X-Ray Diffraction

If a picture is worth a thousand words, then a macromolecular structure is priceless to a physical biochemist. There is no better way to understand how macromolecules function in a cell than to have a visual image of their parts and how they interact. Our first encounter with the components of a living cell usually comes from peering through a light microscope. In a microscope, light is either reflected from the surface of an object or transmitted through the object. In both cases, the scattered light is focused by a series of lenses to form an image of the object (Figure 6.1). In this chap-

Figure 6.1 The light microscope. A microscope forms an image by focusing the light scattered from a sample, such as a melanophore on a slide, through a series of lenses. The image can be magnified infinitely, but the resolution is limited by the wavelength of visible light to about 0.2 microns. [Courtesy of P. McFadden.]
ter, we will discuss methods for seeing the atoms of a macromolecule by X-ray dif-
fraction. This method is analogous in many ways to light microscopy in that X-rays are
scattered by the atoms of the macromolecule. However, X-rays cannot be fo-
cused by lenses to form an image of a molecule. Instead, the X-rays are scattered
from a regular repeating array of molecules (that is, from a single crystal) to give a
pattern that represents the macromolecular order and structure. The structure must
be reconstructed using mathematics as the lens to transform the pattern back into
the original structure. Since the crystals are imperfect and information is lost during
the transformation, the structure that we get is not a true image, as from microscopy,
but a model of the structure (see Chapter 1). Building a model that best fits the ex-
perimental data is similar to solving for the solutions to a set of multiple equations.
We will first discuss the requirements necessary for determining the structure of a
macromolecule at the atomic level (to atomic resolution). This will be followed by a
description of what crystals are and how they are grown, and how X-ray diffraction
from crystals is used to solve structures of macromolecules to atomic resolution.
Finally, we will discuss fiber diffraction, a method that uses the repeating symmetry of
fibrous biopolymers to construct models that represent the average conformation of
these macromolecules without the need to crystallize them.

6.1 STRUCTURES AT ATOMIC RESOLUTION

Although a number of new, higher resolution microscopes have been developed
(Applicacion 6.1), the only method that currently yields reliable structures of macro-
molecules at atomic resolution is X-ray diffraction from single crystals. This tech-
nique requires three distinct steps: 1) growing a crystal, 2) collecting the X-ray dif-
fraction pattern from the crystal, and 3) constructing and refining a structural model
to fit the X-ray diffraction pattern. No one step is any more important than the oth-
ers and all three steps must be completed to determine the structure of a macro-
molecule.

A molecular structure solved to atomic resolution means that the positions of
each atom can be distinguished from those of all other atoms in three-dimensional
space, without the need to apply additional assumptions concerning the structure of
the molecule. The closest distance between two atoms in space is the length of a co-
valent bond. Since the typical covalent bond is approximately 0.12 nm, we need to
"see" two atoms separated by this distance as distinct particles (Figure 6.2). There
are theoretical and practical limitations to resolving a structure to this fine a level.
First, the system must have the atoms of its molecules held rigidly and, second, each
molecule or group of molecules in the system must have identical conformations.
Any fluctuation in the positions of the atoms in the molecule or any significant devi-
ations of molecules from a single conformation would result in an averaging of the
structure, which would blur our vision and thereby reduce the resolution of the tech-
nique. As we will see later, only a single crystal of a molecule has the potential to sat-
ify both of these requirements.
Application 6.1 Atomic Microscopy

Although X-rays can be used to resolve structures to atomic resolution, they cannot be focused to form an image. Recently, the methods of scanning probe microscopy has been developed as a form of atomic microscope. The two techniques, scanning tunneling microscopy (STM) and atomic force microscopy (AFM) work on basically the same principle. Rather than relying on the interaction of light with a molecule, STM and AFM are analogous to a traditional phonograph record player, that is they read the surface topology of an object using a very fine probe (Figure A6.1a). The stage on which the sample is fixed sweeps laterally as the probe tip scans the vertical height of the surface. The stage is a piezoelectric crystal that provides a very precise lateral translation of the sample. In STM, the tip measures the quantum mechanical tunneling current (the leaky current) from the the surface. The resulting fluctuations in current is a measure of the sample height. STM can image the carbon atoms of a graphite surface, and has even been used to manipulate single atoms to form the logo of a famous computer manufacturer. Unfortunately, STM works best with samples that can conduct electrons. Since most organic molecules are poor conductors, macromolecules have not been imaged to atomic resolution by STM.

AFM, on the other hand, does not rely on the electrical properties of an object, but simply the surface topology to form an image. As the tip is dragged or tapped over the sample, the vertical up and down motion is amplified and recorded as the vertical height of the sample. The method has now been developed to visualize samples in solution. The method of magnetic alternating current mode (Magnetic AFM) uses an oscillating magnet to gently tap the surface as it is swept across the probe tip. As a result, high-resolution images of molecules as small as the 126 to 168 base pair circles in Figure A6.1b have been

![Diagram of AFM setup](image)

**Figure A6.1a** Atomic force microscope (AFM). The microscope consists of a glass of mica sample slide that is translated laterally by a piezoelectric crystal stage. The probe tip on the end of a cantilever arm moves up and down as it scans the surface of the sample. This vertical motion is accentuated by a laser and recorded by a photodiode detector. (Adapted from C. K. Mathias and K. E. van Holde (1996). Biochemistry, 2nd ed., Benjamin-Cummings Publishing Co., Menlo Park, Ca.)
recorded (Han et al. 1997). In standard buffered solutions, the circular DNAs have very uniform radii. However, when the 168 base pair circles are exposed to solutions containing Zn\(^{2+}\), the DNA becomes sharply linked, reflecting the axial strain of these larger circles. The average width of the DNA at half height is about 3.5 nm, which suggests that broadening of the sample was less than 1 nm as the tip moves across the sample (the width of a DNA duplex is about 2.0 to 2.5 nm). This is a promising new technique for visualizing samples to near atomic resolution in their native environment, rather than in a crystal.

![AFM images of 168 base pair DNA circles in 1 mM MgCl\(_2\) (a) and 1 mM ZnBr\(_2\) (b). The images were recorded by oscillating the tip at 25 kHz with a vertical amplitude of 5 nm. Image courtesy of S. Lindsay and R. Harrington.](image)


We must now find a radiation source that allows us to see two atoms separated by only 0.12 nm. The limit of resolution \( LR \) of any optical method is defined by the wavelength \( \lambda \) of the incident radiation.

\[
LR = \frac{\lambda}{2}
\]

(6.1)

This is a consequence of the wave properties of light. An extension of the Hessenberg uncertainty principle that results from treating light as a wave states that the position of a particle cannot be fixed to better than about half the wavelength of the radiation used to examine that particle. The resolution of a light microscope, for example, is restricted by the wavelength of visible light (\( \lambda = 400 \text{ nm to 800 nm} \)) to about 200 nm, or 0.2 \( \mu \text{m} \) (about the size of the organelles in a cell). If we require the resolution of the technique to be about 0.12 nm to resolve the atoms of a macromolecule, the wavelength of light required for our atomic microscope would necessarily be <0.24 nm. This falls into the X-ray range of the electromagnetic spectrum (Figure 6.3). However, as we discussed above, X-rays cannot be focused and thus cannot form an image of an object in the same manner as a light microscope. We rely on the constructive and destructive interference caused by scattering radiation from the
Regular and repeating lattice of a single crystal to determine the structure of macromolecules. We must therefore first obtain a crystal of the molecule.

6.2 CRYSTALS

6.2.1 What Is a Crystal?

Crystals are solids that are exact repeats of a symmetric motif. What makes crystalline quartz different from amorphous glass? Both have essentially the same chemical composition. The most telling difference is that the molecules in a crystal are arranged in an orderly fashion, while the molecules of an amorphous solid, like a glass, are disordered. By ordered, we mean regular, symmetric, and repeating (see Chapter 1). This is particularly evident when we compare broken pieces of glass with pieces of a crystal. A glass will shatter to form random shards, with no relationship between the shape of the intact and broken pieces. A crystal, however, can be cleaved very specifically to give fragments that are smaller versions of the original crystal. This can be done almost indefinitely, until we reach the basic unit that describes a crystal, called the unit cell. Thus a crystal can be generated from a molecule.
by repeating a set of translational or rotational symmetry operators indefinitely. This has great utility; to determine the structure of a crystal, we need only determine the structure of the least symmetric component of the unit cell. Thus, in describing a crystal, we need only describe the least symmetric unit and the symmetry that gives us a crystal (Figure 6.4).

Any symmetric system can be reduced to a level that is not symmetric. For example, the $\alpha_2\beta_2$ tetramer of a hemoglobin molecule with identical $\alpha$-subunits and $\beta$-subunits can be constructed by applying a two-fold rotation to an $\alpha\beta$-dimer (Figure 1.33). There is no true symmetry relationship between the $\alpha$- and $\beta$-subunits of this dimer. In a crystal, the level at which there is no symmetry is aptly called the asymmetric unit. By analogy, the $\alpha\beta$-dimer can be considered to be the asymmetric unit of the hemoglobin tetramer in solution.

Starting with the asymmetric unit, we can now apply rotational or screw operators to construct the lattice motif. In the simplest hemoglobin example, a $C_2$ rotational symmetry operator applied to the $\alpha\beta$-dimer would generate the hemoglobin tetramer, which would be the lattice motif of a hemoglobin crystal.
Figure 6.4 Components of a crystal. The asymmetric unit is that part of the crystal that shows no symmetry. A symmetry operator (for example, a \( C_2 \) rotational axis) generates the lattice motif. Repeating this motif by translation generates the corners of the unit cell, which is the basic repeating unit of the crystal lattice.

The lattice motif is translated in three dimensions to form a regular and repeating array, called the crystal lattice, with each repeated motif forming a point within the lattice (Figure 6.4). The lattice points can be connected to form the corners of three-dimensional boxes. These boxes are the unit cells; each unit cell contains all the atoms of the lattice motifs and the asymmetric unit. We see that the edges of this box form the axes of the crystal system. The edges of the unit cell define a set of unit vector axes \( a, b, \) and \( c \), with the unit cell dimensions as their respective lengths \( a, b, \) and \( c \). These vectors need not be at right angles, and the angles between the axes are denoted as \( \alpha \) between the \( bc \)-axes, \( \beta \) between the \( ac \)-axes, and \( \gamma \) between the \( ab \)-axes. The relationship between the lengths \( a, b, \) and \( c \) and angles \( \alpha, \beta, \) and \( \gamma \) of the unit cell axes defines the unique shape and size of the unit cell. However, there are constraints placed on the shape of the unit cell.

A crystal is a stacking of unit cells repeated in three dimensions to build a lattice, leaving no space between the unit cells. This automatically constrains the shape of a unit cell to a parallelepiped, with four edges to a face and six faces to the unit cell. A crystal is constructed by translating the unit cells in three dimensions to fill a volume, the unit cell is constructed by translating the repeating motif (the lattice points), and the lattice motif is generated by applying symmetry operators to the asymmetric unit.

Since all unit cells within a crystal are identical, the morphology of a crystal is defined by the size and shape of a single unit cell (the lengths of \( a, b, \) and \( c \), and the angles \( \alpha, \beta, \) and \( \gamma \) and the symmetry of the motif). Each level of the crystal, with the exception of the asymmetric unit, can be generated using mathematical operators; solving a crystal structure requires only that we determine the configuration of the atoms in the asymmetric unit. In the case of our hemoglobin example, the structure of the entire hemoglobin tetramer in a crystal can be solved by determining only the structure of a single \( \alpha \beta \)-dimer in the asymmetric unit.

Within the constraints placed on the shape of the unit cell, there are 14 unique crystal lattices that can be constructed. These are the Bravais lattices (Figure 6.5). The combination of the 32 symmetry types (the point groups, Chapter 1) along with
the 14 Bravais lattices describe all the possible morphologies of crystals, yielding the shapes and symmetry of 230 different space groups. The space group assigned to a crystal uniquely defines the number of asymmetric units that assemble to form the unit cell of the crystal. Each space group specifies the lattice type and the symmetry of the unit cell.

All symmetry operators described by the various point groups can potentially be incorporated into a crystal lattice. However, there are constraints imposed by the definition of a unit cell and by the asymmetry of most biological polymers. Symmetry operators that invert the configuration of a chiral center are not allowed in crystals of biological macromolecules; mirror symmetry, which relates L- and D-stereoisomers, will not be found in crystals of naturally occurring biological macromolecules (for an exceptional case, see Figure 1.20). Thus, the two allowed types of symmetry operators that are observed in crystals of biological molecules are the rotational and screw symmetry operators. This reduces the 230 potential space groups to only 65 that are relevant to naturally occurring biological macromolecules. The designation for symmetry elements in the crystal are identical to those described in Chapter 1 (see Table 1.4).

The rotational components of crystal symmetry are restricted by the angles α, β, and γ. Since C₃ (or simply designated as 2 for a two-fold rotation) or Z₂₁ symmetry relates two objects by rotation through 180°, these symmetry elements require at least two sets of rectangular faces in a unit cell. Four-fold rotational symmetry requires that at least one set of the faces also be a square. We notice that five-fold rotation is not allowed in standard crystallography. A five-fold rotation or screw axis defines a pentagonal face and, since regular pentagons cannot be packed in three-dimensional space without leaving gaps, we cannot define a unit cell with one face having five edges (this rule has recently been challenged, but so far the controversy has not been resolved). Three-fold rotations require the three unit cell lengths and angles to be identical. This can be accommodated by placing the three-fold axis at the corner of a unit cell or having two adjacent three-fold axes on a face. A six-fold rotational or screw symmetry occurs on a hexagonal face and relates two edges of the unit cell by rotation through 60°. These rotations and their associated screw operators constitute the extent of the symmetry types that are possible for the unit cell of macromolecular crystals. In addition, the presence of two orthogonal symmetry axes automatically defines a third orthogonal symmetry axis. For example, a unit cell having 2₁, 2₂ symmetry (two perpendicular two-fold screw axes) must actually have 2, 2₁, 2₂ symmetry. The symmetry axes in a unit cell need not all intersect in the center. However, if two axes do intersect, the third axis must also intersect. Conversely, for two nonintersecting axes, the third axis must be nonintersecting.

With these rules in mind, we can now discuss some of the 14 Bravais lattices and their space groups in more detail. To start, we must add one qualifier to our definition of a unit cell. Although lattice points are necessarily found at the corners of the unit cell, they are not restricted to the corners. Additional lattice points may be found at the center of the faces or the center of mass of the unit cell. If lattice points are found only at the corners, the unit cell is primitive and the lattice is given the designation P. If lattice points are found at each face of an opposing pair, we have a centered or C
lattice. Lattice points at all six faces define a face-centered or $F$ lattice. Finally, a unit cell containing a lattice point at the center of mass is body-centered and designated as an $I$ lattice. There is a correspondence between the placement of lattice points and symmetry axes. For example, a unit cell that is an orthorhombic lattice ($a \neq b \neq c$, $\alpha = \beta = \gamma = 90^\circ$) can have lattice points at the corners, the centers of the faces or the centers of the unit cell, defining $P$, $C$, $I$, or $F$ as possible crystal lattices.

Together, the lattice type and symmetry of the unit cell define the space group of the crystal. The shorthand abbreviations for space groups incorporate the lattice type ($L$) and the symmetry of the unit cell ($R_T$, where $R$ is the rotation and $T$ designates the translational element of the symmetry operator) in the form $L,R,T,R_T$. Thus, the space group $P2_1/2_1/2_1$ is a primitive unit cell having two-fold screw axes parallel to each of the three crystallographic axes.

We will continue our discussion of Bravais lattices with the unit cell having the lowest potential for symmetry. This is the triclinic lattice, where $a \neq b \neq c$, and $\alpha \neq \beta \neq \gamma \neq 90^\circ$ or $120^\circ$. All edges of the unit cell are of different lengths and all faces are inclined. Since none of the faces are orthogonal or hexagonal, there can be no symmetry axes through any face. Similarly, the lattice can only be primitive. Thus a triclinic unit cell automatically defines a $P1$ space group. If one angle is set at $90^\circ$, then a second angle must also be $90^\circ$, forming a monoclinic unit cell. This requires a space group having a two-fold rotation or screw axis. We notice that even though there are two orthogonal faces, there can only be one symmetry axis. The introduction of a second symmetry axis automatically requires the definition of a third axis of symmetry, which requires a third orthogonal face. That is not allowed in a monoclinic space group. Thus, a monoclinic unit cell can be either $P2$ or $P2_1$. Finally, an additional lattice point can be added to the nonorthogonal face of the monoclinic unit cell, generating a $C$ lattice. Again, an $I$ or $F$ lattice requires that all three unique faces be orthogonal, which is contrary to the definition of a monoclinic unit cell. These restrictions to symmetry and placement of lattice points in the unit cell define the 14 Bravais lattices and the associated space groups (Table 6.1). We see that the lattice type along with the symmetry of the unit cell define the space group of the unit cell. The lengths and angles of the unit cell define the unit cell parameters, and the space group along with the unit cell parameters define the crystal morphology. We will see later that it is the crystal morphology that dictates the general characteristics of the X-ray diffraction pattern. Different crystals that have identical unit cell lengths and angles and are in the same space group are said to be isomorphous. Their X-ray diffraction pattern should also appear to be very similar.

We now understand that all the molecules within the crystal are generated by applying symmetry operators to a single asymmetric unit. Thus a crystal is nothing more than a single asymmetric unit mathematically replicated in three-dimensional space. Alternatively, we can see that all the molecular properties of a crystal can be attributed to those of the asymmetric unit. We can assert, with few exceptions, that all the molecules or groups of molecules that constitute the asymmetric unit in a single crystal suitable for X-ray diffraction studies have essentially identical conformations. Therefore, to solve the structure of a crystal, we need only solve the structure of the asymmetric unit.
TABLE 6.1 SIXTY-FIVE POSSIBLE SPACE GROUPS IN MACROMOLECULAR CRYSTALS

<table>
<thead>
<tr>
<th>Lattice Type</th>
<th>Possible Bravais Lattices</th>
<th>Crystal Shape</th>
<th>Possible Space Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triclinic</td>
<td>P</td>
<td>a ≠ b ≠ c</td>
<td>P1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a ≠ β ≠ γ ≠ 90°</td>
<td></td>
</tr>
<tr>
<td>Monoclinic</td>
<td>P, C</td>
<td>a ≠ b ≠ c</td>
<td>P2, P2₁, C2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a = γ = 90°, β ≠ 90°</td>
<td></td>
</tr>
<tr>
<td>Orthorhombic</td>
<td>P, C, I, F</td>
<td>a ≠ b ≠ c</td>
<td>P222, P2₁,2₂, P2₁,2₁,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a = β = γ = 90°</td>
<td>C222, C2₁,2₂, P2₁,2₂,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>F2,2,2</td>
</tr>
<tr>
<td>Tetragonal</td>
<td>P, I</td>
<td>a = b = c</td>
<td>P4, P4₁, P4₂, P4₃, P4₄,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a = β = γ = 90°</td>
<td>P4₂, P4₁,2, P4₂,2, P4₃,2,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P4₂,2, P4₁,2,2, P4₃,2,2,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>I4,2,2, A2</td>
</tr>
<tr>
<td>Trigonal</td>
<td>P, R</td>
<td>a = b = c</td>
<td>P₃, P₃₁, P₃₂, P₃₃,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a = β = 90°, γ = 120°</td>
<td>P₃₁,2, P₃₁,12, P₃₁,2₁,</td>
</tr>
<tr>
<td></td>
<td>(Rhombohedral)</td>
<td>a = b = c</td>
<td>P₃₁,2, P₃₁,2₁</td>
</tr>
<tr>
<td>Hexagonal</td>
<td>P</td>
<td>a = c ≠ b</td>
<td>P₆, P₆₁, P₆₂, P₆₃, P₆₄,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a = γ = 90°, β = 120°</td>
<td>P₆₁,2, P₆₁,2₂, P₆₂,2,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P₆₂,2, P₆₃,2, P₆₄,2,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P₆₄,2, P₆₄,2₂</td>
</tr>
<tr>
<td>Cubic</td>
<td>P, I, F</td>
<td>a = b = c</td>
<td>P₃₁, P₃₁₂, P₃₁₂₂, P₃₁₃₂,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>a = β = γ = 90°</td>
<td>F₃₁₂, F₃₁₂₂, I₃₁₂, I₃₁₂₂,</td>
</tr>
</tbody>
</table>

6.2.2 Growing Crystals

Now that we know what a crystal is, how can we go about growing a crystal of a macromolecule? There is no straightforward answer to this question because the field of crystallization is more an art than a science. Nonetheless, we can describe the general theory concerning crystallization. Remember that a crystal is a solid. However, macromolecules typically are in aqueous solution. Thus, we must somehow bring the molecule out of solution or precipitate the molecule in an orderly fashion to form a crystal.

A molecule comes out of solution when its concentration exceeds its intrinsic solubility, S° (S° is not actually an intrinsic property, but is dependent on external factors, including temperature, pressure, and the solvent). Growing a crystal requires bringing the concentration of the material in solution to supersaturation (Figure 6.6). The need for supersaturating the solution will become evident when we discuss the mechanism of crystal growth later. The concentration of the molecule can be increased by removing the solvent to decrease the overall volume of the solution. For example, a macromolecule with an intrinsic solubility of 2 mM in solution at 1 mM, can be precipitated by simply decreasing the overall volume to less than half the original volume. For small organic and inorganic compounds, crystals can be grown by simply evaporating solvent from solution.

A second strategy for precipitating a molecule is to decrease S° for a fixed concentration of material. For example, given the same 1 mM solution above, changing the temperature from 27°C to 4°C may be sufficient to reduce the intrinsic solubility
Figure 6.6 Mechanism of crystallization. The initial step in crystallization is the nucleation of a minimum crystal lattice. This is a low probability step that occurs in a supersaturated solution. The crystal grows by adding molecules to the surface of the seed, and occurs at concentrations close to the intrinsic solubility \( S^* \) of the molecule.

of the molecule by one-half. Alternatively, we can affect \( S^* \) by modifying the solvent (for example, by increasing or decreasing the ionic strength of the solution). The optimum solubility of polyelectrolytes, such as polypeptides or polynucleotides, generally falls within a specific range of salt concentrations. If the ionic strength of the solution is increased above this range by addition of salt or by evaporation, or decreased below this optimum range by dialysis, \( S^* \) will decrease and the molecule will precipitate. These are the standard salting in and salting out effects exploited in many purification techniques.

6.2.3 Conditions for Macromolecular Crystallization

How can we selectively grow a crystal out of solution as opposed to producing an amorphous precipitate? There are a number of factors that are important in crystallizing a molecule. Perhaps the most important is the purity of the sample. We can view purity in terms of biochemical purity and structural purity. A sample is considered to be biochemically pure when each macromolecule in the sample has the same molecular formula. This is clearly the ideal situation and it is not generally achieved. In most cases, a macromolecule must be better than 95% pure to produce a crystal.

The term structural purity refers to the conformation of the molecules in a particular sample. As we have seen in Chapters 2 and 3, the energetic difference
between various conformations of a macromolecule can be very small. Even a chemically pure sample may represent a highly heterogeneous population of conformations. Thus, the crystallization solutions and methods are chosen to favor the native conformation of the macromolecule.

We see that crystallization of a macromolecule requires finding a solution that balances solubility with structural homogeneity. These are likely to be under very specific conditions that constitute a minor subset of the conditions that fulfill either criterion. Thus, even though the conditions under which a molecule has a stable conformation and the conditions under which the macromolecule is insoluble are known, it may still be difficult to define the specific conditions under which a molecule will crystallize. For this reason, finding the conditions for crystallization is essentially a shotgun method, in which the molecule is placed under a large number of different solution conditions in the hope that one condition can be found that satisfies both the solubility and conformational requirements for crystal growth.

There is an increasingly large number of macromolecules (both proteins and nucleic acids) that have been crystallized from a large variety of different solutions. However, crystals of nearly all proteins and nucleic acids can be grown from a small subset of these conditions. There are now about 50 different buffer and salt solutions (Table 6.2) that are widely used as starting points to search for crystallization condi-

<table>
<thead>
<tr>
<th>TABLE 6.2 SCREENING SOLUTIONS IN SPARSE MATRIX METHODS FOR CRYSTALLIZING PROTEINS AND NUCLEIC ACIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salt</td>
</tr>
<tr>
<td>Proteins</td>
</tr>
<tr>
<td>None</td>
</tr>
<tr>
<td>0.2 M Na citrate</td>
</tr>
<tr>
<td>0.2 M Na acetate</td>
</tr>
<tr>
<td>Nucleic acids</td>
</tr>
<tr>
<td>12 mM Spermine, 20 mM Mg²⁺, 80 mM Na⁺</td>
</tr>
<tr>
<td>0.5 mM Spermine, 15 mM Mg²⁺, 2 mM BaCl₂</td>
</tr>
<tr>
<td>2 mM CaCl₂, 10 mM Mg²⁺</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
tions for macromolecules. A more scientific approach to crystallizing macromolecules is still desirable (Application 6.2).

To understand the techniques for crystallizing a macromolecule, we need to compare the mechanisms for crystal growth to those for amorphous precipitation. In comparison with the molecule in solution, the highly ordered molecules in a crystal lattice have significantly lower entropy. The external degrees of rotational and translational freedom in a crystal lattice are much reduced from those in solution. For example, two molecules in solution each have complete translational and rotational freedom in three dimensions. If two molecules associate to nucleate the formation of a crystal lattice, their movements are exactly correlated and the external degrees of freedom are reduced by a factor of two. Thus, the difference in entropy between the dimer and monomer states of the two molecules is \( \Delta S = -R \ln(2) = -5.8 \, \text{J/mol K} \). Two molecules, however, do not form a stable nucleation complex. It has been estimated that the contents of at least four unit cells must come together in a highly cooperative manner to form a stable and unique nucleation lattice. This is equivalent to the assembly of 16 asymmetric units for a relatively highly symmetric space group such as \( P2_12_12 \). The minimum \( \Delta S \) for the formation of this nucleation lattice would be \(-R \ln(16) = -23 \, \text{J/mol K}\). In solution, the molecules will have a distribution of conformations, but in the crystal there is only a single conformation. Therefore, there is an additional loss in conformational entropy during crystallization. This latter value is more difficult to estimate.

The growth step of the crystal is envisioned to be single molecules adding to the surfaces of the nucleating lattice. The loss in entropy at these subsequent steps are expected to be less than for the nucleation step. Crystal growth, therefore, occurs in two distinguishable steps: 1) a low-probability nucleation step to form a seed, and 2) a higher-probability growth (or propagation) step to increase its size (Figure 6.6). The two steps in crystal growth are analogous to the steps in the zipper model for structural transitions in macromolecules (see Chapter 4).

We see that nucleation of crystal growth requires a very large driving force. The necessary driving force for nucleation comes from bringing the concentration of macromolecules well above their intrinsic solubility (supersaturation). The subsequent addition of single molecules at the crystal surfaces occurs at concentrations below supersaturation, near the intrinsic solubility. Thus, growth is a series of microequilibrium steps for macromolecules in solution at the surface of the crystal lattice.

Nucleation is the more important of the two steps. If the solution does not reach supersaturation, a nucleus is not formed and a crystal cannot grow. On the other hand, a solution that becomes supersaturated very rapidly will form multiple nuclei that will rapidly deplete the molecules from the crystallization solution. This results in a shower of tiny crystals. There must be a balance between the two extremes in order to grow single crystals that are sufficiently large to provide high-resolution X-ray diffraction data. Thus, the macromolecules in solution must be concentrated in a very well-controlled manner.

A number of methods have been developed to facilitate the process of crystallizing macromolecules. Two widely used strategies are vapor diffusion and microdialysis, both of which are designed to vary the solution environment, reducing the
Application 6.2 Predicting Crystallization Solutions for DNA Fragments

The hit-or-miss approach to crystallizing macromolecules is far from scientific. Sparse matrix approaches present recipes for crystallization solutions that facilitate the search for the right conditions to grow a crystal that is suitable for X-ray diffraction. However, it would be much more satisfying to be able to predict the solutions that will crystallize a macromolecule based on its physical properties. This has been done for at least one class of molecules, DNA fragments that crystallize as left-handed Z-DNA.

Z-DNA is less stable than B-DNA under standard buffer conditions, but it is stabilized by salts and alcohols, both of which are used for macromolecular crystallization. This is because B-DNA is more hydrated than its left-handed counterpart. The amount of salt required to crystallize Z-DNA has been observed to depend on the ability of a sequence to adopt this conformation (see Figure A6.2a). Thus, using a simple calculation of the relative stability of a sequence as Z-DNA relative to its B-form, the salt concentration needed to crystallize a sequence as Z-DNA can be calculated using a linear relationship. We notice that the salt concentration is in log units, which means that a sequence that is slightly less stable than Z-DNA may require a very large increase in the salt con-

![Diagram](https://via.placeholder.com/150)

**Figure A6.2a** Crystallization solutions for Z-DNA predicted from free energies. The effective cation concentration in the crystallization solutions ($\log CS = \log (Z[Z]/[M])$, where CS is the cation strength, $Z$ is the charge, and $[M]$ the concentration of the cation species $i$ in solution is plotted relative to the free energy difference between the Z-DNA form and B-DNA form of each sequence. Free energy difference was calculated from the hydration free energies of the solvent accessible surfaces (see Chapter 4). [Data from Kagawa et al., *Nucleic Acids Res.* 21: 5978-5986.]
centrations, so much so that it may not be possible to crystallize that sequence. By studying the conformation and the form of the DNAs under various solution conditions, the crystallization pathway was mapped to show the conformational and phase transitions caused by the salts and alcohols in the crystallization solutions (Figure A6.2b).

![Crystallization pathway for Z-DNA fragments](image)

**Figure A6.2b** Crystallization pathway for Z-DNA fragments. The solution and solid phase transitions are plotted as a function of the salts (CS) and precipitants (2-methyl-2,4-dimethylpentane diol or MFD) in the crystallization solutions. The arrows indicate three trial crystallization pathways. [Data from Ho et al. (1991)].


Solubility of a macromolecule in a controlled manner. In both cases, the molecule is initially dissolved in a buffered solution, usually containing a mixture of solvents and salts. This sample solution is physically separated from a larger volume of solvent called a reservoir. The sample and reservoir solutions are placed in a closed system, and solvent or salts are removed from (or sometimes added to) the sample solution to render the macromolecules less soluble. In vapor diffusion methods (Figure 6.7), solvents are transferred according to the vapor pressure of the sample versus the reservoir. In this typical system, water is drawn from the sample so that the volume of the sample systematically decreases until the vapor pressure of the water in the sample and that in the reservoir have equilibrated. The two most common vapor diffusion techniques are the hanging drop method, where the sample literally hangs above the reservoir, and the sitting drop method, where the sample sits in a well surrounded by the reservoir. In the microdialysis method, solvent is transferred by equilibrating the osmotic pressures of the sample and reservoir across a semipermeable membrane (see Chapter 13).
Figure 6.7 Vapor diffusion methods of crystallization. In the hanging drop method of vapor diffusion, a sample in solution is suspended above a reservoir, \( R \), that contains a high concentration of a precipitant. The lower vapor pressure of the reservoir draws water from the sample solution, \( S \), to reduce the volume of the sample, \( V_s \), below its initial volume, \( V_{s0} \). Consequently, the concentration of molecules in the sample solution, \( [S] \), increases to above the intrinsic solubility \( S^0 \) of the molecule, resulting in precipitation or crystallization. In the sitting drop method, the sample solution sits in a well rather than hanging suspended, but otherwise the two methods are the same.

With any luck, we obtain a single crystal of a macromolecule that will diffract X-rays. The process of crystallization is but one leg in the triad necessary to solve a structure. With a crystal in hand, we can now collect X-ray data to solve the structure of the crystal.

6.3 THEORY OF X-RAY DIFFRACTION

As we discussed above, the wavelength of X-ray radiation is well suited for resolving atoms separated by the distance of a covalent bond. The energy of a quantum of this radiation is approximately 8000 eV, which is approximately the energy of electrons in their orbitals. This equivalence of energy leads to interactions so that the electrons of an atom will primarily be responsible for the scattering of X-rays. The number of electrons in a given volume of space (the electron density) determines how strongly an atom scatters X-rays. The interference of the scattered X-rays leads to the general phenomenon of diffraction. In this section, we will briefly review the general theory.
of diffraction, extend this to X-ray diffraction, and apply X-ray diffraction to single crystals. We can then consider what can be learned about the crystal morphology and how we can transform the diffraction data into the structure of the molecules in the crystal.

In X-ray diffraction, we treat all electromagnetic radiation as waves. The general theories and consequences of diffraction are applicable to any energy of radiation; we often see diffraction of visible light to illustrate the principles involved in X-ray diffraction from a crystal. We start our discussion by considering the components of diffraction, scattering and interference.

Scattering simply refers to the ability of objects to change the direction of a wave. With visible light, the simplest example is the reflection from a mirror; the mirror simply changes the direction of the light waves. Nonreflecting objects will also scatter light waves. An object placed in the path of a light from a point source cannot cast a sharp shadow because of scattering from its edges. The origin of scattering can be best developed if we start with Huygen’s principle, which states that every point along a wavefront can be considered to be the origin of a new wavefront, as we see in Figure 6.8a. The velocity of this new wavefront is equal to that of the original. For an unimpeded wavefront, the secondary wavefront can be constructed by drawing circles with radius $r = ut$ at points along the starting wavefront and connecting the tangents to each of these circles (Figure 6.8b). Objects placed in the path of a wavefront act as points of propagation for new wavefronts. The entirely new wavefront is called a scattered wave.

If we now place two point objects (A and B) in the path of the wavefront, each of the two points will propagate a new wavefront having identical wavelengths and velocities (Figure 6.9). The offset in the maximum amplitudes of the two waves (their relative phase) depends on the positions of A and B relative to the origin of the initial wavefront. At some position in space, the wave propagating from A will reinforce the scattered wave from B through constructive interference if the two scattered waves are in phase. Alternatively, the wave from A will reduce the amplitude of the wave from B through destructive interference if the amplitudes are out of phase. This is called diffraction; the sum of the two waves propagated from A and B result in an amplitude that is dependent on the relative positions of A and B and is also dependent on where the new wavefronts are being observed. Concomitantly,
if we make several observations of the amplitude of the new wavefront at different positions, we can extrapolate the information to determine the relative positions of the diffracting objects A and B relative to the origin of the initial wavefront. This is how X-ray diffraction is used to solve the structure of molecules in single crystals. In the remainder of this chapter, we will develop the conceptual tools necessary for understanding how the positions of atoms are determined by the diffraction of X-rays.

6.3.1 Bragg's Law

W.L. Bragg developed a simple relationship in 1912 to understand how diffraction relates to the relative positions of point objects in space. To derive Bragg's law of diffraction, we must first slightly modify our conceptual model of diffraction and think of lattice points in the crystal as parallel planes. Stacking a set of reflecting planes at regularly spaced intervals d creates a simple model of a one-dimensional crystal (Figure 6.10). In this model, a wave of X-rays (with wavelength λ) is incident on the reflecting planes at an angle θ. The wave is scattered by reflection from the planes at an identical angle θ. We can ask at this point which values of θ will result in constructive versus destructive interference. We assume that the distance to our point of observation is very large compared to d, so the individual paths of scattered light are essentially parallel. Because we have a large number of planes, we observe constructive interference only when the reflected waves are perfectly in phase (peaks aligned with peaks, node with nodes, and valleys with valleys). This only occurs when
the difference in the length of the path of the incident and the reflected waves of each plane, \( PD \), is equal to some integer \( n \) of the wavelength of the incident X-rays.

\[
PD = n\lambda \tag{6.2}
\]

This path difference is related to the distance separating the reflecting planes by the simple trigonometric relationship,

\[
\frac{1}{2} PD = d \sin \theta \tag{6.3}
\]

Substituting Eq.6.3 into Eq. 6.2, we derive Bragg's law of diffraction

\[
2d \sin \theta = n\lambda \tag{6.4}
\]

or

\[
2 \sin \theta = \frac{n\lambda}{d} \tag{6.5}
\]

This simple relationship makes a profound statement concerning the properties of diffraction. There is a reciprocal relationship between the Bragg angle \( \theta \) and the spacing between the reflecting planes. This means that larger spacing of repeating units in a crystal results in smaller diffraction angles. If we consider now a simple one-dimensional crystal containing a single atom per unit cell along the \( c \)-axis, Bragg's law tells us that we can determine the length of the unit cell along the \( c \)-axis by measuring the Bragg angle. We will treat this topic in greater detail later in this chapter. For an infinite row of scatterers, the only condition at which a reflection will be observed is when the Bragg law condition is met. Under all other conditions, the scattered radiation is annihilated.
Some immediate questions become obvious concerning Bragg’s law. Is reflection an accurate way of depicting a diffraction event? What is reflecting the X-rays? What sits between the reflecting planes (a void)? What happens in three dimensions?

6.3.2 von Laue Conditions for Diffraction

Clearly, treating X-ray diffraction as simple reflection from parallel planes is not an entirely accurate representation of the actual process. First, diffraction is a consequence of scattering from atoms. Atoms scatter X-rays, not because they are shiny spheres but because they are oscillating dipoles. Thus, the scattered waves do not necessarily travel along the single path of a reflected wave; atoms scatter X-rays in all three dimensions. Second, reflection from a plane can only occur when the incident angle \( \theta \) is not equal to zero. Thus, if our one-dimensional crystal of parallel planes is exactly perpendicular to the incident X-ray beam, we predict no reflection. Scattering from the atoms of a crystal has no such constraint. Scattering is a general phenomenon, and reflection from a plane is merely a specific example of it. Still, the data observed during an X-ray diffraction experiments are often called reflections because, as we will see, Bragg’s law correctly describes the conditions of constructive interference in X-ray diffraction. Let us now develop a slightly different model that more accurately describes X-ray diffraction from atoms.

We will start with the simplest case by building a one-dimensional crystal in which single atoms are spaced along the crystallographic c-axis (Figure 6.11). If we treat these atoms as scattering points, then the incident X-rays can be scattered in any direction. We will now orient this crystal with the c-axis perpendicular to the incident radiation. Now imagine scattering from the points in all directions. Only certain directions of the scattered radiation will reinforce; these will be such that the path difference PD between the rays scattered by adjacent atoms corresponds to an integral number of wavelengths. Notice that this condition is no different from that

![Figure 6.11 von Laue conditions for diffraction. A set of scattering atoms arranged in a regular array are spaced by a distance c along the vertical axis. Constructive interference occurs when the angle \( \gamma \) of the scattered beam relative to the crystal axis conforms to the von Laue conditions for scattering. These conditions are analogous to Bragg’s law when reflecting planes at each scattering atom (broken lines) form an angle \( \gamma \) relative to the incident and the scattered beams. The diffraction angle relative to the incident beam is 2\( \theta \).](image)
for deriving Bragg's law. If $\gamma$ is the angle between the direction of the scattered radiation and the row of scatterers, this condition is given by

$$l \lambda = c \cos \gamma$$

(6.6)

where $l$ is an integer and $c$ is the spacing between points (or the length of the unit cell along $c$). If the row of points is very long, complete annihilation of the scattered waves will occur at all other angles. Then, from such a row, radiation will be scattered only along the surfaces of cones, with the conical axis lying coincident with the crystallographic $c$-axis. This is illustrated in Figure 6.12, where the distance to the point of observation is large compared to the size of the crystal and the crystal serves as the point of origin for the scattered X-rays.

If the incident radiation makes an angle $\gamma_i$ other than 90° with the row of scatterers, Eq. 6.6 must be modified to be

$$l \lambda = c \cos \gamma - \cos \gamma_i$$

(6.7)

If we expand this to a two-dimensional array, with spacing $a$ and $c$ along the $a$ and $c$ axes, Eqs. 6.7 and 6.8 must be satisfied simultaneously,

$$h \lambda = a \cos \alpha - \cos \alpha_i$$

(6.8)

Equation 6.8 describes a second set of cones that is now coincident with the crystallographic $a$-axis. We see that reinforcement will occur only where the two sets of cones intersect, which is a line. For a three-dimensional array, such as an orthorhombic crystal with spacings $a$, $b$, and $c$, a third condition must be simultaneously satisfied,

$$k \lambda = b \cos \beta - \cos \beta_i$$

(6.9)

Equations 6.7 to 6.9 are called the von Laue conditions for diffraction. The three resulting cones will intersect at points that are the points of reflection.

Figure 6.12 An incident beam of X-rays causes a set of scattering cones from a one-dimensional crystal aligned along the vertical axis. Each cone makes an angle $2\theta$ relative to the incident beam to conform to the von Laue conditions for diffraction. The intersection of each cone with a piece of flat photographic film is an arc. Each arc is a layer line representing the order of the reflection, the integer index $l$ in Eq. 6.6. In a three-dimensional crystal, each axis of the unit cell generates a set of concentric cones, with the conical axes aligned parallel with the crystallographic axes.
These conditions are satisfied simultaneously only for certain specific orientations with respect to the vector describing the incident beam. Suppose that the crystal is arbitrarily oriented with respect to the incident beam. We can ask the following question: If we consider an angle $2\theta$ between the incident beam and the reflected beam, as illustrated in Figure 6.11 (we will relate this to the diffraction angle later), is there reinforcement of the scattered X-rays in this direction? If so, we have a reflection and the von Laue conditions must be satisfied. To simplify the nomenclature at this point, we write $\alpha = \cos \alpha$, $\beta = \cos \beta$, and so forth. We see that these are the direction cosines for the diffracted and the incident X-ray beams as vectors relative to the crystal axes. The following relationships must apply to the direction cosines of the two vectors

$$\alpha^2 + \beta^2 + \gamma^2 = 1$$

$$\alpha_0^2 + \beta_0^2 + \gamma_0^2 = 1$$

and if the angle between the two vectors is $2\theta$, we have

$$\cos 2\theta = \alpha_0 \alpha + \beta_0 \beta + \gamma_0 \gamma$$

These are standard geometric relationships. Finally, if we square the von Laue conditions, we find

$$\frac{h^2 \lambda^2}{a^2} = \alpha^2 - 2\alpha_0 \alpha + \alpha_0^2$$

$$\frac{k^2 \lambda^2}{b^2} = \beta^2 - 2\beta_0 \beta + \beta_0^2$$

$$\frac{l^2 \lambda^2}{c^2} = \gamma^2 - 2\gamma_0 \gamma + \gamma_0^2$$

or summing,

$$\left( \frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2} \right) \lambda^2 = 1 - 2(\alpha_0 \alpha + \beta_0 \beta + \gamma_0 \gamma) + 1$$

$$= 2(1 - \cos^2 \theta)$$

By a standard identity, $1 - \cos^2 \theta = 2 \sin^2 \theta$, therefore

$$\left( \frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2} \right) \lambda^2 = 4 \sin^2 \theta$$

or

$$\lambda \left( \frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2} \right)^{1/2} = 2 \sin \theta$$

(6.17)

This is the von Laue equation. It uses the defined axes of the unit cell and holds in three dimensions for all planes of atoms in a crystal in which the $a$, $b$, and $c$-axes are orthogonal.
How do Bragg's law and the von Laue conditions relate? To find the relationship, we need only ask what the reflecting planes are and what the Bragg angle represents. Figure 6.11 compares the von Laue row of scatterers for a one-dimensional crystal to the Bragg reflecting planes, which by definition form equal angles with both the incident and the scattered beams. The diffraction angle can now be defined relative to the direction of incident beam. Since the incident beam is fixed, this is a more convenient reference than the scattering plane, which is variable. We readily see that the angle of scattered beam relative to the incident beam is twice the Bragg's angle, \( 2 \theta \). In practice, this is the angle we measure, called the diffraction or scattering angle. Thus, for each value \( l \) in Eq. 6.6, equivalent to \( n \) in Eq. 6.5, there is a series of planes that generate a continuous set of scattered beams. For each value of \( l \), the scattered beams form a set of cones in which the angle of the cone relative to the incident beam is the diffraction angle \( 2 \theta \), and the axis of the cone lies parallel to the direction of the one-dimensional crystal. We notice that \( l = 0 \) conforms to the conditions for diffraction, and yields a plane of scattered X-rays, with \( 2 \theta = 0 \). Under the von Laue equation, \( h, k, \) and \( l \), which are called the Miller indices, define the integer number of wavelengths that result in an observed reflection from a three-dimensional crystal. Thus for a given set of Miller indices \( h, k, \) and \( l \), Bragg's law and the von Laue equation are equal.

\[
\left( \frac{k}{a} + \frac{l}{c} \right)^{1/2} = \frac{2 \sin \theta}{\lambda} = \frac{n}{d}
\]

(6.18)

If we were to record the cones of diffraction from a one-dimensional crystal using a planar sheet of photographic film, the diffraction pattern that we observe would be the intersection of each cone with a plane. This is a distorted picture of the diffraction pattern in that the intersection of a cone with a plane is an arc (Figure 6.12). Each line is called a layer line and is numbered according to the value \( l \). To straighten the layer lines from a one-dimensional crystal, we would need to wrap the film into a cylinder with its long axis lying along the axis of the cone.

As the crystal is expanded to three dimensions, each additional dimension yields a set of cones whose diffraction angle satisfies the von Laue conditions. The observed diffraction from a two- or three-dimensional crystal is the intersections of each series of cones as discussed above. The resulting points of reflection can be seen by comparing the intersection of a film plane with each set of cones from a two-dimensional crystal (Figure 6.13). Each set of cones generates its own set of layer lines, but the actual observed reflections lie at the points of intersection for each layer line; in a three-dimensional crystal, we can imagine a sphere of reflections where each reflection is a point on the surface of a sphere.

6.4 DETERMINING THE CRYSTAL MORPHOLOGY

Before we discuss the complete analysis of each reflection and how this is used to solve the structure of a molecule, we will discuss the general features of the diffraction pattern and how this relates to the morphology of the unit cell. Bragg's law in
Figure 6.13 Recording diffraction data. The reflecting cones from a crystallographic axis can be recorded using a piece of photographic film that is flat, cylindrical, or spherical (shown in the three figures in (a)). The resulting layer lines form one-dimensional arcs on the flat film, and straight lines on both the cylindrical and spherical films (b). The spacing between layer lines increases on the cylindrical film as the order of the layer lines $n$ increases from zero, but remains constant for all layer lines on spherical film. The diffraction from a three-dimensional crystal are points at the intersection of the cones from the three crystallographic axes (c). These points are at the intersection of arcs on flat film, at the intersection between arcs and lines on cylindrical film, and the intersection of straight layer lines on spherical film. A precession photograph mimics spherical film by rolling or precessing a flat piece of film about the crystal axes. This undistorted diffraction pattern.

Eq. 6.5 tells us that the spacing of the indexed reflections are inversely proportional to the lengths of the crystal unit cell. In addition, these spacings are not affected by the number, types, or positions of the atoms in the unit cell. We can therefore use these spacings to directly determine the lengths of the unit cell in the crystal, regardless of the molecules in the crystal. In addition, the orientation of the reflections in the reflection sphere mirrors the orientation of the primary axes in the unit cell (i.e., the angles between the edges of the unit cell). Finally, the symmetry within the unit cell explicitly defines the symmetry of the diffraction pattern. Thus, by observing the spacing and pattern of reflections on the diffraction pattern, we can determine the lengths and angles between each side of the unit cell, as well as the symmetry or space group in the unit cell, together these define the morphology of the
crystal. First, let us consider a simple experiment to record the X-ray diffraction pattern on photographic film. Although film is seldom used today to actually collect data to solve molecular structures, the methods help to illustrate the basic principles of X-ray diffraction. Indeed, many modern X-ray detectors used in crystallography are nothing more than electronic equivalent to photographic film.

Since the relative spacing between reflections is important for determining the lengths of the unit cell axes, we need to record an undistorted diffraction pattern. Ideally, we would use a piece of spherical photographic film to record the diffraction pattern from the sphere of reflections (Figure 6.13). In the real world, however, we are limited to flat sheets of film. Precession photography was designed to circumvent this problem. A precession camera rotates both the crystal and film in concert to give a photograph in which the spacings and the intensities of the diffraction pattern are recorded in an undistorted manner. Each precession photograph can be thought of as a slice through the sphere of reflection. However, no information is obtained for the axis about which the crystal is rotated. Therefore, at least two orthogonal precession photographs are required to analyze a three-dimensional crystal.

The regular pattern recorded on a precession photograph (for example, in Figure 6.14) can be used to define the unit cell parameters of a crystal. Each reflection in the diffraction pattern can be assigned to a unique set of Miller indices from the von Laue conditions for diffraction. For us to learn about the morphology of the crystal unit cell, we must first index the precession photograph in terms of the three Miller indices \( h, k, \) and \( l \). To do this, we take advantage of the inherent symmetry of the diffraction pattern to first define the principal axes of the pattern, and subsequently to define the principal axes of the unit cell. We notice that the von Laue conditions describe diffraction at the angle \( 2\theta \) relative to the incident beam. This treatment should also hold true for \( -2\theta \), which occurs in the opposite direction.

![Figure 6.14 Precession photograph of the tetragonal crystal of lysozyme. The photograph was recorded along the four-fold symmetry axis. The photograph is indexed using the vertical and horizontal primary axes shown. An alternative set of primary axes for indexing is indicated along the diagonal. In this latter case, the crystal unit cell will be defined to be larger than the set chosen. The distance between 10 diagonal layer lines is smaller than that of 10 vertical layer lines, which corresponds to a larger unit cell along the diagonal. [Courtesy of P.A. Karplus.](image)
(sin(−θ) = −sin θ). For all space groups including the lowest symmetry triclinic P1, the two halves of the reflection sphere should be symmetric. This is known as Friedel’s law, and the reflections related by Friedel’s law are called Friedel pairs. Friedel’s law simply states that a reflection with Miller indices h,k,l should be identical for one at −h,−k,−l. Thus if we assign a principal axis for all values of h with k = l = 0 (called the h 0 0 axis), any reflection h will have spacings and intensities identical to that of −h. Similarly, the intensities of reflections along k will be identical to that of −k for h = 0; the same is true for l and −l. Therefore, we can find the three principal axes by simple inspection of the symmetry in the diffraction pattern. The intersection of the three axes is the origin (h = 0, k = 0, l = 0).

Using Friedel’s law, we can see by inspection where the three principal axes are in the diffraction pattern, but what are their identities? The h 0 0 axis corresponds to the a-axis of the unit cell, the 0 k 0 axis to the b-axis, and the 0 0 l axis to the c-axis. For the low symmetry triclinic unit cell, the a-axis of the crystal unit cell is defined as the shortest of the three axes. The reciprocal relationship for diffraction automatically defines the h 0 0 axis as having the largest spacing between diffraction layers. Similarly, the c-axis is the longest edge of a triclinic unit cell, making the 0 0 l axis the one with the smallest spacing between layer lines. Thus the k 0 0 axis in the precession photograph has the largest spacing between reflections, the 0 0 l axis has the smallest spacing, and the 0 k 0 axis has intermediate spacing. Other lattices will have well-defined rules for assigning a, b, and c, and thus for indexing the Miller indices according to the symmetry of the unit cells, as listed in the International Tables of Crystallography (Hahn 1989).

Having indexed the reflections, we can now define the geometric parameters of the unit cell, which are the lengths and the angles between each principal axis. The angle between the 0 0 l and the 0 k 0 axes defines the angle α of the unit cell, between h 0 0 and 0 0 l defines β, and between h 0 0 and 0 k 0 defines γ. As discussed above, the spacing between observed reflections on the precession photograph is the length of the arc between two reflections on the sphere (see Figure 6.15). Indeed, to calcu-

**Figure 6.15** Relationship between spacing of reflections on a precession photograph and the spacing of reflecting planes in a crystal lattice. In precession photography, the film is rolled along a sphere. The spacing between two reflections on the photograph represents the length D of an arc at the surface of the sphere. Knowing the distance r from the crystal to the film, D is related to the diffraction angle θ by trigonometry and Bragg’s law.
late the diffraction angle $2\theta$, we need only measure the distance between two adjacent layer lines on the film $D$, and know the distance between the film and the crystal $r$. We see that

$$D = (2\pi r) \frac{2\theta}{2\pi} = 2\theta$$

(6.19)

Since we are only considering one dimension, we can use Bragg’s law, Eq. 6.5, and it is now easy to calculate each length $d$ of the unit cell by

$$\sin \left( \frac{D}{2\theta} \right) = \frac{\lambda}{2d}$$

or

$$d = \frac{\lambda}{\sin \left( \frac{D}{2\theta} \right)}$$

(6.20)

In addition to these geometric parameters, the diffraction pattern recorded on a precession photograph can also give the symmetry of the crystal and thus the space group to be assigned to the crystal. There are numerous different combinations of conditions to define space groups. We will not attempt to go through all of these, but instead we will give a simple example of how the space group can affect the intensity of the diffraction pattern.

Let us consider the simple one-dimensional crystal with regularly spaced points along the $c$-axis, just as we did to define the von Laue conditions for diffraction. The diffraction pattern will show mirror symmetry according to Friedel’s law, but no additional special conditions are placed on the intensities. In the one-dimensional case, this means that the intensities at $l = n$ will be identical to those at $l = -n$.

Now let us add a simple symmetry element to the crystal, a $2_1$ (two-fold screw) axis parallel to the $c$-axis of the unit cell (Figure 6.16). For each atom $A$ in the unit cell, there will be a symmetry-related atom $A'$ at one-half the unit cell along the $c$-axis, rotated $180^\circ$ relative to $A$. Diffraction from $A$ and $A'$ will be a single resultant reflection. For the $00l$ axis, the reflections result from a projection of $A$ and $A'$ onto the crystal $c$-axis (this will become evident later). We notice that this symmetry relationship defines the phase and intensities of the resulting reflection from atom $A$ and its symmetry mate $A'$ as being identical, and the unit cell appears to be exactly one-half the length of the actual unit cell. The reciprocal relationship of the diffraction pattern to the spacing of lattice points in the crystal defines the diffraction angle, and thus the spacing of the diffracted spots as twice that expected for the actual unit cell when the crystal is aligned perpendicular to the incident beam. In terms of the Miller indices, there is a special condition: Only the $l = 2n$ (the even integer reflections) are observed, while all odd reflections ($l = 2n + 1$) cancel and are unobserved. When extrapolated to a three-dimensional crystal having a $2_1$ axis parallel to the $c$-axis, the symmetry imposes this special condition on the intensities of the reflections only on the $00l$ principal axis. All reflections off this primary axis will show intensities at all values of $l$. 
Figure 6.16 Systematic absences caused by a \( 2_1 \) axis along the crystallographic c-axis. A set of atoms \( A_n \) are spaced by a distance \( c \) along the vertical axis. A two-fold screw axis generates a set of symmetry related atoms \( A' \) that are rotated \( 180^\circ \) and translated by \( c/2 \) relative to the atom \( A \). The resulting scattered beam appears to come from a unit cell that is half the length of the actual unit cell. The corresponding diffraction angle will be twice that expected from the unit cell and, therefore, the reflections along the principal axis \( 0 \ 0 \ l \) will be spaced twice the distance expected. This appears as the absence of reflections at all odd values of \( l \) along this axis in the Weissenberg photograph.

We see that this is an exclusionary condition. That is, the special condition tells us that no reflections should be observed for the odd reflections along the \( 0 \ 0 \ l \) axis. It does not, however, guarantee that a reflection will be observed for all even \( l \) indices along this axis; thus, the observation that has no odd \( l \)-indexed reflections along \( 0 \ 0 \ l \) missing defines a \( 2_1 \) symmetry axis that is parallel to c-axis of the unit cell. A crystal in the \( P2_1 \) space group will therefore have all the odd-indexed reflections along one of the principal axes missing. If the odd-indexed spots along an orthogonal axis are also missing, then there is an additional orthogonal \( 2_1 \) axis. Notice that this additional \( 2_1 \) symmetry axis automatically defines a third \( 2 \) or \( 2 \) symmetry axis that is orthogonal to the first two. Thus the monoclinic \( P2_1 \) space group is extended to an orthorhombic \( P2_12_12_1 \) or \( P2_12_12_1 \) lattice.

Each different space group specifies its own unique set of special conditions for observed and unobserved reflections along the principal and diagonal axes. For the space group \( P2_12_12 \), reflections are permitted at \( h = 2n, k = 2n, \) and \( l = 2n \) (i.e., all even \( h, k, \) and \( l \) reflections) along the principal axes \( h \ 0 \ 0 \ k \ 0, \) and \( 0 \ 0 \ l, \) respectively. At the odd indices along these axes, the intensities of the reflections are zero;
these reflections are **systematic absences** in the diffraction pattern. Thus, the space group for a particular crystal can be determined by testing the pattern of observed and unobserved diffraction reflections in the diffraction pattern against the special conditions, as specified in Hahn (1989).

### 6.5 SOLVING MACROMOLECULAR STRUCTURES BY X-RAY DIFFRACTION

The model that we developed above places a single atom, and thus a single Bragg reflecting plane, within the unit cell of a crystal. The crystals that we are interested in will certainly have more than a single atom in a unit cell (upwards of 10,000 for a hemoglobin crystal). How does this affect the treatment of diffraction by Bragg's law or the von Laue conditions? What can we say about the crystal morphology? Let us consider the more complicated situation of two atoms in a unit cell. In the Bragg law formulation, we simply space pairs of reflecting planes in a regular array to represent two atoms in the unit cell of a crystal. The reflections from each pair of planes can be summed to form a new resultant wave that can be considered a single reflection from a single plane. Since each pair of planes is regularly spaced within the crystal, the relative phasing of the summed reflection is the same for each pair of reflecting planes. Thus the diffraction angle is the same as before but related to the regular spacing of the reflecting pairs. Stated differently, the diffraction angle of a crystal is inversely proportional to the length of the unit cells in a crystal, regardless of the number of atoms that are in the unit cell.

However, the phase of the resultant wave reflected from the pair of reflecting planes may not be the same as that of the incident radiation, and the phase from each individual plane that makes the pair may not be that of the resultant wave. In addition, the amplitude of the resultant wave may not be identical to that of the incident radiation. Similarly, the atoms in a macromolecule affect the amplitude and phase of the unit cell reflections. This information gives us the structure of the molecule. However, in order to determine the positions of the atoms in the molecule (that is, to solve the structure of the crystal), we must deconvolute each reflection into the phase and amplitude contributions from each of the individual reflections from each atom in the molecule. (As we will see later) is the primary problem in trying to solve the structure of molecules in a single crystal by X-ray diffraction.

To solve the structure of a molecule within the crystal, we must determine the elemental type and position of each atom of that molecule. For a crystal of a biological molecule, this generally includes solving the structure of the ordered solvent molecules as well as the macromolecule. Defining the atom type means that we determine the number of electrons for that atom, while defining the *atomic position* means that we determine the \((x, y, z)\) coordinates of each atom in the unit cell. We will see later that in reality, this involves determining the electron density at a position in the unit cell. The most straightforward definition of atomic position is to use a set of \(x, y,\) and \(z\) values in the standard orthogonal Cartesian coordinate system. However, this becomes cumbersome when the unit cell is not orthogonal. A more
convenient method is to define \((x, y, z)\) coordinates in terms of fractions of the unit cell lengths. *Fractional cell coordinates* are the fractional distances of atoms between the origin and the next unit cell, and therefore have values between 0 and 1. The position of an atom in the standard coordinate system (with \(nm\) as the unit of length) can be defined as \((xu, yv, zu)\).

To understand how both the position and type of atoms can be determined, we must return to the beginning of our discussion on X-ray diffraction and relate these two properties to the amplitudes and the positions of the reflections in the X-ray diffraction pattern.

### 6.5.1 X-Ray Scattering in Reciprocal Space: The Scattering Vector

Before we discuss X-ray diffraction at the atomic level, we need to define some new terms. To simplify the discussion, we will concentrate on orthorhombic and higher-symmetry crystal lattices. The von Laue equation, Eq. 6.17, says that we will observe reflections at scattering angle 2\(\theta\) for certain values of \(h/a, k/b,\) and \(l/c\). In the notation of reciprocal space, these certain values are written as \(a^*, b^*,\) and \(c^*\), where for orthorhombic and higher symmetries \(a^* = 1/a, b^* = 1/b,\) and \(c^* = 1/c\). That is, in reciprocal space the unit cell axes are \(a^*, b^*,\) and \(c^*\) with lengths \(a^*, b^*,\) and \(c^*\) in 1/mm (as opposed to \(nm\)). The general relationships between the reciprocal crystal lattice and the crystal lattice in real space for monoclinic and triclinic lattices are described in Table 6.3.

The beauty of the reciprocal space concept is that, with the general relationships, the von Laue equation (Eq. 6.18) can be rewritten as

\[
\lambda(h^2a^2 + l^2c^2 + l^2a^2)/2 = 2 \sin \theta = \frac{n\lambda}{d_{AB}} \tag{6.21}
\]

which is valid for all crystal symmetries. Furthermore, it is easy to visualize Bragg’s Law. In Figure 6.17, we denote the lattice of a two-dimensional crystal in reciprocal space; we simply plot their reciprocal axes \(a^*\) and \(c^*.\) We can define some point \(O\) that lies at the origin of the unit cell and intersects the incident X-ray beam. Placing a point \(A\) along the path of the incident X-ray, the distance from \(A\) to \(O\) is \(1/n\lambda.\)

From this point \(A,\) a circle is inscribed with a radius length of \(1/n\lambda.\) Finally, placing a point \(B\) diametrically opposed to the origin defines the three reference points from which we can define the lengths and angles for scattering in reciprocal space.

When the incident X-rays are in a direction that causes scattering, other reciprocal lattice points will lie on the circumference of the circle. We have placed a lattice point \(L\) in Figure 6.17 at a lattice point that intersects the circle. The distance \(OL\) between the reciprocal lattice points causing scattering is \(1/d\) in the Bragg notation. Since \(OLB\) is inscribed in a semicircle, it must be a right angle. Then, by definition, \(\sin(OLB) = \sin(OAB) = (1/d)/(2n\lambda)\) where \(\theta\) is defined in Figure 6.17 as the angle \(OLB.\) But after rearranging, this is simply Bragg’s Law \(n/\lambda = 2d\sin \theta.\) Thus \(\theta\) is the Bragg angle and \(BL\) lies on a Bragg reflection plane. We also see that \(OL\) is the vec-
### Table 6.3: Relationship between Unit Cell Parameters in Real Space and Reciprocal Space

<table>
<thead>
<tr>
<th>Lattice type</th>
<th>Real Space</th>
<th>Reciprocal Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orthorhombic</td>
<td>a ( a^* = \frac{1}{a} )</td>
<td>a ( a^* = \frac{1}{a} )</td>
</tr>
<tr>
<td>and higher</td>
<td>b ( b^* = \frac{1}{b} )</td>
<td>b ( b^* = \frac{1}{b} )</td>
</tr>
<tr>
<td>symmetry</td>
<td>c ( c^* = \frac{1}{c} )</td>
<td>c ( c^* = \frac{1}{c} )</td>
</tr>
<tr>
<td>( \alpha = 90^\circ )</td>
<td>( a^* = \frac{a}{90^\circ} )</td>
<td>( a^* = \frac{a}{90^\circ} )</td>
</tr>
<tr>
<td>( \beta = 90^\circ )</td>
<td>( b^* = 90^\circ )</td>
<td>( b^* = 90^\circ )</td>
</tr>
<tr>
<td>( \gamma = 90^\circ )</td>
<td>( c^* = 90^\circ )</td>
<td>( c^* = 90^\circ )</td>
</tr>
<tr>
<td>V</td>
<td>( V^* = \frac{1}{V} = a^*b^<em>c^</em> )</td>
<td>V ( V^* = \frac{1}{V} = a^*b^<em>c^</em> )</td>
</tr>
<tr>
<td>Monoclinic</td>
<td>a ( a^* = \frac{1}{a \sin \beta} )</td>
<td>a ( a^* = \frac{1}{a \sin \beta} )</td>
</tr>
<tr>
<td></td>
<td>b ( b^* = \frac{1}{b} )</td>
<td>b ( b^* = \frac{1}{b} )</td>
</tr>
<tr>
<td></td>
<td>c ( c^* = \frac{1}{c \sin \beta} )</td>
<td>c ( c^* = \frac{1}{c \sin \beta} )</td>
</tr>
<tr>
<td>( \alpha = 90^\circ )</td>
<td>( a^* = \frac{90^\circ}{90^\circ} )</td>
<td>( a^* = \frac{90^\circ}{90^\circ} )</td>
</tr>
<tr>
<td>( \beta = 180^\circ )</td>
<td>( b^* = 180^\circ - \beta )</td>
<td>( b^* = 180^\circ - \beta )</td>
</tr>
<tr>
<td>( \gamma = 90^\circ )</td>
<td>( c^* = 90^\circ )</td>
<td>( c^* = 90^\circ )</td>
</tr>
<tr>
<td>V</td>
<td>( V^* = \frac{1}{V} = a^<em>b^<em>c^</em> \sin \beta^</em> )</td>
<td>V ( V^* = \frac{1}{V} = a^<em>b^<em>c^</em> \sin \beta^</em> )</td>
</tr>
<tr>
<td>Triclinic</td>
<td>a ( a^* = \frac{bc \sin \alpha}{V} )</td>
<td>a ( a^* = \frac{bc \sin \alpha}{V} )</td>
</tr>
<tr>
<td></td>
<td>b ( b^* = \frac{ac \sin \beta}{V} )</td>
<td>b ( b^* = \frac{ac \sin \beta}{V} )</td>
</tr>
<tr>
<td></td>
<td>c ( c^* = \frac{ab \sin \gamma}{V} )</td>
<td>c ( c^* = \frac{ab \sin \gamma}{V} )</td>
</tr>
<tr>
<td>( \alpha = 90^\circ )</td>
<td>( \cos \alpha^* = \frac{\cos \beta \cos \gamma - \cos \alpha}{\sin \beta \sin \gamma} )</td>
<td>( \cos \alpha^* = \frac{\cos \beta \cos \gamma - \cos \alpha}{\sin \beta \sin \gamma} )</td>
</tr>
<tr>
<td>( \beta = 90^\circ )</td>
<td>( \cos \beta^* = \frac{\cos \alpha \cos \gamma - \cos \beta}{\sin \alpha \sin \gamma} )</td>
<td>( \cos \beta^* = \frac{\cos \alpha \cos \gamma - \cos \beta}{\sin \alpha \sin \gamma} )</td>
</tr>
<tr>
<td>( \gamma = 90^\circ )</td>
<td>( \cos \gamma^* = \frac{\cos \alpha \cos \beta}{\sin \alpha \beta} )</td>
<td>( \cos \gamma^* = \frac{\cos \alpha \cos \beta}{\sin \alpha \beta} )</td>
</tr>
<tr>
<td>V</td>
<td>( V^* = a^<em>b^<em>c^</em> \sqrt{1 - \cos^2 \alpha^</em> - \cos^2 \beta^* - \cos^2 \gamma^* + 2 \cos \alpha \cos \beta \cos \gamma} )</td>
<td>V ( V^* = a^<em>b^<em>c^</em> \sqrt{1 - \cos^2 \alpha^</em> - \cos^2 \beta^* - \cos^2 \gamma^* + 2 \cos \alpha \cos \beta \cos \gamma} )</td>
</tr>
</tbody>
</table>


The distance between Bragg planes is called the scattering vector, and is denoted by S. In Figure 6.17, S has components \( a^* \) and \( 5c^* \), so that \( S = a^* + 5c^* \). In three dimensions, for the general Bragg plane \((h k l)\),

\[
S = ha^* + kb^* + lc^* = (ha^*, kb^*, lc^*)
\]  

(6.22)
Thus the components of $S$ are simply any value of $(ha^*, kb^*, lc^*)$ that describes a reflection. There is a scattering vector for each reflection. Also, the length of $S$ is $||S|| \approx 1/d_{hkl}$. Finally, $AL$ in Figure 6.17 must be the direction of the scattered X-rays in real space.

We notice that the angle $OAL$ is $2\theta$. Using this angle, we could derive the equations for the von Laue conditions for diffraction (which is left as an exercise for the student at the end of the chapter). When extended to three dimensions, the scattering vector $S$ is the intersection between the vertices of the reciprocal lattice and a sphere having a radius of $n\lambda$ (Figure 6.18). Thus, each vector intersects the surface of the sphere at a single point, defining the Ewald sphere or the sphere of reflections in reciprocal space for each value of $n$. We see that the scattering vectors are discrete, each specified by a unique set of Miller indices $h$, $k$, and $l$. As $n$ increases, the volume of the sphere increases and encompasses the reflections that intersect the surfaces of the Ewald spheres at lower values of $n$. Rotating the crystal allows a different set of lattice points to intersect with the sphere to cause scattering. Thus reflections will come into the sphere and exit the sphere, allowing us to measure reflections from various parts of reciprocal space.

In an X-ray diffraction experiment, the intensity of each reflection is given by the intensity of a single scattering vector $I(S)$. All the information for the structure
of the molecule is in \( l(S) \), and it is the decomposition of this information that is essential for solving the structure of the crystal. To understand the process of and the problems inherent in solving the structure of a macromolecule from single-crystal X-ray diffraction, we must understand how the molecular structure defines the measured quantity \( l(S) \).

6.5.2 The Structure Factor

Thus far we have determined the dimensions and the shape of the unit cell by examining the X-ray diffraction pattern. If the molecule consists of a single scatterer at the origin (i.e., at each lattice point) of the unit cell, we have solved the structure. However, a macromolecule has many atoms located far from the origin. To treat the scattering from a macromolecule, we will start by describing diffraction from a discrete scatterer at the origin, see how the diffracted X-ray is affected when the atom is displaced from the origin, and finally treat the scattering from multiple atoms.

As we recall, a point placed in the path of a wavefront acts as the origin of a new wavefront. Thus if we treat light as a wave, an atom that scatters X-rays can be thought of as a point of origin for the scattered wave (Figure 6.19). If the atom is at the origin of a unit cell, then we can describe the observed amplitude of the scattered
Figure 6.19 Propagation of waves. A point placed at the origin $O$ of the unit cell propagates a wave with a maximum amplitude $E_o$ (a). At some point $x$, the instantaneous amplitude is observed as $E_i$. If the atom is displaced from the origin by a distance $r_1$, the amplitude of the wave is observed to be different from that propagated from the origin (b). The wave propagated from a second scatterer at a distance $r_2$ from the origin will have an observed amplitude $E_s$. The wave resulting from both scatterers has an amplitude that is the sum of the two waves ($E = E_i + E_s$), which is dependent on the phase difference $\Delta\phi$ between the two scatterers.

The wave $E$ at some point along an axis $x$ and some time $t$ distant from the origin by the standard equation for the propagation of a wave as a cosine function (see Chapter 8 for a more detailed treatment of light).

$$E = E_o \cos \left(\frac{2\pi}{\lambda} \left(ut - \frac{x}{\lambda}\right)\right)$$  (6.23)
The equation for a wave that is shifted in phase by some fraction of a wave \( \phi \) is

\[
E = E_0 \cos 2\pi \left( \frac{\nu t - x}{\lambda} + \phi \right)
\] (6.24)

For molecules consisting of many scattering atoms, we would need to sum the amplitudes of all waves. To make the addition of multiple waves mathematically easier, we will reformulate Eq. 6.24 in terms of a complex function. We can accomplish this by using the relationship

\[
A \cos x + iA \sin x = Ae^{ix}
\] (6.25)

where \( i = \sqrt{-1} \). Equation 6.25 can then be rewritten as

\[
E = E_0 e^{2\pi i (\nu t - x/\lambda)}
\] (6.26)

where it is understood that we will always take the real (physically meaningful) part of the function.

We will now displace the atom from the origin by a distance \( r \). This distance is simply the length of the vector that defines the atom position relative to the origin of the unit cell, \( r_i = |r| = (x_1a + y_1b + y_1c) \). Equation 6.26 for the first atom becomes

\[
E_1 = E_0 e^{2\pi i (\nu t - x/\lambda - r)}
\] (6.27)

Notice that the amplitude observed at any position \( x \) and at any time \( t \) is some fraction of the maximum intensity of the incident X-rays, \( E_0 \).

Now, let us place a second identical atom at some distance \( r_2 \). It acts as a single point of propagation for a new wave having an amplitude

\[
E_2 = E_0 e^{2\pi i (\nu t - x/\lambda + r)}
\] (6.28)

The relative positions of the two atoms in space can be defined as a single variable \( \Delta \phi = r_2 - r_1 \), thus Eq. 6.28 can be rewritten as

\[
E_2 = E_0 e^{2\pi i (\nu t - x/\lambda + r_1 + \Delta \phi)} = E_0 e^{2\pi i \Delta \phi}
\] (6.29)

The observed amplitude for the scattering from the two atoms is simply the sum of the two waves,

\[
E = E_1 + E_2 = E_0 (1 + e^{2\pi i \Delta \phi})
\] (6.30)

We see that the amplitude of the summed waves is increased or decreased by a factor related to the difference in positions of the two atoms, which results in a phase shift between the two summed waves. For two waves that are exactly in phase, \( \Delta \phi = 0 \), and \( E = E_0 (1 + e^{i \phi}) = 2E_0 \). That is, the two waves add and show constructive interference. For two waves that are out of phase by 180°, \( \Delta \phi = 1/2 \), and \( E = E_0 (1 + e^{i \phi}) = E_0 (1 + \cos \pi + i \sin \pi) = 0 \). These waves exactly cancel from destructive interference. The observed amplitudes will generally fall somewhere in between these two extremes.

If the two atoms are different types of elements, each atom will have a different number of electrons occupying a given volume in space (this is the electron density at that point in space). The intrinsic amplitude of the scattered X-ray from each
type of atom, $E_a$, is dependent on the electron density of the scatterer. The higher the electron density at any point in space, the higher the amplitude of the scattered light. We can include this into the equation for wave propagation by including an atomic scattering factor $f$ for each atom. In crystallography, a structure factor $F$ is defined as the amplitude of each scattered beam observed at specific values of the Miller indices $(h k l)$ that is analogous to Eq. 6.29 and includes all atoms in the unit cell,

$$F(hkl) = F(S) = \sum_{j=1}^{N} e^{2\pi i \mathbf{S} \cdot \mathbf{r}_j}$$

(6.31)

where $N$ is the number of atoms in the unit cell and

$$\mathbf{S} \cdot \mathbf{r}_j = (h a^* + k b^* + l c^*)(x_j + y_j + z_j) = h x_j + k y_j + l z_j$$

(6.32)

We see that, by analogy to $E$, the equation for the structure factor $F(S)$ is composed of two parts. The atomic scattering factor $f_j$ defines the maximum amplitude of the scattered X-ray if that atom is placed at the origin of the unit cell ($\phi = 0$) and is dependent only on the type of atom that is the scatterer. This is a real number. The exponential factor contains the information relating the relative phase of the scattered wave, and is a complex number. Thus we make use of Eq. 6.25 and write Eq. 6.31 in terms of the real $(re)$ and imaginary $(im)$ parts.

$$F(hkl) = F(S) = \sum_{j=1}^{N} e^{2\pi i \mathbf{S} \cdot \mathbf{r}_j} = \sum_{j=1}^{N} [f_j \cos(2\pi S \cdot r_j) + i \sin(2\pi S \cdot r_j)]$$

(6.33)

We can visualize the real and imaginary components of any complex number by plotting these components on orthogonal axes in an Argand diagram. This can be done for each reflection with Miller indices $(h k l)$. Summing the components over all atoms $j$ in Eq. 6.33 gives the net $F_{re}$ and $F_{im}$ of $F(S)$. These are shown in Figure 6.20, where we see that the net phase $\phi$ is the angle between $F(S)$ and $F_{re}$.

In addition to being a convenient method to illustrate the relationship between the real and imaginary components of the structure factor, we will later use the Ar-
and diagram to demonstrate how the phase angle for a macromolecule can be estimated by measuring the structure factors of crystal modified by heavy atoms. Since the two components of \( F(S) \) are orthogonal, they can be treated independently and can be summed separately, as in Eq. 6.33.

So far, we have treated the diffracting elements as fixed and discrete points in space. As discussed above, X-rays are scattered by electrons, and we know from quantum mechanics that electrons should be treated as a probability distribution in space. This means that X-ray scattering is dependent on the electron density \( \rho \), the number of electrons per unit volume. At any point in the unit cell \( \mathbf{r} \), there will be an electron density, \( \rho(x, y, z) = \rho(\mathbf{r}) \). If more electrons occupy a given volume of space, that volume will show stronger diffraction. For the volume element \( dV = dx\,dy\,dz \), we can write the structure factor in terms of electron density

\[
F(S) = \rho(\mathbf{r}) e^{2\pi i \mathbf{S} \cdot \mathbf{r}} dV
\]  

(6.34)

If we consider the space in a crystal as having a continuous probability of finding electrons, then the structure factor is described by integrating over the volume

\[
F(S) = \int V \rho(\mathbf{r}) e^{2\pi i \mathbf{S} \cdot \mathbf{r}} dV
\]  

(6.35)

This is essentially equivalent to Eq. 6.33 for discrete atoms, because positions where atoms lie are positions that will show the highest probabilities of finding electrons. Similarly, heavier atoms have higher numbers of electrons for a given volume in space. The \( e^{2\pi i \mathbf{S} \cdot \mathbf{r}} \) factor depends on the phase of the scatterers, and so depends on the positions of those scatterers in space. Equations 6.31 and 6.35 tell us that if we know the type and positions of each scatterer relative to the origin of the crystal (that is, if we know the structure of the molecule), we can calculate \( F(S) \). This means that we can see how well the final model of the macromolecular structure in the crystal fits the observed X-ray diffraction data. Returning to the original question, how do we determine the structure of the molecules from the structure factor?

We recognize Eq. 6.35 as a Fourier series. In this case, the structure factor, which is a function of the scattering in reciprocal space, is written in terms of the electron densities in real space. A diffraction pattern of any kind (whether from X-ray or electron diffraction) is a representation of reciprocal space and looks nothing like the objects scattering the radiation. The Fourier transform as given in Eq. 6.36 gives us a function that describes the electron density at any particular point in real space in terms of the scattering vector in reciprocal space

\[
\rho(\mathbf{r}) = \frac{1}{V} \int V^* e^{-2\pi i \mathbf{S} \cdot \mathbf{r}} F(S)
\]  

(6.36)

where \( V^* \) is the volume element in reciprocal space, and \( V \) is the volume of the unit cell.

Since diffraction occurs in a regular pattern and, according to the von Laue conditions, at discrete points, the electron densities can be calculated from a sum of
the $F(S)$ for all Miller indices $(h k l)$. This is a Fourier series for discrete reflections as opposed to the continuous Fourier transform of Eq. 6.36.

$$\rho(r) = \frac{1}{NV} \sum_{h=1}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F(h k l) e^{-2\pi i r \cdot S}$$

(6.37)

When the electron densities calculated from Eq. 6.37 are displayed for the $(x, y, z)$ positions in the crystal unit cell, we generate an electron density map for the unit cell of the crystal (Figure 6.21). This map shows the positions and the types (the number of electrons for a given volume) of each atom in the asymmetric unit of the crystal.

A convenient method for interpreting structural information from an electron density map is to plot the map as a set of contours, as in a geographical topography map (Figure 6.21a). Each set of concentric contours represent peaks of electron density, normally centered at an atom or at groups of atoms, depending on the resolution. The map can be constructed to present the electron density of the unit cell in three-dimensional space by first drawing contours representing the identical levels

![Figure 6.21](image)

**Figure 6.21** Electron density maps. The electron density calculated from the Fourier transform in Eq. 6.37 can be represented by (a) a contour map or as (b) a set of chicken wires. In this figure, the electron density of the heme binding pocket of myoglobin is shown. In (a), four sections of the contour map are overlapped to show the electron density at the heme and the surrounding amino acid residues. An enlarged view of this same set of electron densities are shown in (b) as surrounding the model of the heme (solid lines).
of electron density in one plane of the unit cell. This plane represents one section of the map. The three-dimensional unit cell is sectioned along the axis perpendicular to the contour plane. When the sections are stacked, we generate a three-dimensional view of the electron density in the unit cell. A molecular model can thus be derived from the electron density map by connecting the centers of each peak with lines representing the chemical bonds in the molecule. With the advent of three-dimensional computer graphics, such contouring techniques are being supplemented by a simpler “chicken wire” representation of the electron density map (Figure 6.21b). Both types of representations have their utility in solving a macromolecular structure.

Equations 6.31 and 6.35 tell us that all atoms and their associated electrons contribute to the structure factor of each observed reflection, and all observed reflections must therefore be used to completely determine the types and positions of each atom in space. Because we will never in practice be able to measure all the possible reflections, we must actually truncate the Fourier series in Eq. 6.37 at some finite values of \((h \ k \ l)\). This means we can never have a completely accurate description of the contents of the unit cell. If we can measure the structure factors for a large number of scattering vectors \((S)\), then we can readily calculate the electron density at any point in space and thus solve the structure of the crystal. There is a problem, however. Although \(F(S)\) can be calculated from a known set of atomic positions, it cannot be directly measured in a diffraction experiment. We can measure the intensity of the diffracted X-ray for each reflection, and the structure factor is related to the intensity; but, as we will see in Section 6.6.3, some critical information is lost in the intensities that precludes directly solving the atomic structure of a crystal from the observed X-ray diffraction data.

### 6.5.3 The Phase Problem

We showed above that the structure factor can be derived starting with a wave description of light propagating from two or more scattering points. Knowing the structure factor, we can calculate the electron density at any position in the crystal unit cell and thus solve the structure of the molecules within the crystal. However, there is a problem in measuring the structure factor. We have devices that detect both the amplitude and phase for light of long wavelengths. In Chapter 12, we discuss nuclear magnetic resonance spectroscopy and show how the Fourier series measured in the microwave region can be transformed directly into the NMR spectrum. Similarly, we see in Chapter 9 that infrared absorption can be detected as a Fourier series that can be transformed directly into the infrared spectrum. Unfortunately, the devices that we have available to detect short-wavelength light measure total energy. This will depend on the time period of energy collection and the energy per unit time, which we call the intensity \(I\) of light. The total energy collected will be proportional to the intensity, and it can be shown (for instance, see Halliday et al.) that the intensity of a light wave is proportional to its amplitude \(E\), squared. Thus we have the amplitude information for each structure factor, but we have lost the phase information. Referring back to the Argand diagram in Figure 6.20, we know the magnitude of the...
structure factor \( |F(S)| \), but not its direction. In terms of the complex notation of Eq. 6.33

\[
|S(S)| = |F(h k l)|^2 = |F(S)|^2 = |F(S)\rangle F^*(S) = F_{\alpha r} + F_{\delta r}^*
\]

where \( F^*(S) \) is the complex conjugate of \( F(S) \), which is

\[
F^*(S) = \sum_{j=1}^{N} e^{-2\pi i x j k} = F_{\alpha r} - iF_{\delta r}
\]

How critical is this loss of information? If we go back to the original equations for \( F(S) \) (Eqs. 6.31 and 6.35), we see that the phase information is \( \phi = 2 \pi S \cdot \mathbf{r} \) for the diffracted X-rays. We recall that the quantity \( S \cdot \mathbf{r} = (h x + k y + l z) \). Thus the information corresponding to the absolute positions of the atoms within the unit cell are lost, leaving only the part of \( F(S) \) that relates to the number and types of scatterers translated to the origin of the unit cell of the crystal. This becomes clear when we calculate the electron density maps using only \( |F(h k l)| \) from the X-ray diffraction data or using only the phase of \( F(S) \) (see Figure 6.22). We see that in this situation, it is impossible to directly solve the structure of the crystal using the information from \( \sqrt{I(S)} = |F(S)| \). A simple analogy in mathematics is to take \( \sqrt{4} = |x| \). In this case, \( x = \pm 2 \). We know the magnitude of \( x \), but do not know its direction. This is known as the *phase problem* in crystallography.

Why bother with X-ray diffraction if we apparently cannot directly solve structures using this method? Obviously, the technique would not be useful and would

![Figure 6.22](image)

Electron density calculated from the two components of \( F(h k l) \). In (a), the electron density of a DNA crystal was calculated using only \( |F(h k l)| \) from the X-ray diffraction data. The map does not fit the model of the DNA structure, but resembles the pattern expected for the Patterson function. In (b), the same map was calculated using only the phase information for \( F(h k l) \) with \( \langle n(h k l) \rangle \) set at 1.0 for all reflections. The resulting map very closely resembles the dC-dG base pair in the structure. This demonstrates the importance of the phasing information over the magnitude of the structure factor.
not have become such a widely used and powerful technique (and we would not spend a full chapter discussing the method) if there were not ways around this problem. First of all, we should emphasize that the phase information for each atom is contained within the intensity data \( I(S) \) collected, or else the diffraction pattern would have identical scattering intensities at all \( (h k l) \). It is the net phase of \( F(S) \) that cannot be readily derived directly from \( I(S) \). In fact, each \( F(S) \) can be calculated knowing the atoms in the molecules and their positions in the unit cell, since this gives the individual contributions to \( F(S) \), as in Eq. 6.33, which can be summed to give \( F(S) \) and the phase of each reflection. In other words, if we have an approximate model for the molecular structure in the crystal, we can simulate an approximate X-ray diffraction pattern. This is one of the keys to solving the structure of a crystal.

In the following sections, we will discuss the general methods for solving the phase problem, and thus for solving crystal structures. These include molecular replacement, direct methods, and isomorphous replacement. In molecular replacement, a known structure of a related molecule is used as the initial model for the unknown structure. This method has been generally useful for solving the structures of various nucleic acids and of structural variants of known protein structures. This method, however, will not work for a completely unknown structure.

Direct methods use a number of tricks to solve the phase problem, either by trying all possible phase combinations for each \( S \) and simply finding that combination that best fits the overall data to solve the structure, or by using the phase information for each atom inherent in the intensity data to retrieve some information concerning the relative positions of atoms in the crystal (this is known as the Patterson method). These methods are currently useful only for directly solving the structures of small molecules (on the order of 100 to 300 nonhydrogen atoms) because of the exponential growth in the phase problem as the size of the molecule increases. We will discuss the Patterson method in some detail, since it is a useful first step for the more general method of isomorphous replacement.

In isomorphous replacement, specific atoms in the crystal are tagged with heavy atoms, atoms that have very high electron densities and therefore can strongly perturb the X-ray diffraction pattern. Once the positions of these heavy atoms are located within the crystal (using direct methods or the Patterson method), the overall phase of the original molecule can be estimated. This is a general technique for solving the structures of completely unknown structures of macromolecules including nucleic acids, proteins, and complexes of these.

At the end of this section, we will discuss some of the newer methods for solving the phase problem, including anomalous dispersion with synchrotron sources.

**Molecular Replacement.** As we have shown above, once we have a model for a structure, we can calculate \( F(S) \) for all values of \( (h k l) \) for that structure. If the structure that we are trying to solve is very similar to one that has been previously determined, we can simply use those elements within the known structure that are common to both to generally “phase the data.” This is perhaps the simplest method for solving the structure of a macromolecule, and it has been applied generally to
solve structures as small as oligonucleotides and as large as virus particles. There are a number of specific criteria that must be met before molecular replacement will work. The first is that the previously determined structure must be very similar to the one we are trying to solve. If they are dissimilar to any great degree, the phasing information will not be sufficient to give good electron density maps. This limits the types of crystal problems that are amenable to this method. One class of problems for which molecular replacement is an appropriate method is the determination of the structure of a mutant protein from the structure of the native protein, or the structure of homologous proteins from different species. Another class of problems for which molecular replacement has been successfully used is in solving the structures of double-helical oligonucleotides. In this latter case, the chemical and physical properties of the molecules are very well understood from other experimental methods. In addition, the regularity of the structures allows reasonable models to be built entirely from the symmetry of the structures (see Chapter 1).

In these situations the number of atoms that differ between the known and unknown structures is small, while the number of those that are identical both chemically and in spatial location is large. The common regions supply both the atomic scattering factors (from the chemical formula) and the phase information (from the atomic coordinates of the atoms) for a large number of scatterers. In the case of a mutant protein structure solved from a native structure, the two proteins may differ by less than one amino acid out of a hundred. Thus the calculated structure factors can be very accurate. We can simply exclude the atoms that are not shared between the known and the unknown structures and calculate an electron density map for the regions that differ. This is known as a difference electron density map (or an omit map), since atoms that are dissimilar are omitted from the F(i) calculation (see Figure 6.23). The quality of the difference map depends on the accuracy of the phases calculated from the starting model. With a good difference map, however, we can use the residual electron density to build those parts of the unknown structure to fit the observed densities.

The other constraint to molecular replacement is that the model must sit in the correct orientation and absolute position within the unit cell. If both of these criteria are not met, even a perfect model will not properly phase the diffraction data. This is particularly critical when using a model derived from other methods (e.g., from structures determined by NMR), or from a structure having very different unit cell parameters. Obviously, the simplest case is one where the model will be used to phase the new data comes from a structure that sits in exactly the same orientation and position within the crystal and the crystals are isomorphous.

When the simplest case does not hold, then the starting model must somehow be placed in the new unit cell and properly positioned. This is normally accomplished by using a series of rotation and translation functions to "fit" the model to the electron density. Again, the success of this method depends on how closely the initial structure matches the unknown structure.

The Patterson Function. Contained within the observed intensities of the X-ray diffraction data is the phase information of the individual atoms, and thus in-
formation on the relative positions of the atoms in the unit cell. In trying to solve the crystal structure, why do we not simply use the observed intensities to construct a Fourier series that will be some function of the atomic positions? This is the basis for the Patterson function \( P \). Using Eq. 6.38, the Patterson function can be defined as

\[
P(x, y, z) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} [F(h, k, l)]^2 e^{-2\pi i (hx + ky + lz)}
\]

(6.40)

Interestingly, if the transform of \( F(S) \) is \( \rho(r) \), the transform of \( F^*(S) \) is \( \rho(-r) \). The resulting Patterson map has the form,

\[
P(x, y, z) = \sum_{j=1}^{N} \sum_{k=1}^{N} \rho_{j}(r_{k}) \delta(-r_{k})
\]

(6.41)

The Patterson function has peaks in a Patterson contour map that correspond to the vector difference between the atomic positions (Figure 6.24).

We can see from this that there is a loss of information as we go from atoms in real space to the vector differences derived from the Patterson function. A very real indicator of this lost information is found in the symmetry of a Patterson map. The symmetry of a crystal unit cell can be described by one of 230 different space groups. The symmetry of a Patterson map, however, is limited to only 24 space groups. These 24 Patterson space groups can be generated by first removing all the translational elements of the symmetry operators from the original crystal space group. Thus, two-fold screws become two-fold rotational axes, and so on. In addition, all Patterson maps are centrosymmetric, therefore, there is always a symmetry axis at the origin.

What does this mean in terms of the positional information available from the Patterson function? The physical description of this function is that each peak is
Figure 6.24 Patterson maps of two atoms in a unit cell. (a) Two unique atoms $A$ and $B$ in a unit cell are displaced from the origin by the distance vectors $r_A$ and $r_B$. (b) The Patterson map for the two atoms shows two cross vectors, one for the vector from $A$ to $B$ and the other from $B$ to $A$. The two self-vectors ($A$ to $A$ and $B$ to $B$) result in two contours at the origin of the map. (c) Two additional atoms, $A'$ and $B'$, are generated in a crystal with two-fold rotational symmetry. Although there are still only two unique atoms, there are now four additional cross vectors. (d) This is a Harker section in the Patterson map, the additional cross vectors are $2r_{A}$ and $2r_{B}$. This allows us to determine $r_{A}$ and $r_{B}$ (or the atomic coordinates of $A$ and $B$) directly from the Patterson map.

Patterson map corresponds to a distance vector separating two atoms. Consider the simple case where two atoms $A$ and $B$ are located at positions $(x_A, y_A, z_A)$ and $(x_B, y_B, z_B)$ in the unit cell of a crystal (Figure 6.24). For $i = A$ and $j = B$ in the Patterson function, the difference vector $(r_i - r_j) = (r_A - r_B) = (x_A - x_B, y_A - y_B, z_A - z_B)$. Similarly, for $i = B$ and $j = A$, $(r_i - r_j) = (r_B - r_A)$. Finally, for $i = A$ and $j = A$, and for $i = B$ and $j = B$, the distance vectors are zero, placing two vectors at the origin of the Patterson map. This means that if we plot the Patterson function in a manner similar to an electron density map, there will be three peaks: two corresponding to a distance vector separating $A$ from $B$ and $B$ from $A$, called the cross vectors, and the third corresponding to the self-vectors at the origin. The magnitudes of the peaks are related to the product of the electron density of the two atoms separated by the distance vector. If two different vectors are coincident, the Patterson peaks are additive. The relative magnitudes would thus be $1:1:2$ for the peaks at $((x_A - x_B, y_A - y_B, z_A - z_B), (x_B - x_A, y_B - y_A, z_B - z_A), and$
We see that for a molecule having a large number of atoms, the self-vectors at the origin very quickly become the dominant feature of a Patterson map, and the number of cross-vectors grows as the square of the number of atoms.

There are a number of important concepts that must be stressed in understanding the utility and the limitations of the Patterson function. First and foremost, the function defines relative distances and not absolute positions. However, this does not prevent us from using the Patterson method to solve structures. Consider an atom \( A \) that sits at the end of a distance vector \( r \) from a two-fold symmetry axis. There would be a symmetry-related atom \( A' \) at some distance \(-r\) from the axis. The distance separating \( A \) from \( A' \) is \( r - (-r) = 2r \). Thus, in the Patterson map, there would be a single vector projecting from the origin with a length of \( 2r \). To use this information in determining the absolute position of an atom in the unit cell, we must find a section of the Patterson map that reflects the symmetry of the unit cell. The Patterson map itself does not necessarily assume the symmetry of the unit cell, but certain sections (called the Harker sections) reflect some of the symmetry elements of the original space group of the crystal (Figure 6.24). For space groups containing rotational or screw symmetry axes, these occur at the origin of \( a, b, \) and \( c \), or some fraction of the unit cell along \( a, b, \) and \( c \).

The usefulness of the Harker section is that distance vectors for symmetry-related atoms lie in this plane. For example, if we have a single atom \( A \) at some position \((x_A, y_A, z_A)\), and there is a two-fold screw parallel to the crystallographic \( a \)-axis, then a symmetry related atom \( A' \) must sit at a position \((x_A + 1/2, y_A, z_A)\). The difference vector relating \( A \) and \( A' \) is \((1/2, 0, 0)\). Thus at this Harker section in the \( b-c \) plane at \( 1/2 \) along the \( a \)-axis of the unit cell, there will be a Patterson peak corresponding to the coordinates \((2y_A, 2z_A)\). The absolute coordinates \( y_A \) and \( z_A \) of atom \( A \) can thus be determined directly from the Patterson peak in the Harker plane. Notice, however, that we cannot determine the coordinate \( x_A \) from this Harker section. This requires additional information, which could come from the Patterson peaks in a perpendicular Harker section if this is allowed by the symmetry of the unit cell.

Why can we not simply use the Harker planes to determine the coordinates in three-dimensional space of all the atoms in a molecule? For a single atom in the asymmetric unit, this would be a trivial problem, and for a small number of atoms it is a tractable problem. However, as the number of atoms increases, so do the number of symmetry-related atoms and correspondingly the number of Patterson peaks in the Harker sections. At some point it becomes impossible to resolve the individual peaks in the Patterson map, even from three perpendicular Harker sections. Thus, the Patterson method is really useful only for locating a small number of atoms within the unit cell. How can this be a useful method in solving structures of macromolecules? We will see in the next section that by determining the positions, and thus the phases, of a few particular atoms, the overall phase of the protein can be estimated.

For crystals of nucleic acids, there are special situations where the Patterson map can go a long way in solving the structure of the molecule. Because the base pairs of the DNA duplex are stacked along an axis, the large number of redundant
atom-atom distances between base pairs will tend to dominate the Patterson map. In this case, we can specifically determine the orientation of the base pairs relative to the crystal axes, determine the helical rise of the structure (and thus the general conformation of the molecule), and with some luck the position and orientation of the molecule within the unit cell (Figure 6.25). This can provide a starting model that can readily be used to solve the structure of all the atoms in the molecule by molecular replacement.

**Multiple Isomorphous Replacement.** The final method that we will discuss in detail for solving molecular structures (particularly those of macromolecules) is isomorphous replacement. The *isomorphous* part of this method indicates that we will be using crystals whose unit cells are nearly identical in shape and size and are in identical space groups. The *replacement* part of the method indicates that we will be substituting, or in many cases adding heavy atoms to the molecules in the unit cell. The general method is as follows.

1. The X-ray diffraction data for a crystal of a protein or nucleic acid is collected. This is called the *native* data set, since it comes from the unmodified crystal. The structure factors from the native crystal will be called $F_{\text{p}}$, from the protein crystallographer's nomenclature, although it can apply to crystals of any biological molecule.

2. Isomorphous crystals are obtained of the identical molecule in which a strongly diffracting atom (a heavy atom, such as a metal or a halogen) is attached at a specific location. The X-ray diffraction data for these crystals are
the heavy atom derivative data sets. The structure factor for the modified crystals will be called $F_{\text{HN}}$, since it reflects the structure of the starting native molecule and the heavy atom(s) in the crystal.

3. A derivative data set is appropriately scaled to the native set, and a difference data set representing the scattering from only the heavy atom is obtained by subtraction. The structure factors of the difference data $F_{\text{HN}}$ reflect only the scattering of X-rays by the heavy atom.

4. The $F_{\text{HN}}$ are used to determine the positions and the phases of the heavy atom in the unit cell.

5. This process is repeated for at least one additional heavy atom derivative.

6. The phases of at least two heavy atom derivatives are used to estimate the phase for the native data set to solve the structure of the macromolecule in the native crystal.

We start this discussion with the isomorphous heavy atom crystals. There are three criteria for obtaining a successful heavy atom derivative. First, the procedure should not significantly affect the structure of the crystal or the molecules within the crystal (that is, the crystals must remain isomorphous with the native crystal). Second, the modifications must be specific. This means that we know what amino acid or nucleic acid residue is being modified. Finally, the number of heavy atoms within the asymmetric unit should be small. This makes it easier to locate the positions of the heavy atoms in the unit cell.

There are a number of methods for specifically introducing one or more heavy atoms into a macromolecular structure and maintaining the characteristics of the native crystal. The classical method is to start with the native crystal of the unmodified macromolecule and soak metals or other atoms into the crystal (Application 6.3). A list of common derivatives that are specific for certain residue types in proteins and polynucleotides are listed in Table 6.4. Any dramatic changes in the morphology of the crystal becomes obvious if the crystal cracks during the soaking procedure. More subtle modifications either to the crystal morphology or the macromolecular structure are indicated when the unit cell parameters and space group are determined from the X-ray diffraction pattern of the modified crystal.

Another method for introducing heavy atoms is to make specific modifications to the molecule prior to crystallization. Many proteins that are being crystallized are recombinant proteins (their genes have been cloned and their gene products produced in a system for expressing proteins). In these cases, it is straightforward to introduce a mutation that, for example, replaces a serine residue at the surface of the protein with a cysteine. This introduces a site that can readily bind mercury at a specific site. An alternative method for introducing a heavy atom at a very specific site in the protein is to express the polypeptide with selenium-methionine in place of the standard sulfur-containing amino acid. This substitution does not appear to significantly affect the structure or function of most proteins.

In the case of synthetic DNA or RNA fragments, brominated or iodinated nucleotides (5-bromocytosine or 5-iodouridine) can be readily incorporated into the sequence during synthesis. Such strategies have been used to solve the structure of oligonucleotides and complexes of oligonucleotides with proteins or drugs.
Application 6.3 The Crystal Structure of an Old and Distinguished Enzyme

The enzyme urease, which catalyzes the hydrolysis of urea, was first isolated in crystalline form by James B. Sumner in 1926. It was in fact the first enzyme to be crystalized, demonstrating that enzymes were chemical compounds with distinct identities. It was also identified as the first enzyme to utilize nickel in catalysis. The crystal structure of urease, however, was not solved until 1995 in the laboratory of P. Andrew Karplus (Jabri et al. 1995), some 70 years after the first crystals were obtained. The enzyme for structural analysis was isolated from the bacterium *Klebsiella aerogenes*, rather than jack bean. However, the two forms of the enzyme are homologous, with more than 50% of their sequences being identical.

The structure was solved by multiple isomorphous replacement, with 40 heavy atom compounds screened to find five workable derivatives of the crystal (Table A6.3). The phasing of the diffraction data was aided by anomalous scattering from the heavy atoms and incorporation of selenomethionine into the protein by growing bacteria in media containing the selenized amino acid. The resulting structure showed the enzyme quaternary structure to be a trimer of heterotrimeres. Three complexes of the α, β, and γ subunits associate to form a triangular unit (Figure A6.3). In addition, an unusual, although not unique, carbamylated lytine was observed to bridge the two nickels in the active site. The structure raises some interesting questions, including how this active site evolved. The structure of the catalytic site shows great similarity to that of adenosine deaminase, a zinc-containing enzyme. The two enzymes also share some similarities in their catalytic mechanisms, including a metal-coordinated water that attacks the amide carbon of the substrates, and the tetrahedral intermediate that releases an ammonia to form the product. This is a clear case of divergent evolution of the catalytic sites dictated by the requirements of the substrate rather than convergent evolution driven by the requirements of the catalytic mechanism.


### TABLE A6.3 CRYSTALLOGRAPHIC DATA AND RESULTS FOR UREASE

<table>
<thead>
<tr>
<th>Crystal</th>
<th>Resolution of Data</th>
<th>Number of Unique Reflections</th>
<th>Final R factor</th>
<th>Nonhydrogen Protein atoms</th>
<th>Solvent Molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native</td>
<td>2 Å</td>
<td>58,134</td>
<td>18.5%</td>
<td>6002</td>
<td>215</td>
</tr>
<tr>
<td>Apo-enzyme</td>
<td>2.8 Å</td>
<td>20,532</td>
<td>18.4%</td>
<td>5944</td>
<td>157</td>
</tr>
<tr>
<td>HOEtGcCH4CO3Na</td>
<td>3.3 Å</td>
<td>11,027</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EsCh</td>
<td>3.3 Å</td>
<td>12,210</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOEt(Ch3COO)2</td>
<td>2.5 Å</td>
<td>28,709</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C(CH3)OCCH3</td>
<td>2.4 Å</td>
<td>29,672</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(CH3)3PBr(Ch3COO)</td>
<td>2.4 Å</td>
<td>23,496</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Se-Met</td>
<td>3.5 Å</td>
<td>20,332</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data from Jabri et al. (1995).
Figure A6.3 Quaternary structure of utnase from the bacterium Klebsiella aerogenes. The structure of the protein consists of a heterodimer of α, β, and γ subunits to form a unit with a molecular weight of 83,000 mg/mol. Three of these units then associate to form a larger trimer.

<table>
<thead>
<tr>
<th>TABLE 6.4 HEAVY ATOM DERIVATIVES FOR MACROMOLECULAR CRYSTALS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Heavy atom</strong></td>
</tr>
<tr>
<td>AgNO₃</td>
</tr>
<tr>
<td>K₂PdI₂ (Br or Cl)</td>
</tr>
<tr>
<td>Hg acetate</td>
</tr>
<tr>
<td>p-chloromercuric benzene sulphoneitate (PCMBS)</td>
</tr>
<tr>
<td>Se</td>
</tr>
<tr>
<td><strong>Nucleic Acids</strong></td>
</tr>
<tr>
<td>Cu</td>
</tr>
<tr>
<td>Pt</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>Br</td>
</tr>
</tbody>
</table>
Once a native data set and one or more derivative data sets have been collected, a difference data set is derived. The structure factors of the reflections in this difference data set are defined exclusively by the heavy atoms in the isomorphous crystals. The heavy atoms can be located within the crystal by the Patterson method described above.

At this point, we can see how all this information comes together to give an estimate of the phases for the native data set. We start with the native data set, from which we can calculate \( F_p \) for each reflection. Notice that since we do not have the phase information for this data, we only know their magnitude \( |F_p| \). In an Argand diagram, we can define the length of each \( F_p \) vector but not its angle; therefore, all possible phases for the structure factor are represented by a circle circumscribed by the vector \( F_p \) (Figure 6.26).

For one heavy atom derivative, we can similarly define the magnitude but not the phase of each \( F_{pH} \). This vector defines a second circle in the Argand diagram. The difference here is that the origin of the second circle does not coincide with the origin of the \( F_p \) circle. It is shifted by a vector, and that vector is \( F_H \). Since the positions of the heavy atoms are determined, both the magnitude and the direction of \( F_H \) are known and the relative positions of the two Argand circles can be found. This shift in the origin allows the second circle to intersect the first. The two intersection points represent the vector sum of \( |F_p| + |F_{pH}| \) to give \( |F_{pH}| \). Thus, these points define two possible phases for each structure factor \( F_p \) (Figure 6.26). Data for an additional isomorphous derivative must be obtained to resolve the ambiguity in the phase angles. Obviously, the more heavy atom derivatives that are used, the more accurate will be the estimate for the phases of the native data set.

Using these estimated phases for the native data set, an approximate electron density map can be drawn. A rough model for the structure of the macromolecule

![Figure 6.26](image-url) Estimating phases from multiple isomorphous replacement. The magnitude of the structure factors for the native protein \(|F_p|\) and one heavy atom derivative \(|F_{pH}|\) each define circles for all possible phases of the reflections. The structure factor \( F_{pH} \) of the heavy atom derivative shifts the two circles relative to each other. The intersection of the two circles defines two possibilities for \( F_p \).

A second heavy atom derivative \( F_H \) is used to distinguish between the two possible phases for \( F_p \).
can be constructed by connecting regions of continuous high electron density in this initial approximate electron map. In proteins, the initial features that become obvious are the regular helical structures, and these are fit by placing a polypeptide chain into the well-defined regions of the electron density map. This process is known as chain tracing. As additional structural information is added to this initial model, the phasing information improves and thus the electron density map improves. Once a majority of the backbone has been properly traced, the side chains of the larger amino acids will become evident in the improved maps. At this point, the primary structure of the molecule becomes an invaluable tool for completing the structure.

**Structure Refinement.** The general methods described to this point yield an initial model that grossly fits the measured diffraction data. It is clear that the quality of the calculated electron density map is strongly dependent on the quality of the data and on the model. Short of growing new crystals and collecting better diffraction data, we cannot change the quality of data. However, we can improve the model by refining the structure to better fit the electron density. In this refinement process, the model undergoes subtle changes to improve its details. We might think that we could simply fit the model visually to the calculated electron density maps. However, there is a problem with electron density maps that are calculated using the structure factors derived from the observed intensities \( F(hkℓ) \). We recall that only \( |F(hkℓ)| \) can be determined from the intensity data; the phasing information that we used in \( |F(hkℓ)| \) was calculated from the model, \( F_{\text{calc}} \). This means that the model strongly affects the maps.

A more objective criterion for evaluating how well a model fits the data is the R factor

\[
R = \frac{\sum |F(hkℓ)| - |F_{\text{calc}}|}{\sum |F(hkℓ)|} \tag{6.42}
\]

In this case, only the magnitudes of the observed and calculated structure factors are compared.

Generally, R factors range between 70% (0.7) for a random fit to a value approaching 0% for an ideal fit (Figure 6.27). An ideal fit is not achievable, primarily because of the limitations inherent in the quality of the data and imperfections in the crystal. For macromolecules, a general rule is that a model refined to an R factor less than 20% (0.2) indicates a good fit, and thus can be presented as being a solution to the structure. To achieve this level of correlation between the model and the data, the structure of the macromolecule and the structure of a large portion of the solvent (water, ions, etc.) must be well defined. Unfortunately, an artificially low R factor can be achieved for a poor model by simply adding more solvent to account for the electron density in the unit cell. This is a case of overfitting the data.

We see that solving the structure of a macromolecule from X-ray diffraction data is highly dependent on the construction of molecular models and on the quality of these models. There are a number of steps where model bias can creep in, but several objective criteria are available to evaluate the quality of the resulting models.
Other Methods for Phasing X-Ray Diffraction Data. Of the three methods that we have discussed, multiple isomorphous replacement by heavy atoms is the most common technique for solving a previously undetermined macromolecular structure. The usual method for determining the positions of the heavy atoms is from difference Patterson maps. However, it is possible to determine the positions of heavy atoms by other means. For example, the difference data set derived by subtracting the native data from derivative data should reflect the structure of only the heavy atoms. This is simply a problem of determining the positions of a small number of atoms within a large unit cell and can, in theory, be solved by direct methods. However, in practice, direct methods approach is limited to very high-quality difference data.

An alternative approach to determining the positions of heavy atoms is to take advantage of the properties inherent to these atoms. Atoms with high electron densities not only scatter X-rays, they also absorb X-rays. When the energy of the radiation used in the diffraction experiment can be absorbed by (that is, is near the absorption edge of) the scattering atom, the assumption that the atomic scattering factor of the phase is no longer valid. This leads to a phenomenon known as anomalous dispersion, which we observe as a breakdown of Friedel’s law — then the difference in intensities between Friedel pairs can be used to determine the phase of the heavy atoms. We will discuss anomalous scattering in some detail, because an extension of this method can be used to estimate the phase of the macromolecule using only a single heavy atom derivative.

Our assumption that $f$ is phase-independent comes from treating the scatterers as unperturbed electrons. In reality, the interaction of light with any particle requires some change in the properties of that particle. Classically, an electron that scatters X-rays will be induced to vibrate or oscillate. When the wavelength of the radiation approaches the absorption edge of an atom, the electron will resonate at its absorp-
tion frequency $\nu$. This is not significant for the light atoms that constitute the majority of scatterers in macromolecules, but it becomes important for heavy atoms, such as those used in isomorphous replacement. For these heavy atoms, $f$ can be treated as the sum of three terms: $f_0$ (the atomic scattering factor for a spherically distributed electron), and the perturbations $f'$ and $i f''$ to the real and imaginary components of the atomic scattering factor. Thus, for these atoms even the atomic scattering factor includes a component that will affect the phase of the scattered radiation.

The correction terms affect each member of the Friedel pair differently.

$$f_+ = f_0 + f' + if''$$

(6.43)

and

$$f_- = f_0 - f' - if''$$

(6.44)

for the atom scattering factors of $F(hkI)$ and $F(-h-k,-l)$, respectively. As we can see from the Argand diagram in Figure 6.28, $f'$ and $if''$ are symmetric about the real axis while $f''$ are symmetric about the imaginary axis. Thus, $f_+$ is not equal to $f_-$, and $F(hkI)$ is different from $F(-h-k,-l)$. These differences in intensity are small, but can be easily measured with the very sensitive X-ray detectors used today. An accurate determination of the anomalous scattering for a heavy atom derivative can thus lead to estimates for $f'$ and $if''$, which can lead to the determination of the phase and position of the heavy atom. With this, the standard methods of multiple isomorphous replacement can be used to estimate the phases of the macromolecule.

Recently, the phenomenon of anomalous dispersion has been successfully extended in a method called multiple-wavelength anomalous dispersion (MAD) to allow the phasing of macromolecules using only a single heavy atom derivative. MAD takes advantage of the dependence of $f'$ and $if''$ on the wavelength of the incident radiation. The closer the wavelength of this radiation is to the absorption edge of the scattering atom, the stronger the anomalous dispersion. Thus two different wavelengths result in two different values of $if''$, which in turn gives us two different pieces of phasing information from the heavy atom. This is the same as having two independent heavy atom derivatives. If the protein being studied is a metalloprotein.
(such as the copper-containing protein hemocyanin), MAD can be applied to the native crystal and the diffraction data can be phased using the inherent heavy atom. However, the phase information is not as strong as with two derivatives with truly different atomic coordinates. The phasing power of MAD can be improved by using multiple wavelengths. This requires an X-ray source that is tunable to a wide spectrum of wavelengths, which currently limits MAD to data collected at synchrotron sources.

6.5.4 Resolution in X-Ray Diffraction

We started this chapter by describing X-ray diffraction from single crystals as the only reliable method for determining structures of biological molecules to atomic resolution. At this point, it is obvious that diffraction is not the same as microscopy. We can thus ask the following questions: What does resolution mean in X-ray diffraction, and can all crystal structures be determined to atomic resolution? Since we must transform incomplete data, the structures obtained are models of the molecules in the crystals and not the molecules themselves. The resolution of a structure is thus dependent on the data used to solve the structure. We will show that resolution in diffraction is related to the diffraction angle of the observed data.

The theoretical limit to resolution using X-ray diffraction is inherent in Bragg's law of diffraction, Eq. 6.5. We remember that resolving a structure to atomic resolution means that we can determine the unique positions of the closest spaced atoms. In its simplest conception, this would be the ability to resolve crystals in which the unit cell contains a single atom and these unit cells have a spacing d of about 0.1 nm. We can therefore use d as a measure of the resolution limit in a diffraction experiment as (d_{40})_{max}. Bragg's law tells us that d decreases as sin θ increases for a given wavelength of radiation, λ. We discussed resolution earlier in terms of the Heisenberg uncertainty principle. Bragg's law says that the best we can do in X-ray diffraction is n = 1 and sin θ = 1 to give d = λ/2. This occurs at θ = 90°. Thus data collected at higher diffraction angles are higher resolution, with the maximum resolution at θ = 90°. The same logic also applies to the von Laue equations, Eq. 6.17. The two wavelengths typically used in X-ray crystallography are the Kα bands emitted when electrons bombard a copper or a molybdenum anode (Table 6.5). Since λ = 0.154 nm for X-rays from copper radiation, the best resolution that can be obtained from this source is d = 0.077 nm, while for molybdenum radiation d = 0.036 nm.

<table>
<thead>
<tr>
<th>Table 6.5 X-ray Radiation and Resolution</th>
<th>CuKα</th>
<th>MoKα</th>
</tr>
</thead>
<tbody>
<tr>
<td>λ</td>
<td>0.15418 nm</td>
<td>0.07107 nm</td>
</tr>
<tr>
<td>(d_{40})_{max} = 1/2</td>
<td>0.07709 nm</td>
<td>0.03554 nm</td>
</tr>
</tbody>
</table>

Now we ask whether all single crystals will provide a sufficient number of data points to resolve a structure to atomic resolution. This depends on several factors that are intrinsic to the instrumentation for collecting diffraction data, as well as to the crystal itself.

Since the highest resolution data corresponds to a Bragg angle $\theta = 90^\circ$ (or a diffraction angle relative to the incident beam of $2\theta = 180^\circ$), this data is reflected directly back to the X-ray source; therefore the highest resolution data cannot be collected. The practical limit in terms of instrument design is $2\theta = 110^\circ$, which corresponds to $\sin \theta = 0.82$, or $d = 0.048 \text{ nm}$ resolution for CuK$_\alpha$ radiation. This is the wavelength generally used in solving the structures of biological macromolecules.

Why do we not use molybdenum radiation, which has a higher theoretical resolution limit of $d = 0.043 \text{ nm}$? The problem lies in a different resolution problem, that is, in the ability to resolve closely spaced reflections in the X-ray diffraction pattern. Although we have treated the reflections from X-ray diffraction as single points, in reality they are normal distributions representing the quantum mechanical probability that a photon will be observed at a particular point in space. This means that each reflection has a width in real space, usually at least 1° across at half-height.

The separation between two reflections must therefore be greater than 1° for us to be able to distinguish them as discrete points. This places a limitation on the longest edge of the unit cell that can be studied using either Cu or Mo radiation. For example, a unit cell with a maximum length of 5 nm (a small protein), will have reflections separated by about 2° (as estimated from Bragg’s law for CuK$_\alpha$ radiation). That same crystal using Mo radiation shows reflections separated by only about 0.8°, which would not allow the reflections to be resolved. Thus crystals of biological molecules, with their large unit cell dimensions, require the use of the longer wavelength of X-rays, which is typically CuK$_\alpha$ radiation.

We have shown above that atomic resolution depends on the diffraction angle of the reflections used to determine the parameters for each atom in the crystal unit cell. Now we ask: How much data is required to obtain atomic resolution? The number of reflections $N$ is equal to the number of reciprocal lattice points contained in the volume ($V^{-1}$) of the sphere of reflections (recall that the sphere of reflections is defined in reciprocal space and therefore its volume is in units of $1/\text{nm}^3$).

$$N = \left(\frac{4}{3}\pi \left| \mathbf{S} \right|^2 \right) / V^{-1}$$

or

$$N = \left(\frac{4}{3}\pi \right) V / d^3$$  \hspace{1cm} (6.45)

in real space.

Finally, let us ask: At what point do we have enough data to solve a structure to atomic resolution? To solve a structure, we must assign values to four parameters for each atom in the molecule. To solve the structure to atomic resolution, these parameters must be determined independently. This means that at least one reflection must be collected for each atomic parameter.
The atomic parameters that are determined from X-ray diffraction are the $(x, y, z)$ coordinates of the atom, and some indication of the type of electron distribution at this position. This fourth parameter is often the temperature factor, or $B$ factor. The $B$ factor was originally introduced as a measure of the thermal motion of the atom. For a given atom of $e$ electrons, a higher $B$ factor indicates that these electrons occupy a larger volume. Assigning the properties of atoms to a single parameter assumes that the thermal motions of atoms are isotropic. In fact, the vibrations of atoms in a molecule are anisotropic, and thus the best model for a molecule should include the thermal motion in all three directions in space ($B_x$, $B_y$, and $B_z$). This is observed only in the highest resolution structures, because it dramatically increases the number of parameters required to define an atom. For macromolecules, the $B$ factor is usually treated as being isotropic.

In addition to thermal motion, the $B$ factor also reflects how often an atom is positioned at a particular point in space. This is the partial occupancy of an atom at a specific position. For example, an atom may be very rigidly held in place and thus occupy a very well-defined volume in space. However, if it sits in two different positions in different molecules in the crystal, the occupancy at any one position represents only half the electrons of the atom. This results in a lower partial occupancy and a higher $B$ factor for that atom. In practice, the $B$ factor is a measure of both thermal motion and partial occupancy, which reflects the overall disorder of the atom.

Thus, we see that at least four unique reflections are required to define the parameters of each atom in the asymmetric unit. For 330 atoms, which occupy a volume of about 6 nm$^3$, this translates to a minimum of $N = 1440$ reflections. From Eq. 6.45, we see that this number of reflections can be collected at $d = 0.28$ nm or at $2\theta = 34^\circ$. This diffraction angle is significantly lower than we would expect for atomic resolution from Bragg's law. Our calculation, however, greatly overestimates the number of reflections that can be used to solve a molecular structure because not all the reflections in the sphere of reflections are unique.

The number of unique data points that can be collected for a specific resolution limit is restricted by the symmetry of the diffraction pattern. For example, Friedel's law tells us that $F(h k l) = F(\bar{h} \bar{k} \bar{l})$, indicating that even for a crystal with the lowest symmetry ($P1$), only half the data (or one hemisphere of the sphere of reflections) is unique. Additional symmetry may reduce this even more. For example, for a crystal with $V = 25$ nm$^3$ (as seen in crystals of deoxyribonucleotides), we would expect 104,720 reflections at 0.1 nm resolution, of which 52,360 would be unique for a $P1$ lattice. For the higher symmetry lattice $P2_12_12_1$, only 13,090 would be unique. This larger number of data points for a low symmetry space group of a given unit-cell volume is required because there are fewer asymmetric units contained in this lattice and, thus, more atoms that need to be solved within the asymmetric unit. Fortunately, for the same volume, molecules in higher symmetry lattices require fewer reflections to solve their structure because of the symmetry relationships among the molecules. The two effects (the number of unique reflections and the number of atoms in the asymmetric unit) just balance. Indeed, a crystal with $V = 25$ nm$^3$ in the $P2_12_12_1$ lattice would have four asymmetric units in the unit cell, with each typically
containing about 350 atoms. In the P1 lattice, there would be a single asymmetric unit containing more than 1400 atoms. Thus, for a P222, lattice containing 350 atoms, there are potentially 13,090 unique reflections or 9 per parameter at \( d = 0.1 \) nm, if all the reflections are reliably collected to this level.

What prevents us from solving all single crystal structures to atomic resolution? We are first limited by the quality of the crystal. Thus far we have treated crystals as exact repeats of unit cells. In these unit cells, the molecules are rigid and have identical conformations. We showed very early on in this chapter: that any deviation from this ideal case would necessarily reduce our ability to resolve the atoms of the structure. In fact, the crystals of macromolecules are not perfect. The molecules are not entirely static, and are not exactly identical across all unit cells. This often results in parts of molecules being unresolved, even if most of the structure is well defined and of high quality. Thus, the structure that is solved has been averaged over the unit cells of the crystal, and the resolution to which a structure can be solved depends on just how close the unit cells are to being identical. This is the primary limitation to the resolution of the data that can be collected. We can work the logical backwords and show that this loss in resolution due to time or spatial-averaging of the structure will result in fewer reflections observed at high diffraction angles. Indeed, any amount of disorder will tend to broaden and reduce the intensity of all the reflections observed from the crystal. To minimize thermal disorder, crystallographic data can be collected at low temperatures (to liquid-nitrogen temperatures).

The resolution of the reflections are also limited by the size of the crystal. We can simply state that the larger a crystal is, the more repeating scatterers there are in the path of the X-ray beam. This translates into higher intensities for all reflections, and thus more data that can be collected. This brings us back to the initial problem in X-ray crystallography, that is, to growing crystals. Crystallographers will go to great lengths to grow larger and better crystals, including going into space. Crystals grown in the microgravity environments of the U.S. space shuttle or the Russian MIR space station are typically more ordered and up to 50% larger than those grown on Earth. Unfortunately, it is unlikely that extraterrestrial crystallography will become a routine method of growing macromolecular crystals in the near future.

6.6 FIBER DIFFRACTION

A single crystal has all its molecules and the atoms in the molecules arranged in well-ordered and repeating arrays. Can X-ray diffraction provide structural information on molecules that do not form ordered crystalline arrays? The answer is yes, if there is inherent symmetry within the molecule itself to cause constructive and destructive interference of the diffracted X-rays. We observe this in long helical biopolymer fibers. We discuss fiber diffraction here, first for historical reasons. The structural features of the helices of biopolymers (the \( \alpha \)-helix from Linus Pauling’s work and the DNA duplex from Franklin and Wilkins as interpreted by Watson and Crick) were first determined from fiber diffraction studies. Second, the method still
is applicable for providing low-resolution structures of macromolecular complexes that are not amenable to single-crystal X-ray diffraction, including prions and sickle-cell hemoglobin fibers.

We can think of the molecules in oriented fibers as being aligned and therefore ordered in one dimension. If the molecules have a regular secondary structure, then the inherent symmetry of the helices will generate repeating units that are aligned along the fiber axis. We can therefore treat an exact repeat of the helix (see Chapter 1) as a crystalline unit cell with well-defined lattice points. The molecules within the fiber, however, are randomly rotated relative to each other. Thus, the unit cells are rotationally disordered and averaged across the width of the fiber. When treating fiber diffraction in the same manner that we do single-crystal diffraction, we first need to investigate the unit cell, the general properties of the diffraction pattern from this unit cell, and the effect of the individual residues on the diffraction pattern. We can then see how the diffraction pattern can provide us with information on the structural parameters of the helix, including the helical symmetry, pitch, and radius.

In fiber diffraction, the repeat of the helix produces layer lines along the long axis of the fiber (the z-axis), like that of a one-dimensional crystal (Figure 6.12). These are the meridional reflections (Figure 6.29). The molecules in a fiber may also

![Figure 6.29](image)  The fiber diffraction photograph of B-DNA. The diffraction photograph of the lithium form of a DNA fiber (recrystallized at 46% humidity) shows the helical X expected for helical structures and 10 layer lines spaced according to n/P in nm⁻¹ between the origin and the exact repeat of the pattern. This indicates that the fiber is B-DNA. [Courtesy of R. Longridge].
pack in a semicrystalline array around the diameter to produce a set of layer lines that lie perpendicular to the z-axis. These are the equatorial reflections. In the remainder of this section, we will see how the meridional reflections provide information on the pitch and symmetry of a helix, while the equatorial reflections give us the radius of the helix.

### 6.6.1 The Fiber Unit Cell

The packing of the symmetric units (helices) in the fiber of a biopolymer is essentially the packing of infinitely long cylinders. The regular and repeating unit is not a box defined by a parallelepiped but a series of stacked repeating cylinders. It is more convenient, therefore, to define the unit cell in cylindrical coordinates (Figure 6.30). In this case, the unit cells are all aligned relative to a single axis (the helix axis, which is defined as the z-axis). Any point within this unit cell has coordinates $r$, $\phi$, and $z$, representing the distance from the helix axis, the rotation about the helix axis, and the vertical distance along the helix axis. Since diffraction from fibers must conform to Bragg's law for constructive and destructive interference, these coordinates can be treated in reciprocal space by defining the reciprocal lattice parameters $R$, $\Psi$, and $Z$ to describe the scattering vector, $s$ (Figure 6.30). Since the helices exactly repeat along the z-axis, the parameter $Z$ will define the spacing of layer lines in the diffraction pattern in reciprocal space, just as the length of the c-axis of a unit cell defines the spacing of layer lines along the c*-axis in reciprocal space for the one-dimensional crystal in Figure 6.11.

Because the fibers are not crystalline, there are very few reflections in a fiber diffraction pattern. Although there is far from enough information to determine the positions of atoms in the helix, we can show how detailed information on the fundamental structure can be derived for a helical biopolymer by discussing the type of patterns that we expect for different types of helices, and comparing these to patterns observed for actual helices of polypeptides and nucleotides.

We will start by deriving the diffraction pattern for a continuous helix based on the diffraction from a line that is wrapped around the stacked cylindrical unit cells of the fiber (Figure 6.31). However, the helices of biopolymers are not continuous.

![Figure 6.30 Cylindrical unit cell in real space and in reciprocal space. A point in a cylindrical unit cell is defined by the parameters $r$ for the vertical length, $\phi$ for the radius, and $\psi$ for the angle of the point from the z-axis. This defines a vector $s$ for the position of a point relative to the origin of the unit cell. In reciprocal space, the analogous parameters are $R$, $\Psi$, and $Z$ to define the scattering vectors $s$ relative to the origin.](image-url)
They contain discrete residues (amino acid, nucleotide, or saccharide building blocks) that are regularly spaced along this continuous helix. A more representative fiber diffraction pattern for this discontinuous helix will then be derived by introducing a set of spaced lattice planes that are regularly spaced like the residues along the helix axis.

6.6.2 Fiber Diffraction of Continuous Helices

To derive the fiber diffraction pattern for a continuous helix, we start by defining the structure factor for the helix. The most general expression for the structure factor in this polar coordinate system \( F(R, \Psi, Z) \) is related to the electron density of an infinite solid cylinder by the Fourier transform,

\[
F(R, \Psi, Z) = \int_0^\infty \int_0^{2\pi} \int_{-\infty}^{\infty} \delta(r, \phi, z) e^{i[R \cos \Psi \sin \phi - \Phi - R \cos \Psi + Z]} dr d\phi dz \quad (6.46)
\]

This is greatly simplified if we assume that the helix is a line wrapped on the surface of the cylinder: Then lattice points occur only at the radius of the helix, \( r = r_H \). At all other values, \( r \neq r_H \) and \( \delta(r, \phi, z) = 0 \). This means that \( F(R, \Psi, Z) \) is independent of \( r \). Using a trigonometric identity for \( \phi \) and \( \Psi \),

\[
F(R, \Psi, Z) = \int_0^{2\pi} \int_{-\infty}^{\infty} \delta(z) e^{i[R \cos \Psi \sin \phi - \Phi + Z]} dz d\phi \quad (6.47)
\]

The unit cell has a length equal to the pitch of the helix, \( P \). Therefore, any point along \( z \) can be described in terms of the fractional unit cell relative to the pitch \( z = n/P \). If we index the reflections along \( Z \) relative to the reciprocal lattice, \( Z = n/P \) (where \( n \) is the index for the Bragg reflection or observed layer line), then the integral simplifies to a form that is dependent only on \( \phi \).

\[
F(R, \Psi, n/P) = \int_0^{2\pi} e^{i[R \cos \Psi \sin \phi - \Phi + n \phi]} d\phi \quad (6.48)
\]
This is similar in form to the Bessel functions, $J_n(x)$, which describe the amplitudes of dampened sine and cosine functions for the variable $x$.

$$J_n(x) = \left(\frac{1}{2\pi}\right) \int_0^{2\pi} e^{ix \cos \theta + in\theta} d\theta$$  \hspace{0.5cm} (6.49)

For $x = 2\pi n R$, the structure factor has the form

$$F(R, \Psi, \frac{n}{P}) = J_n(2\pi n R)e^{i(\phi + n\psi)}$$  \hspace{0.5cm} (6.50)

The order of each Bessel function, $n$, is equivalent to the index of each layer line of the diffraction pattern (the zeroth-order Bessel function describes the intensity distribution for the $n = 0$ layer line; see Figure 6.32). As the index $n$ increases to higher orders, the maximum amplitude shifts to higher values of $2\pi n R$ (equivalent to the scattering vector $S$). For each value of $n$, the amplitude decreases as $2\pi n R$ increases.

To construct the fiber diffraction pattern for a continuous helix, we start by defining the meridional axis $Z$ and the equatorial axis $R$, remembering that an X-ray diffraction pattern reflects the symmetry of the unit cell in reciprocal space (Figure 6.33). The intensity at each layer line $n$ is calculated as $I = F(r, \Psi, n/P) F(r, \Psi, n/P)^*$ with the spacing between layer lines equal to $1/P$ along the meridian axis. For each layer line $n$, $I_n(s) = I_{-n}(s)$, so the pattern shows mirror symmetry across the meridian axis. Similarly, $I_n(s) = I_{-n}(s)$ so the pattern

**Figure 6.32** Bessel functions, $J_n$

Bessel functions are shifted to higher values of $x$ and are reduced in intensity $J_n^2$ as the order of the function $n$ is increased. The intensity $I(S)$ along the layer lines in a fiber diffraction pattern is proportional to $J_n^2$. 

\[ x = 2\pi n R \]
Figure 6.33 The diffraction pattern of continuous and discontinuous helices. The fiber diffraction pattern of a continuous helix (a) has layer lines spaced by $lP$ and a helix $X$ at an angle $\delta$ relative to the meridian axis. The pattern for a discontinuous helix, having an integral number of residues in an exact repeat, is derived by the intersection of the fiber diffraction from the continuous helix and the reciprocal lattice of the unit cell. The spacing of the lattice lines indicates the position of the repeating $\beta$-pattern. The pattern for the discontinuous helix repeats the helical $X$ at the repeat $c$ of the helix, in this example, $c = 3$ residues per turn for a $3\beta$ helix. The order of the Bessel function at each layer line is defined by $n = l - cm$, where $l$ is the index of the layer line and $m$ is an integer indexing the repeat. The original repeat of the helical $X$ pattern has $m = 0$; the first repeat has $m = 1$, and so on. [Adapted from C. R. Cantor and P. R. Schimmel, Biophysical Chemistry (1980), Part II, W. H. Freeman and Co., New York, p. 799.]

also has mirror symmetry across the equatorial axis. The maximum intensities at all layer lines form the familiar X pattern associated with helical structures (the helical X). A line that connects the maximum intensity of the layer lines forms an angle $\delta$ relative to the meridian axis. For a continuous helix, this $\delta$ is related to the inclination angle that the helix relative to the z-axis is. Indeed, the diffraction pattern of the continuous helix is drawn as a helical X through the origin of the reciprocal lattice, with each arm of the X placed at an angle equal to $\delta$. The intersection of this X with each layer line indicates the position at which we would expect to observe a reflection from the Bessel function.

6.6.3 Fiber Diffraction of Discontinuous Helices

Although the diffraction pattern is regular for a continuous helix, it does not repeat as we would expect for a repeating structure such as a helix of a biopolymer. This arises because continuous helices are dependent only on the pitch and the radius of
the helix. The pattern will not repeat until we define distinct unit cells and repeating lattice motifs within the unit cells that can act as reflecting planes. This is analogous to the repeated lattice motifs of the single-crystal unit cell that act as reflecting planes in Bragg's law.

We will incorporate the exact helical repeat \( c \), as previously defined for a helix in Chapter 1, into the function that describes the spacing of the layer lines, \( Z \). For the continuous helix, \( Z = n/P \). For the discontinuous helix, we define a new index for each layer line \( l \), such that, \( Z = lp \), with \( p \) being a function of \( n \) and \( c \). This means that the \( X \) pattern of the continuous helix will exactly repeat itself along the meridional axis.

There are two distinct situations that must be considered in accounting for the helical repeat. In the simplest case, the helix exactly repeats in one turn (e.g., where the helical symmetry is explicitly \( c \), as defined in Chapter 1). The value of \( l \) for this helix is \( l = cm + n \), where we use \( m \) to index the helical repeat. Thus, we can determine which order Bessel function contributes to each layer line by trying all the possible values of \( m \) that satisfy the equation \( n = l - cm \). Take the simple case of the \( 3_{10} \) helix for a polypeptide chain, which, as we recall from Chapter 1, has \( 3 \) helical symmetry. The Bessel functions that contribute to the \( l = 0 \) layer line are \( |n| = -3m \), which are 0, 3, and 6 for \( m = 0, 1, \) and 2. The intensity distribution at each layer line will be dominated by the Bessel function having the lowest \( |n| \), or \( n = 0 \) for \( l = 0 \) in this case. Similarly, for \( l = 1 \), \( |n| = 1 - 3m \), which are 1, 2, and 5 for \( m = 0, 1, \) and 2. The dominant Bessel function is \( n = 1 \), and the layer line will be spaced at \( Z = 1/P \).

This is no different from what is expected from the continuous helix. However, for \( l = 2 \), the dominant Bessel function is \( |n| = 1, 2, \) and 5 (for \( m = 1, 0, \) and 1). Thus the intensity distribution at \( l = 2 \) is predicted to be exactly identical to \( l = 1 \). We can similarly show that the pattern at \( l = 3 \) is identical to \( l = 6 \). Thus the overall diffraction pattern exactly repeats at \( l = c \). This is presented diagrammatically in Figure 6.33.

If we start with the \( X \) pattern of the continuous helix at \( l = 0 \), we can generate the pattern for the discontinuous helix by simply repeating the helical \( X \) at \( l = c \) and \( l = -c \). These represent \( l = mc \) for \( m = \pm 1 \). We can therefore interpret the index \( m \) as representing the scattering vector for one unit cell repeated along the helical axis both above and below the original unit cell, in reciprocal space. This is analogous to the Miller indices \( (h, k, l) \), which count the number of reciprocal lattices along the \( a \)-, \( b \)-, and \( c \)-axes of the unit cell in a single crystal. If each layer line is spaced as \( Z = 1/P \), then the pattern is repeated at \( c/P = 1/h \) (where \( h \) is the helical rise).

In the more complex situation, the helix does not exactly repeat in a single turn. The \( \alpha \) helix, for example, has a helical symmetry of 18\( \beta \), or 18 residues in 5 turns of the helix of the exact repeat. If we represent the helical symmetry more generally as \( cT \), where \( c \) is the number of residues per exact repeat and \( T \) the number of turns in the exact repeat, it can be shown that the layer lines of the diffraction pattern will be regularly spaced by \( Z = 1/TP \). Each layer line \( l \) will have a pattern of intensities defined by the Bessel functions with order \( n = (l - mc)/T \). Since the order of the Bessel functions must be an integer, the trick is to find values of \( m \) that yields integer values for \( n \) from this equation. In these cases, the diffraction pattern is still a repeat of the helical \( X \) from the continuous helix with the repeat pattern centered at
Figure 6.34 Fiber diffraction pattern for a helix with a nonintegral number of residues per turn. An alpha-helix, which has 3.6 residues per turn, has a helical symmetry of 16. The resulting fiber diffraction is predicted to be the intersection of the pattern from the continuous helix (a) with a reciprocal lattice representing the number of residues in the exact repeat of the helix (18 in this case). The resulting pattern therefore has 18 layer lines, with the spacing between lines equal to \( 1/3P \). In the alpha-helix, this spacing is \( 1/5P = 0.37 \text{ nm}^{-1} \). [Adapted from C.R. Cantor and P.R. Schimmel (1980), *Biophysical Chemistry, Part II*, W.H. Freeman and Co., New York, p. 799.]

The layer line \( l = c \) for \( m = \pm 1 \) and spaced at \( Z = 1/b \). Each successively higher order of the Bessel function (from 0 to 1 to 2, etc.) occurs at \( l = nV \) for the helical \( X \) at \( m = 0 \). In other words, the actual repeating pattern reflects the number of turns of the helix required to exactly repeat the helix (Figure 6.34).

The practical application of these formulas is that the X-ray fiber diffraction pattern can be used to determine the structural properties of a helical fiber, including the helical symmetry and pitch. As with single-crystal diffraction, a reflection will be observed when the reciprocal lattice intersects the sphere of reflections. However, for a fiber, the lattice points lie only along the helix axis. In a typical fiber diffraction study, the helix axis must be tipped in order to observe the repeating pattern along the meridional axis (Figure 6.35). The angle at which the fiber is tipped allows the reciprocal lattice for one repeat to intersect the sphere of reflection to cause scattering. Otherwise, the meridional reflections will not repeat in the diffraction pattern, and we cannot determine the repeating unit of the helix.

The radius of the helix \( r \) can be determined from the packing of the cylindrical unit cells. Since this is perpendicular to the helix axis, the packing of the helices affects the reflections along the equatorial axis. The most efficient packing of cylinders is a hexagonal array (Figure 6.36a). There are two distinct repeating patterns in this
Figure 6.35  Recording a fiber diffraction pattern. The fiber diffraction pattern of a helix represents the intersection of the fiber lattice with the sphere of reflection in reciprocal space. If the fiber is aligned exactly perpendicular to the incident X-ray beam, the exact repeat of the lattice may not intersect the sphere, and therefore the diffraction pattern may not show the layer line for the exact repeat along the meridional axis. The fiber must be tipped to bring the repeat of the lattice onto the sphere to record the repeating pattern.

(a)

(b)

Figure 6.36  Equatorial reflections. The equatorial reflections of a fiber diffraction pattern result from the semicrystalline packing of the helices along the diameter. (a) The helices are regularly spaced by their radius $r$ and by $\sqrt{3}r$. (b) The reflections along the equator of the fiber diffraction pattern, therefore, are spaced by $1/\sqrt{3}r, 1r, 2/\sqrt{3}r, 2r$, and so forth.
array: one related to \( r \), the other to \( \sqrt{3}r \). Again, since the diffraction pattern reflects the helical parameters in reciprocal space, we would expect to observe reflections along the equatorial axis spaced at \( R = 1/\sqrt{3}r, 1/r, 2/\sqrt{3}r, 2/r \), and so on (Figure 6.36b).

Fiber diffraction has a venerable history. Before crystallography, it was the only way to use X-ray diffraction to study the structures of biopolymers such as nucleic acids. Although the data were severely limited, models for biopolymers were made that were consistent with the data. Of course, this did not prove that the model was correct, but many models proved to be very good, and the method contributed significantly to our understanding of biopolymer structures. For example, the structure of the B-DNA duplex derived from fiber diffraction very accurately describes the mechanisms of semiconservative replication. Still, the only reliable information concerning the helical structure are the pitch, the exact repeat, and the radius. We cannot, for example, determine the handedness of the helix, let alone the position of every atom in space.

**EXERCISES**

6.1 Cryoelectron microscopy can resolve structures to 0.7 nm.
   a. What is the approximate wavelength \( \lambda \) (in nm) and energy (in eVs) of the radiation used in the method?
   b. What type(s) of structure(s) might you be able to resolve using this method (such as helices, domains, subunits of hemoglobin, and so on)?

6.2 Starting with the equation for the molecular structure factor \( F(h) \),
   a. Show that the intensity of reflections along the \((h 0 0)\) axis will be zero for all odd values of \( h \) when a two-fold screw axis is aligned along the crystallographic \( a \)-axis. [Hint: start by placing an atom at a specific set of \( x, y, \) and \( z \) coordinates (in fractional unit cells), and generate its symmetry related atom in the unit cell.]
   b. Show that Friedel's law is correct.

6.3 For an atom with atomic coordinates \( x = 1.51 \) nm, \( y = 0.35 \) nm, and \( z = 0.50 \) nm in a tetragonal unit cell with lengths \( a = b = 20.0 \) nm, \( c = 30.0 \) nm,
   a. Calculate the \( x, y, z \) coordinates in fractional unit cell lengths.
   b. If this is the only unique atom in the unit cell, calculate the phase angle for the \( h = 2, k = 3, l = 4 \) reflection.

6.4 Derive the von Laue condition for diffraction for the two-dimensional reciprocal crystal lattice in Figure 6.17.

6.5 The following is the Patterson map of a molecule composed of identical atoms in a plane. Each contour represents the intensity of the peak at each position.
a. How many atoms are in the molecule?
b. Draw the structure of the molecule, and show how this structure would give the Patterson map shown above.

6.6 The precession photograph in Figure 6.14 was recorded with a crystal to film distance of 7.65 cm (the spacing between reflections was measured to be 0.15 cm). The third dimension for this crystal is 3.777 nm.

a. Calculate the unit cell dimensions of this crystal.
b. Propose one possible space group for this crystal.
c. If the third axis has 4, symmetry, what systematic absences would you expect to see?
d. Estimate the number of unique reflections one would expect for a data set at 0.133 nm resolution.

6.7 You collected X-ray diffraction data from a protein crystal and from an isomorphous crystal of a derivative having a single selenomethionine derivative. Contours for the selenomethionine heavy atom were observed at the Harker sections of the difference Patterson map at \( x = 0.43, y = 0.22, z = 0.5 \) and at \( x = 0.43, y = 0.5, z = 0.35 \) (in fractional unit cell coordinates).

a. What are the atomic coordinates of the heavy atom?
b. Calculate \( F_{\text{p}} \) for the reflection \( h = k = l = 2 \).

6.8 For a \( \pi \)-helix (4.4 residues per turn):

a. Write the helical symmetry of this helix for full repeats.
b. Predict the lowest order Bessel function that would contribute to each layer line starting from 0 to \( l = \) full repeat.
c. Equatorial reflections were observed at 0.020, 0.023, 0.035, and 0.040 nm\(^{-1}\). What is the radius of the helix?
For A-DNA (11 bp/turn, 0.26 nm rise):

- **a.** Predict the two lowest-order Bessel functions that will contribute to layer lines \( l = 0 \) and \( l = 5 \).
- **b.** Calculate the spacing between layer lines in reciprocal space and in real space (assuming a fiber to film distance of 12 cm).
- **c.** The diameter of A-DNA is approximately 2.2 nm. Estimate the spacing of the equatorial reflections in \( \text{nm}^{-1} \).

Sketch the fiber diffraction pattern expected for A-DNA.

**REFERENCES**

**Crystallization of macromolecules**


There are a number of factorial or sparse matrix methods to screen initial crystallization solutions for proteins and nucleic acids. Some are described by:


**X-ray diffraction**

The texts with the most thorough description of the theories and methods of X-ray crystallography:

- **The general methods of crystallography as applied to proteins.**
Fiber diffraction


Some papers that describe recent applications of fiber diffraction to study protein and nucleic acid structures include:


5.13 Both allele a and a' code for a subunit of mass about 30,000 Da. However, the a product is capable of associating (probably to a dimer); the a' product is not. Thus, the heterozygote exhibits both a' monomers and a dimers. The dimer must be rather compact, for the $s$ value of 4.6 is somewhat larger than expected for a mass of $\sim 60,000$. The product could possibly be a trimer, if of extended structure.

**CHAPTER 6**

6.1 $\lambda = 1.4 \text{ nm}, E = 1.42 \times 10^{-19} \text{ J} = 885 \text{ eV}$

6.3 a. $x = 0.076, y = 0.018, z = 0.017$  
   b. $2\pi(0.27) = 97.2^\circ$

6.5 a. 5  
   b. 

6.7 a. $x = 0.22, y = 0.11, z = 0.18$  
   b. $F_R = f_R$

6.9 a. $n = 0; n = 5$  
   b. $0.329 \text{ nm}^{-1}$  
   c. $0.26 \text{ nm}^{-1}; 0.45 \text{ nm}^{-1}; 0.53 \text{ nm}^{-1}; 0.91 \text{ nm}^{-1}$

**CHAPTER 7**

7.1 Defining an element of surface area, $2\pi r^2 \sin \theta d\theta$ over which scattering intensity is constant, we have

$$\int_0^\pi 2\pi r^2 \sin \theta d\theta = I - I_0$$

Substitution of $I$ from Eq. 7.15, followed by integration gives

$$I - I_0 = (16\pi^2) R_s I_0$$

In most cases, $I - I_0 \ll I_0$, so we can use the approximation

$$-\tau = \ln \left( \frac{I}{I_0} \right) \approx \frac{I - I_0}{I_0} \approx -\frac{I - I_0}{I_0}$$

Therefore $\tau = (16\pi^2) R_s$.

7.3 Calculating the weight-average molecular weight of the mixture, and taking

$$\% \text{ error} = \frac{M_w - M_p}{M_p} \times 100$$

where $M_p$ = protein molecular weight, we obtain about 50% for the error.

7.5 a. With $\rho(r) = \rho_0$, integration by parts immediately gives the result

$$F = \frac{4\pi n_0 \hbar^2}{h^3} \left[ \sin (hR) - (hR) \cos (hR) \right]$$