9.03 Introduction to Macromolecular X-Ray Crystallography

Esko Oksanen and Adrian Goldman, University of Helsinki, Helsinki, Finland

9.03.1 Introduction

Macromolecular crystallography is a powerful method for investigating the atomic structure of proteins and nucleic acids and thereby unravelling the molecular mechanisms of their functions. To visualize atoms, light must have a wavelength about the size of an atom, so that the atoms can diffract light, identical with cells and organelles that scatter visible light. For this purpose X-rays are required. However, since the interaction of X-rays with matter is weak, scattering is also weak unless there are multiple molecules that are ordered in the same way, which means we require crystals.

We therefore describe the basis of macromolecular crystallography and provide a summary of how to understand the results of a crystallographic experiment. We start with a mathematical description of what a crystal means in terms of symmetry; this applies to all crystals, whether macromolecular or not. Later, we describe how protein crystals grow by using the hanging drop and sitting drop vapor diffusion methods; this explains why protein crystals are so fragile and scatter X-rays very weakly.

The phenomenon of diffraction and its description as a Fourier transform (FT) is explained. The measured intensity of the diffracted X-rays related to the FT of the electron density, and the electron density – seen as an electron density map – is related to the (inverse) Fourier sum of the intensity of the diffracted X-rays. As we can only measure their intensity, we do not know the phases of the diffracted X-rays; we have to determine them to solve the structure. Therefore, three principal methods are used, two experimental approaches (isomorphous replacement and anomalous scattering) and one based on known structures (molecular replacement).

Next, we describe how, the 'electron density map' is improved and interpreted in terms of the atoms of a structure. The inconsistency between this structure and the experimental data is later minimized to make sure it is as accurate as possible. Finally, we provide an example of a crystal structure from recent literature and summarize the various statistics reported in papers on crystal structure.

9.03.2 Why Crystallography?

The standard approach to study minute details is microscopy, in which light scattered by the specimen is focussed onto the image plane by a lens. The smallest observable detail, however, is limited to half of the wavelength of light (~300–700 nm). The inter-atomic distances in organic molecules are ~0.1–0.2 nm; therefore, we cannot observe atoms under a light microscope but require light of a shorter wavelength. Photons of this wavelength are X-rays, and so an 'X-ray microscope', if it existed, would allow us to visualize atoms. Unfortunately, the refractive index of X-rays is so small that the lenses required to focus X-rays in an X-ray microscope are impossible to make. The scattered radiation, however, still contains the information about the structure of a molecule. It is not practical to image a single molecule because >99% of the X-rays pass straight through, hence to obtain any appreciable signal a macroscopic sample is required.

This imposes a further condition: the molecules must line up with each other in a well-defined spatial arrangement; that is, they must form a crystal. From such a crystal, the structure of a macromolecule may be determined by using single-crystal X-ray diffraction. The method is essentially the same for all biological macromolecules or complexes. However, as most of the structures determined are of proteins, we often refer this method as protein X-ray crystallography.

What topics can be addressed by X-ray crystallography? In recent years, the structures of important biological systems have been solved, for example, β_2-adrenergic receptor, which upon binding to adrenaline or nor-adrenaline causes the 'fight-or-flight' response[1] and the structure of the ribosome, which translates the messenger RNA in the cell into protein.[2] The ribosome is also an important drug target; the macrolide antibiotics like erythromycin bind to the 50S ribosomal subunit, and the structure of the ribosome explains how they work, and how mutations in the bacterial ribosome lead to antibiotic resistance.

The understanding of the degradation of natural products such as camphor has been greatly enhanced by understanding the catalytic cycle of the cytochrome P-450 enzyme P-450cam in structural detail.[3,4] These enzymes catalyze the addition of O_2 to nonactivated hydrocarbons at room temperatures and pressures – a reaction that requires high temperature to proceed in the absence of a catalyst. *O*-Methyltransferases are central to the secondary metabolic pathway of phenylpropanoid biosynthesis. The structural basis of the diverse substrate specificities of such enzymes has been studied by solving the crystal structures of chalcone *O*-methyltransferase and isoflavone *O*-methyltransferase complexed with the reaction products.[5] Structures of these and other enzymes are obviously important for the development of biomimetic and thus environmentally more friendly approaches to natural product synthesis.

9.03.3 Protein Crystals

What is a crystal? We need a mathematical description of a crystal in order to understand, even in a qualitative way, how crystals diffract X-rays and why we see the patterns we do?

Let us start from a crystal of salt, sugar, diamond – or even the enzyme inorganic pyrophosphatase (**Figure 1**). Why do they have sharp edges and regular faces? The reason is that crystals are macroscopic objects in which the constituent atoms or molecules arrange themselves in the same manner. This means that an ideal crystal consists of a series of repeated units (the unit cell, see below), with each unit containing the same arrangement of atoms inside it – known as the motif (**Figure 2**). Exact motions – translations – relate the atoms to each other; clearly they form an array – a lattice of indistinguishable points – and we can move from one point (O in **Figure 2**) to an indistinguishable point (P in **Figure 2**) along a straight line. Such a vector between two equivalent points is known as a lattice vector. Three noncoplanar lattice vectors, called the basis vectors[1] **a**, **b**, and **c** define a coordinate system. Any point in the crystal can then be referred to by a vector $(x\mathbf{a} + y\mathbf{b} + z\mathbf{c})$, where x, y, z are the coefficients in the **a**, **b**, and **c** directions – just as in a three-dimensional Cartesian coordinate system. Unlike a normal Cartesian system, the basis vectors are neither necessarily perpendicular to each other nor of the same length, and so they do not enclose a cuboid, but a general prismoid shape.

[1] Vectors are marked with bold type so **a** is a vector and a is a scalar.

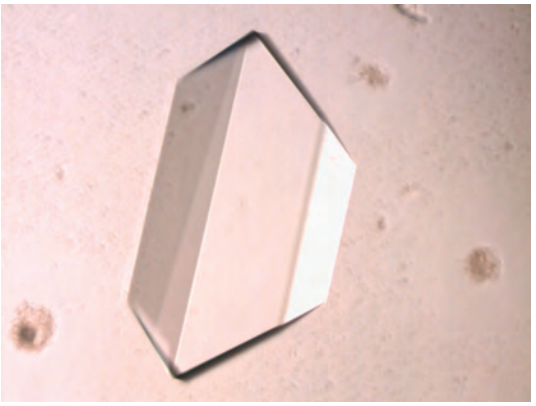

Figure 1 A crystal of yeast inorganic pyrophosphatase grown by temperature-controlled batch crystallization.[23] The longest dimension of the crystal is ~700 μm.

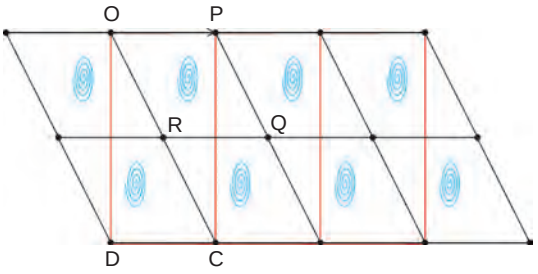

Figure 2 A schematic representation of a protein crystal in which the motif (protein molecule) is shown as a spiral in cyan. The smallest unit cell (OPQR) is shown in black; it is a rhombus: lozenge-shaped with all sides equal. A larger rectangular unit cell (OPCD) is shown in red. This cell leaves one lattice point in the middle and is known as *centered*.

This prismoid box is called the unit cell, and the entire crystal can be constructed by translating the box parallel to its edges. It is possible to choose many different unit cells, as shown in **Figure 2**, but we normally choose the smallest unit cell. However, in some cases a more convenient, but larger, cell is chosen as seen in **Figure 2**, where the rectangular centered cell (red) would be chosen. Such choices make cells easier to understand. Clearly, a rectangle picked up and rotated 180° looks the same; this is equivalent to the statement that a rectangular unit cell has twofold symmetry: that is, 90° angles and unequal edges. The same is true of the lozenge-shape in **Figure 2** but it is not as obvious.

Up to this point, our description has been general; it applies to crystals of rhenium chloride as accurately as to crystals of the ribosome. In addition to the translational symmetry (periodicity) that is inherent in the definition of a crystal, other symmetry can occur, but the kinds that can occur are restricted to crystals of biological macromolecules. Because the molecules are chiral, the symmetry operations in crystals must not change the handedness of the molecule, and so mirror planes, inversions, and 'glide planes' (sliding mirror planes) do not occur. This leaves only rotations and 'screws' (helical-type symmetry, sliding rotations).

Rotations of 60°, 90°, 120°, or 180° are the only ones allowed,[2] corresponding to six-, four-, three-, and twofold rotations. In addition, screw axes can occur, where the molecule is rotated by the same angles – 60°, 90°, 120°, 180° – and translated by a fraction of one of the lattice vectors **a**, **b**, or **c**. These have symbols like 2_1 (a 180° rotation followed by a translation of 1/2 of a lattice vector) or 4_3 (a 90° rotation followed by a translation of 3/4 of a lattice vector). These symmetry operators, lined up through the entire crystal, are the crystallographic symmetry operators.

[2] Other rotations are not compatible with the translational symmetry, which is the essence of a crystal. For a detailed explanation, see, for example, Giacovazzo *et al.*[6]

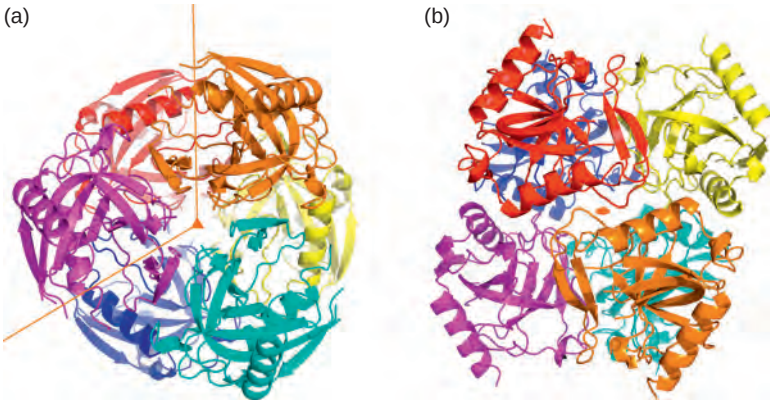

Figure 3 The *Escherichia coli* pyrophosphatase hexamer in a standard protein representation: the spirals are α-helices and the arrows are β-strands. The unit cell axes are marked with orange lines and each monomer has a different color (a) viewed along one of the twofold axes. The twofold rotation axis, marked with an ellipse, relates the orange and red monomers, and the yellow and purple monomers (and the hard-to-see cyan and blue monomers) to each other. (b) A view along the threefold axis, marked with a triangle. The threefold axis relates the orange, purple, and cyan monomers to each other, as well as the yellow, red, and blue ones to each other.

The presence of rotational or screw symmetry means that the unit cell has internal symmetry. Therefore, only part of the unit cell, known as the asymmetric unit, is needed to uniquely define the unit cell. (The asymmetric unit may also contain more than one molecule, related by movements – symmetry operations – that are not part of the crystal symmetry – noncrystallographic symmetry operators. This can be very important in determining the protein structure, as discussed in Section 9.03.9.3).

The convention we have described implies a hierarchy. We can generate the unit cell from the asymmetric unit by applying the various additional crystallographic symmetries: rotations, screws; and we can generate the entire crystal by translating the unit cell parallel to its edges – by its lattice vectors. Indeed, multimeric proteins sometimes crystallize so that the asymmetric unit contains only one monomer and the other monomers in the biological multimer are related by crystallographic (rotational) symmetry operators. One such example is hexameric *Escherichia coli* pyrophosphatase,[7] which has D_3 (32) point group symmetry[3] and crystallizes in space group R32, where each lattice point also has D_3 symmetry. The point group describes the rotational symmetries of an object – be it a molecule or something else. For *E. coli* pyrophosphatase, the point group 32 means there is a threefold (120°) rotation perpendicular to a twofold (180°) rotation. In the crystal, the center of the hexamer and a lattice point with the same symmetry coincide and therefore the asymmetric unit of the crystal contains a monomer. Crystal symmetry then relates the monomers in the hexamer (**Figure 3**).

The combination of rotational and translational symmetry defines the space group of the crystal. It is shown that 235 space groups exist, but only 65 allow the handedness of the molecule to be preserved, and so only 65 can occur in macromolecular crystallography. The space groups are numbered, but are commonly referred to by their symbols, such as $P2_12_12_1$. The most common in macromolecular crystallography are $P2_12_12_1$, P1, $P2_1$, and C2.

9.03.4 Obtaining Protein Crystals

A crystal is a very precisely ordered aggregate that represents the thermodynamically most favorable state under the conditions of crystallization. Since attractive forces between protein molecules are not very specific, successful crystallization requires both a very pure protein sample (typically $\approx$ 99% pure) and a careful search for the right conditions. In addition to favorable thermodynamics, crystal growth must also be kinetically favored over nonspecific aggregation. This often means a relatively slow process and while some protein crystals grow in hours, most take weeks to months to form.

[3] The point group, which is D_3 in the Schönflies notation used for example in molecular spectroscopy, is called 32 in the International (or Hermann–Mauguin) notation used by crystallographers.