter 8, the heavy atom, whose position in the crystal can be determined by what are known as Patterson techniques (see Chapter 9), provides a reference wave. In a derivatized crystal the resultant diffraction intensities represent the sum of this heavy atom-produced reference wave interfering with the wave arising from all of the atoms in the protein. Just as the relative phase of a specific sound wave can be deduced by “beating” it against a reference sound wave of known phase, or for light waves using interferometry, the same is done for the native diffracted wave. The mathematical construct for obtaining the phase information requires measurement of the native diffraction intensities and the corresponding intensities from crystals independently derivatized at least at two unique sites in the crystal. It is known as a Harker diagram (Harker, 1956). From Harker diagrams for each of the reflections that comprise a complete diffraction pattern, the necessary phases, or at least reasonable approximations, can be obtained. This is discussed in more detail in Chapter 8.

## THE ELECTRON DENSITY

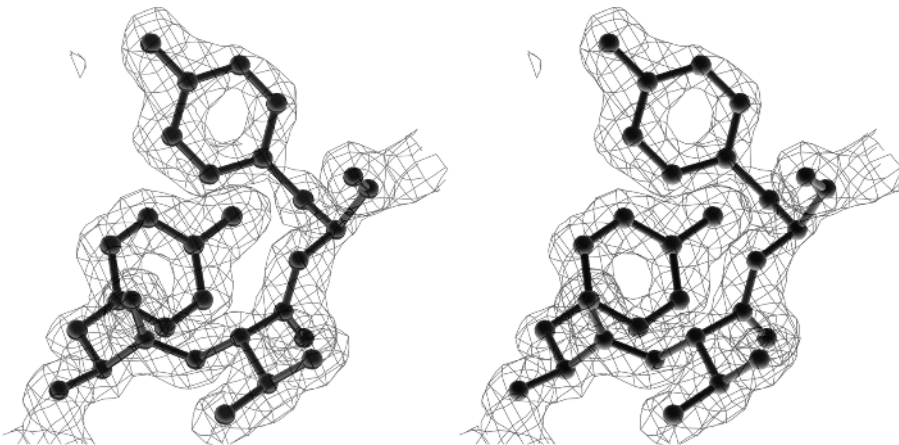
To produce an electron density image of the molecules comprising the unit cell contents of a crystal from the measured native structure amplitudes, and the approximate phase angles derived by isomorphous replacement (or by other methods, see Chapter 8), the value of the Fourier synthesis  $\rho(x, y, z)$  must be computed at every point  $(x, y, z)$  in the unit cell. The unit cell, as we will see in Chapter 3, is the actual repeating unit of the crystal. The electron density  $\rho(x, y, z)$  is a three-dimensional function and is continuous throughout the unit cell. A good approximation to this density continuum can be obtained by computing the value of  $\rho(x, y, z)$  on a grid of points whose separations are sufficiently small. The value of  $\rho(x, y, z)$  is calculated, for example, on grid points separated by distances  $\Delta x$  and  $\Delta y$  over a particular plane of constant  $z$  of the unit cell. The  $z$  coordinate is incremented by  $\Delta z$ , and  $\rho(x, y, z)$  is computed on all grid points in the plane  $(x, y, z + \Delta z)$ .

Figure 1.15 contains a composite of several planes from an electron density map of a protein. By continually increasing the final coordinate by  $\Delta z$ , the electron density map is built up from the series of two-dimensional planes. The individual sections are plotted on some transparent material after contour lines have been drawn around areas within certain density limits. The result is a topological map of the electron density presented on sequential planes of the unit cell as a series of contour levels. When the individual planes are stacked in consecutive order, a three-dimensional electron density image is created. This is discussed in more detail in Chapter 10. Currently, however, the presentation of the electron density is considerably more sophisticated. We use automated computer graphics systems to present detailed density images in three-dimensional space as in Figure 1.16.

Although measurement of, in many cases, the hundreds of thousands of X-ray reflections that are necessary to compute the Fourier synthesis of a macromolecule has in the past been an extremely time-consuming endeavor, this has, in recent years, become a far less arduous task. This is due to the advent of very rapid data collection devices based on area detectors,



**FIGURE 1.15** Electron density from a monoclinic unit cell of the Gene 5 DNA Unwinding Protein crystal, lying between  $y = 0.125$  and  $y = 0.250$ , is projected onto a single plane. Superimposed upon this electron density is a portion of the atomic model of the Gene 5 Protein. Electron density planes, like those shown here, are the images obtained directly by X-ray diffraction from computed Fourier syntheses.



**FIGURE 1.16** A more sophisticated presentation of electron density, in virtual three dimensions, is possible using computer graphics. In this stereo diagram, two tyrosines separated by a valine residue are superimposed upon their density in a  $1.8 \text{ \AA}$  resolution electron density map of the serine protease from penicillium cyclopium.

and high-flux density synchrotron X-ray sources. These, combined with enormous advances in computer technology, have now made it possible to solve the structure of virtually any protein that can be crystallized with, in the best of cases, no more than a few days to a week of concentrated effort. This time frame is continuing to shrink even now.

When the structure of a protein has been determined so that its constituent atomic positions are known to less than an angstrom, it can be precisely refined (see Chapter 10) by applying difference Fourier and various nonlinear least squares procedures (Hendrickson and Konnert, 1981; Sussman et al., 1997; Brunger et al., 1987). With these procedures, atomic coordinates are adjusted in increments of  $\Delta x$ ,  $\Delta y$ ,  $\Delta z$  so that chemical groups are made to approach ideal geometry, but in a manner that simultaneously minimizes the difference between the diffraction pattern calculated from the model structure, which we will see is possible, and the intensities actually observed experimentally. This can be done because the Fourier transform is symmetrical and allows calculation of either side of the transform from the other. By computing many refinement cycles, each accompanied by small optimizing shifts in the atomic coordinates, and using the difference between the calculated and observed diffraction pattern as a guide, convergence can ultimately be attained at a highly precise structural model having estimated errors in the atomic coordinates of less than a tenth of an angstrom. Structures refined in this way may then be utilized to carry out crystallographic experiments designed to elucidate the biochemical properties of the protein (see McPherson, 1987).