I690/B680 Structural Bioinformatics Spring 2006

Protein Structure Determination by X-ray Crystallography

Suggested Reading

(1) "Crystallography Made Crystal Clear" by Gale Rhodes, Academic Press, 1993.
(2) "Proteins: Structures and Molecular Properties" by Tom Creighton, Freeman, 1992, Section 6.1, pages 202-216
(3) Van Holde, Johnson, Ho. "Principles of Physical Biochemistry", 2nd Ed., Prentice Hall, 2006, Chapter 6

There are two experimental techniques commonly used to determine the 3D structures of biopolymers at close to atomic resolutions. They are:

(1) X-ray Crystallography [applies to crystalline samples]
(2) Nuclear Magnetic Resonance (NMR) Spectroscopy [applies to samples in solution]

We will discuss X-ray crystallography in this leture and NMR spectroscopy in the next lecture.

1. Overview

- When X-rays are shone through a crystal, the atoms in the crystal diffract the X-rays. The diffracted X-rays can be detected and the observed pattern can be interpreted to provide information about the arrangement of the molecules in the crystal and the arrangement of atoms within the molecule.

- The steps in determining the structure of a protein by X-ray crystallography are:
  - Express/purify the protein
  - Crystallize the protein
  - Collect diffraction data
  - Determine the space group, unit cell dimensions, and number/symmetry of molecules per unit cell
  - Solve the "phase problem"
  - Calculate an electron density map
  - Build a model of the molecule to fit the electron density map
  - Refine the model

- We will focus our discussion on the phenomenon of X-ray diffraction and the relationship of the observed data to the organization of the crystal and the structure of the molecule.

2. Sample Preparation and Crystallization

- Crystallization of proteins involves screening a wide range of solution conditions to find those conducive to crystal formation. This necessitates a large amount (at least several milligrams) of highly pure and homogeneous material. *E. coli*

expression systems are commonly used because heterogeneous post-translational modifications are not generally a problem with *E. coli*.

- Typical parameters that are varied during crystallization trials are:
  - Protein concentration
  - Buffers and ions
  - Ionic strength
  - pH
  - Temperature
  - Precipitants

3. Description of a Crystal

- Crystals are ordered arrays of molecules repeating many times in 3 dimensions
  - The repeating unit in the crystal is called the **unit cell**.

- Creighton Figure 6.2, page 204 shows examples of unit cells and the other symmetry elements discussed below

- In some cases there is also **internal symmetry** within the unit cell (e.g. a 2-fold rotation axis).
  - Symmetry operations that also map one unit cell onto another are referred to as **crystallographic symmetry** operations
  - Other symmetry operations are referred to as **non-crystallographic symmetry** operations
  - The smallest unit that can be repeated an integral number of times to yield the unit cell is called the **asymmetric unit**.

- There are 65 possible ways in which the asymmetric unit can be repeated many times to form a crystal. These different arrangements of the crystal are called **space groups**.
  - Not all space groups are possible with proteins (because some involve mirror symmetry, which is not compatible with proteins having all L-amino acids)

- Proteins in crystals are typically highly **hydrated**, i.e. a large proportion of the crystal is water, so the crystals can be quite soft.

- There is convincing evidence that, in most cases, the structure of a protein in a crystal is very similar to its (average) structure in solution, but crystallized proteins are less flexible than proteins in solution.

4. Diffraction from Crystals

- The typical experimental setup is shown in Rhodes, Figure 2.5, page 11.

- We will discuss why X-rays are diffracted from a crystal to give **reflections** in specific positions on the X-ray film/detector.

- **Bragg's Law** describes the condition for constructive interference of light waves reflected from parallel surfaces.
  [See Van Holde Figure 6.10 (page 261, 1st ed. Or page 293, 2nd ed.) and Rhodes, Figure 4.7 (page 49)]

  $$\mathbf{2d\sin\theta = n\lambda}$$

  d is the distance between the two surfaces
  θ is the angle at which the light beam impinges upon the surfaces
  λ is the wavelength of the light
  n is an integer (1,2,3…)

- Crystals can be thought of as parallel planes of atoms. We will look at some figures demonstrating this.

- Thus, if an X-ray impinges on a plane of atoms in the crystal at an appropriate angle, it will be diffracted, interfere constructively with parallel beams, and result in a detectable diffracted X-ray (a reflection).

- Where will this diffracted X-ray appear on the X-ray detector/film? We will look at a diagram in which the position of the reflection is mapped onto the surface of a sphere whose radius is $1/\lambda$ (the inverse of the X-ray wavelength). We will conclude that:

  - **The spacing of reflections is inversely related to the plane-spacing in the crystal**.
  - **The dimensions of the crystal lattice (unit cell size, shape and symmetry) dictate the positions of the observed reflections**

- Imagine a crystal with several possible planes. Imagine moving an X-ray around that crystal. At certain angles one gets constructive interference and a reflection will be observed.
- Now think about it differently. Shine an X-ray at a particular angle to the crystal. The X-ray will hit the different planes at different angles and some will diffract constructively, giving reflections at certain angles from the source beam.
- Now rotate the crystal around an axis perpendicular to the source beam. Again, some planes will diffract constructively, giving reflections at certain angles from the source beam, but these reflections will be different from those obtained before you rotated the crystal.

- In practice one collects a series of diffraction pictures with the crystal positioned at different angles to the beam. One can then analyze the positions of reflections in the resulting diffraction pattern to deduce the plane spacing in the crystal, i.e. the space group and unit cell geometry.

5. Reflection Intensities and Phases

- The above discussion concludes that the positions of the reflections are dependent on the unit cell dimensions and space group but they do not provide any information about the structures of the molecules that are located within the unit cells. For this it is necessary to analyze the intensities and phases of the X-ray reflections.

- In discussing the condition for constructive interference (Bragg's Law) we have considered groups of parallel planes spaced a distance d apart, where d is related to the dimensions of the unit cell. We will now consider a crystal in which each unit cell contains one molecule with 3 atoms, we will assume that the interplane distance satisfies Bragg's Law.
  - One can draw a separate set of parallel planes through EACH of the 3 atoms. The three sets of parallel planes are also parallel to each other.
  - X-rays reflected from a plane through one type of atom will be **in-phase** with X-rays reflected from a parallel plane through the **same type of atom**. Thus, there will be constructive interference between these reflected rays and the possibility of detecting a reflection.
  - But X-rays reflected from a plane through one type of atom will NOT be in phase with X-rays reflected from a parallel plane through a different type of atom (e.g. the blue and the red atoms in the figure).
  - In fact, **the reflected X-ray at the Bragg angle (θ) will be a convolution of the reflected X-rays coming from each set of parallel planes, i.e. each type of atom**. In some (but not all) cases, the interference will result in a detectable reflected X-ray (a reflection on the X-ray detector).
  - We will look at Excel simulations of three X-rays and consider the properties of these X-rays and their resultant. Note the following features:
    1. All three waves have the **same frequency** – the frequency of the X-ray source
    2. The three waves have **different intensities and phases**
    3. The **intensities** depend upon how efficiently the atoms reflect X-rays, i.e. the **electron density** on the plane from which the X-ray is reflected.
    4. **The relative phases of the three waves depend upon the relative positions of the three different sets of planes, i.e. the positions of the atoms in the molecule!**
    5. The resultant of these three waves is another wave with the same frequency as the component waves and with intensity and phase that depend on those of the component waves.

6. Detecting Intensities and Phases: The Phase Problem

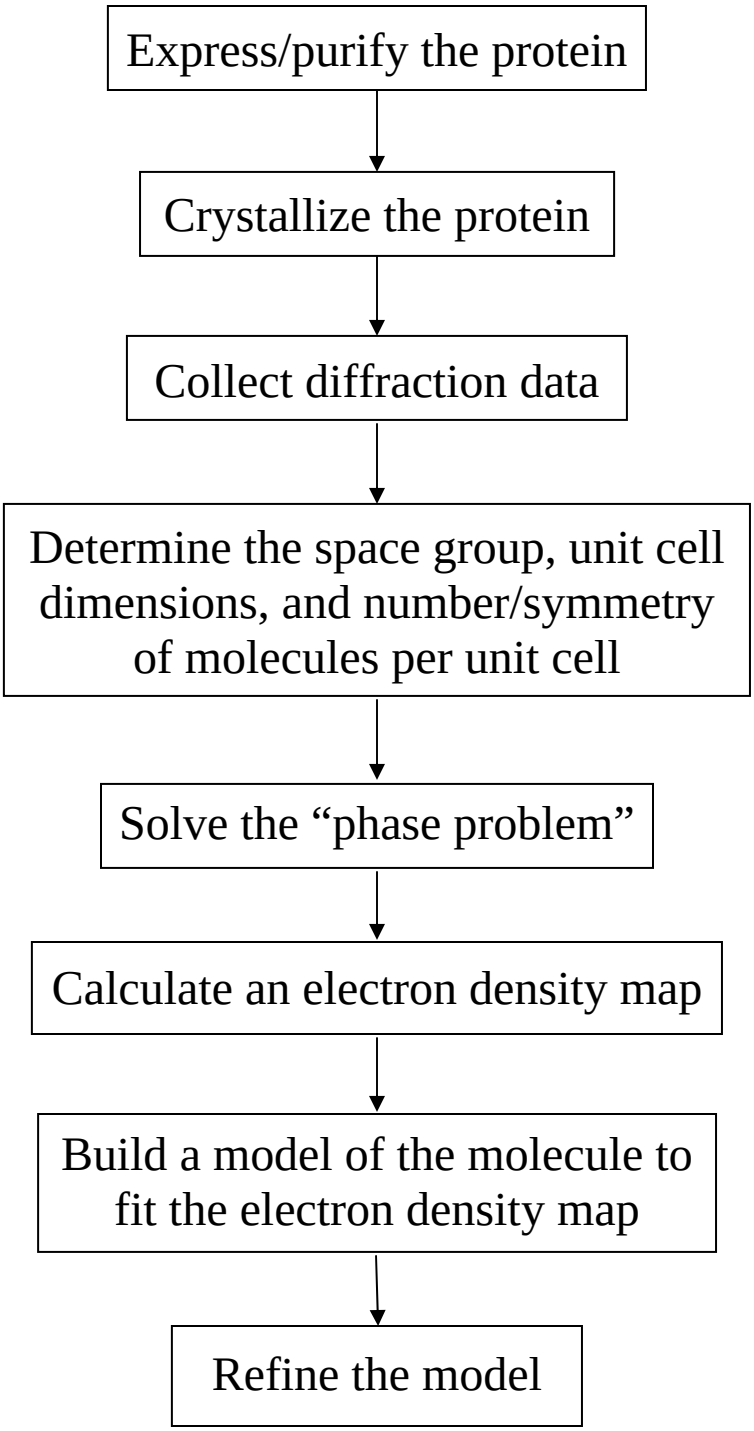
- The above analysis demonstrates that the X-ray we detect is a resultant of X-rays emanating from each atom-type in the molecule.
- **If we could measure the intensity and the phase of the resultant X-ray, then:**
  - **we could deconvolute it into its component waves using the mathematical process called Fourier transformation or Fourier analysis; and**
  - **we could analyze each component wave to deduce the electron density at each position in the molecule (or the unit cell)**
- **We can determine the intensity of the resultant X-ray from the intensity of the reflection on the X-ray detector.**
- **However, the reflection does not contain any information about the phase of the resultant. This missing information is referred to as the phase problem!**

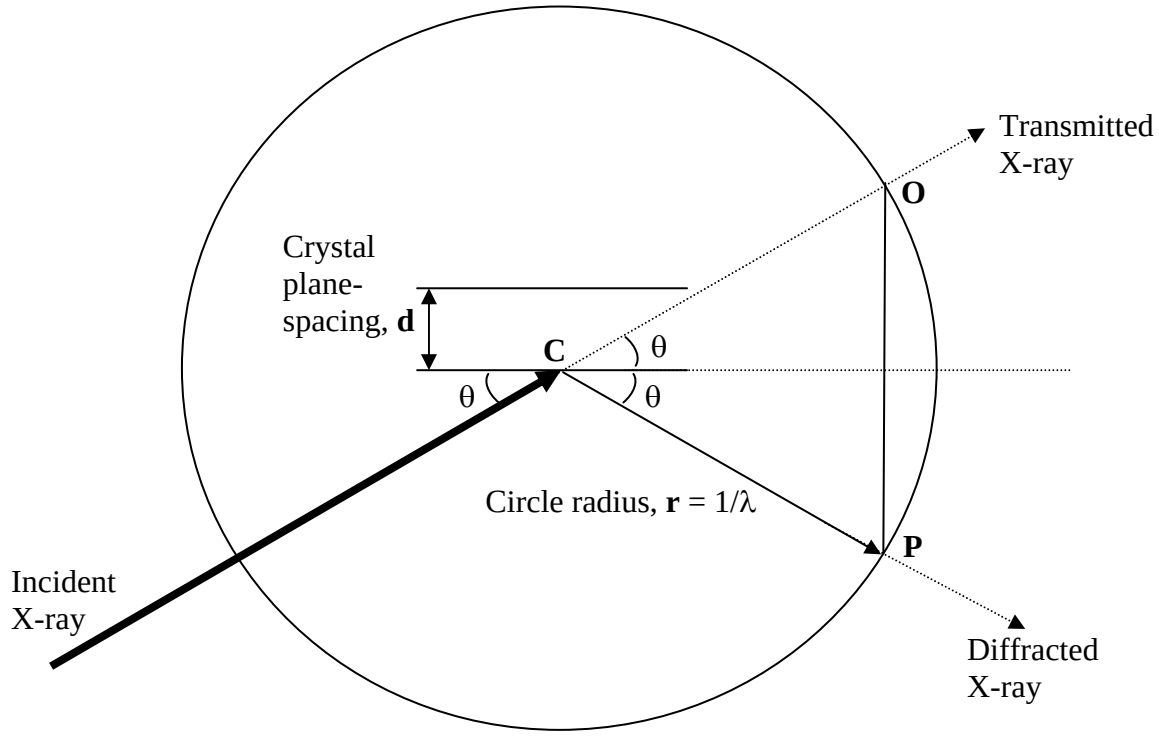
- Crystallographers have developed several methods for **solving the phase problem**, i.e. determining the phases of the reflected X-rays. We will not discuss them in detail, but you should at least recognize that the following approaches are used.

  - Molecular Replacement. If the structure of a similar protein is known, this can be used to build a model of your protein then comparison of this model with the observed diffraction intensities can allow the phases to be solved.
  - Multiple Isomorphous Replacement. A series of heavy atom derivatives of the crystals are prepared and diffraction data collected n the derivatives. As long as the molecular structure and unit cell dimensions remain essentially unchanged, the diffraction patterns with and without heavy atoms can be compared to extract phases for the various reflections.
  - Multi-wavelength Anomalous Dispersion (MAD). Typically samples are prepared in which methionine residues are replaced by selenomethionine. Diffraction data are collected at several wavelengths (at a synchrotron) and the scattering of X-rays by the selenium atom can then be used to obtain phases.

7. Fourier Analysis and Solving the Structure

- Once phases have been determined, it is a straightforward matter to convert the intensities and phases of the reflected X-rays to the electron densities at each position in the unit cell. This is accomplished using Fourier transformation.
- The resulting electron density map and the known protein sequence are used to build (then iteratively refine) molecular models that are in best agreement with the experimental data.


Express/purify the protein
Crystallize the protein
Collect diffraction data
Determine the space group, unit cell dimensions, and number/symmetry of molecules per unit cell
Solve the “phase problem”
Calculate an electron density map
Build a model of the molecule to fit the electron density map
Refine the model



Distance (**OP**) from origin **O**, to reflected point **P** is related to the angle θ and the radius of the circle shown (**r**) by the equation:

$\sin\theta = (\mathbf{OP}/2)/\mathbf{r} = \mathbf{OP}/2\mathbf{r} \Rightarrow \mathbf{OP} = 2\mathbf{r}\sin\theta$

If the radius of the circle is set at $1/\lambda$, then $\mathbf{OP} = (2\sin\theta)/\lambda$.

But from Bragg's Law, diffraction will only occur if $2\mathbf{d}\sin\theta = n\lambda \Rightarrow (2\sin\theta)/\lambda = n/\mathbf{d}$.

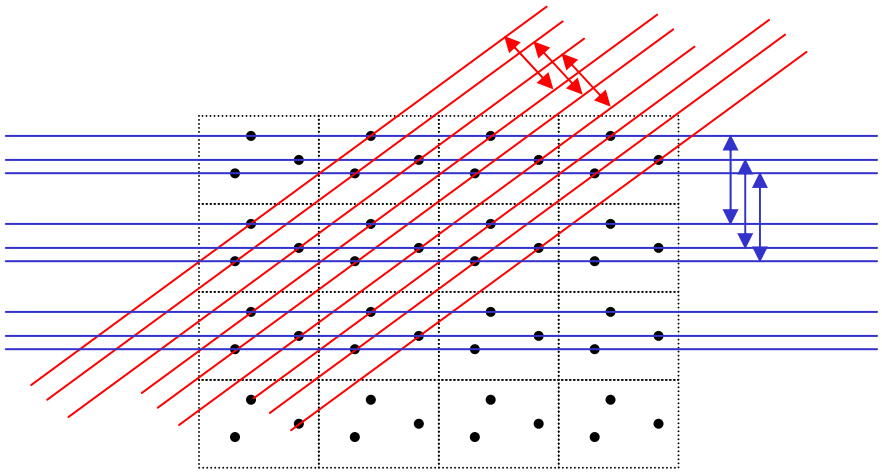
Therefore, reflections will occur only at points that follow the equation:

$$\mathbf{OP} = n/\mathbf{d}.$$

The spacing of reflections is inversely related to the plane-spacing in the crystal.

i.e. the dimensions of the crystal lattice (unit cell size, shape and symmetry) dictate the positions of the observed reflections.

d=1.1λ
Diffraction
condition (n=1)
θ = 27.0°



d=1.5λ
Diffraction
condition (n=1)
θ = 19.5°