X-ray crystallography

CS/CME/BIOPHYS/BMI 279
Fall 2015
Ron Dror

X-ray crystallography is the most common method for determining three-dimensional structures of proteins and other molecules. It involves making a crystal of the molecule to be imaged (with many copies of that molecule packed in a regular three-dimensional grid or "lattice"). The crystal is then exposed to a very bright x-ray beam. Some of the x-rays shine right through the crystal, but some are scattered (deflected) by electrons in the crystal. These scattered x-rays form a diffraction pattern, which is basically a set of bright spots in a regular arrangement (see lecture slides for examples). The diffraction pattern is three-dimensional, in the sense that the bright spots are distributed throughout a volume. To image them all with a planar imaging instrument (which captures a 2D image at each point in time), one needs either to rotate the imaging device or to rotate the crystal itself. Once the three-dimensional diffraction pattern has been reconstructed, it can be used to reconstruct the electron density of the crystal (and thus the three-dimensional structure of the crystallized molecule), but this requires solving a challenging computational problem.

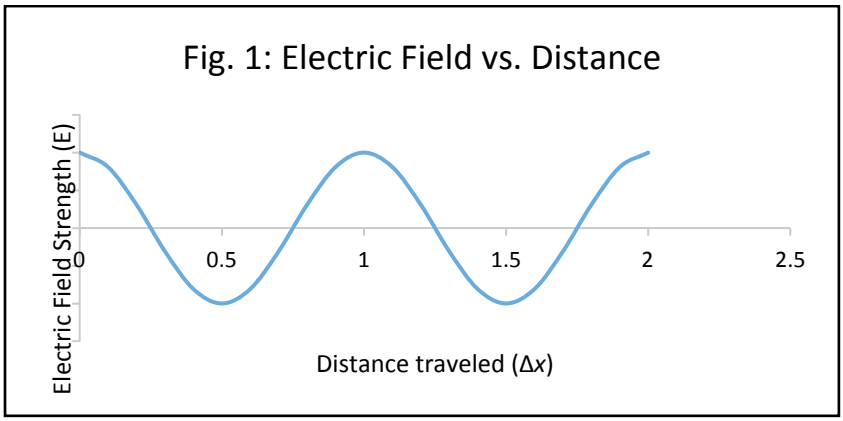
1. X-Ray Crystallography: How it works

X-Rays are Electromagnetic Waves

X-rays (like visible light and other types of radiation) are oscillations of electromagnetic fields propagating in space, known as **electromagnetic waves**. At any instance in time, along the direction at which the x-ray travels, the electric field strength *E* can be described as a function of the distance travelled, Δx:

$$E = E_0 \cos(2\pi\Delta x/\lambda)$$

where E_0 is the **amplitude** and λ is the **wave length**. The x-rays used in crystallography generally have wavelengths in the 0.1 – 10 nm range. The argument of the cosine function, namely the quantity $\phi = (2\pi\Delta x/\lambda)$ here, is called the **phase**. (fig. 1)



X-Ray Crystallography Experiment

When an x-ray beam goes through a crystal, a proportion of it is scattered by electrons. The x-rays following different paths through the crystal but arriving at the same point on the imaging device (detector) will interfere with each other and produce patterns on the detector / screen / photograph. X-ray waves that are in phase with one another will add constructively (i.e., add to form an even stronger wave), but those that are sufficiently out of phase with one another will tend to cancel one another out. This interference phenomenon is referred to as **diffraction**, and it causes the pattern of bright spots.

It turns out that the diffraction pattern is closely related to the Fourier transform of the electron density: the brightness of the diffraction pattern at each point is proportional to the magnitude of the Fourier transform coefficient at that point in Fourier space. Then the brightness of each spot in the diffraction pattern specifies the *magnitude* (scaling factor) of one Fourier component (i.e., 3D sinusoid) of the electron density. (Any pattern of electron density can be written as a sum of 3D sinusoids, in a unique way.) Unfortunately, the *phase* (shift) of each Fourier component cannot be measured directly in any straightforward manner. If we knew both magnitudes and the phase for each Fourier component, then we could determine the electron density by just summing up the shifted and scaled sinusoids, but because we don't know the phases, figuring out the electron density is challenging.

You are not responsible for understanding *why* the observed diffraction pattern ends up being the Fourier transform of the electron density (or more specifically, the magnitude of each Fourier coefficient), but for those who are curious, here's a brief explanation. (This explanation is slightly different—and more mathematical—than the one given in class, but the two are consistent. Both cut some corners for brevity.)

Consider x-ray paths that originate from the same point source and arrive at the same point on the detector,

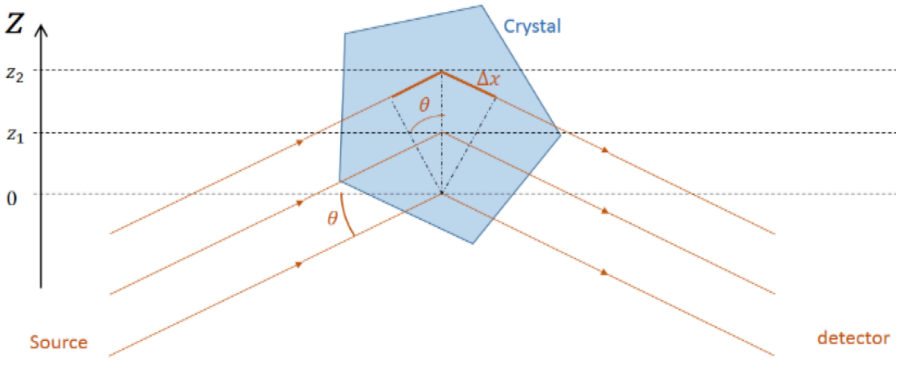


Fig. 2: Bragg diffraction

but in the process were scattered by electrons at different locations inside the crystal. If both the source and the detector are far away in comparison to the size of the crystal, these paths are approximately parallel to each other (Fig. 2). Define the bisector of the incident and detecting directions as the z axis (Fig. 2). We'll assume for now that all electrons in the crystal lie along the z-axis (i.e., we're starting off with a one-dimensional crystal, for simplicity). From the source to the detector, an x-ray scattered at $z = z_n$ travels a

distance $\Delta x_n = 2z_n \cdot \sin\theta$ longer than one scattered at $z = 0$. Therefore a phase difference will be introduced:

$$\Delta\phi_n = \frac{2\pi\Delta x_n}{\lambda} = 2\pi \cdot z_n \cdot \frac{2\sin\theta}{\lambda}$$

Since the electron density (ρ) is different everywhere inside the crystal, more x-rays are scattered at some locations than others (the higher the density of electrons at a location—i.e., the higher the likelihood of finding an electron at that location—the higher the likelihood an x-ray passing through that location will be scattered). Suppose the electric field oscillation of the scattered x-ray at position *z* is proportional to $\rho(z)$. Then at the detector, we can determine the electric field strength (at a particular instance in time) by summing over all incident light paths. The sum will take the following form (where z_0 = 0):

$$E \propto \rho(z_0)\cos\left(2\pi \cdot z_0 \cdot \frac{2\sin\theta}{\lambda}\right) + \rho(z_1)\cos\left(2\pi \cdot z_1 \cdot \frac{2\sin\theta}{\lambda}\right) + \rho(z_2)\cos\left(2\pi \cdot z_2 \cdot \frac{2\sin\theta}{\lambda}\right) + \cdots$$

You may notice that this expression, as a function of $\left(\frac{2\sin\theta}{\lambda}\right)$, is similar to a one-dimensional Fourier transform of the electron density of the crystal $\rho(z)$. This generalizes to three-dimensional crystals: by measuring the diffraction intensity at different θ and different orientations of the z axis, a three-dimensional Fourier transform of the crystal structure can be obtained.

2. Solving the Phase Problem

To determine the electron density of the crystal (and thus the positions of the atoms in the molecule), we need to determine phases for each of the Fourier components (sinusoids). This computational problem is referred to as **phasing** and generally consists of two steps:

(1) First, come up with an approximate solution for the structure and thus an approximate set of phases. This is called **initial phasing**.
(2) Then adjust the phases to achieve an estimated structure that better agrees with the observed diffraction patterns. This is called **phase refinement**, and it's usually performed in an iterative manner.

Initial Phasing

The most common way of determining initial phases is called **molecular replacement**. Here, one starts with a rough guess of the structure—usually based on the known structure of a homologous protein—and uses it to determine approximate phases. To do this, one must search over all the possible ways in which the protein might be positioned in the unit cell (for example, different rotations of the protein relative to the axes of the crystal lattice) and find the one that best fits the observed diffraction pattern.

In some cases—particularly when one doesn't have a good guess of the approximate structure of the protein—one needs some additional experimental to solve the phasing problem. Many techniques have been devised to get such information, with different strengths and weaknesses. As an example, one might replace a few of the atoms of the protein with heavier atoms. The difference between the diffraction

patterns observed before and after replacement provides information about phases. (You're not responsible for knowing about this technique.)

Phase Refinement

During the phase refinement process, one adjusts the structure to better fit the observed experimental data. The set of potential protein conformations is far too large to search exhaustively. Thus, one generally employs a Monte Carlo search method (in particular, simulated annealing). Also, one generally wishes to focus on protein conformations that seem likely given what is known about the physics of proteins. Thus, one generally attempts to find a structure that minimizes a sum of two functions. The first function increases with the difference between the experimentally observed diffraction pattern and the one predicted from the chosen structure. The second function is a molecular mechanics force field, similar to those used in molecular dynamics; it is larger for high-energy conformations than for low-energy conformations.