Lecture #13: 11 May 2004
Topics: Protein Structure Determination
Scribe: Minli Zhu

We acknowledge that protein structure determination is lagging behind protein sequencing. In this lecture, two kinds of methods are introduced to determine the structure of protein. They are X-ray Diffraction Crystallography and NMR Spectroscopy. For each method, there are several techniques involved.

1 X-ray Diffraction

1.1 X-ray Diffraction Crystallography

X-ray crystallography utilizes the fact that X-rays are diffracted by crystals. X-ray can be diffracted by the electron cloud of an atom because it has proper wavelength: roughly 10^{-8} cm. X-ray scatter off the periodic group of molecules or atoms in the crystal to form a diffraction pattern that is strong enough to be sensed. Then, the electron density in a crystal lattice cell can be reconstructed based the pattern. Extra phase information must be obtained either from similar solved structures or from supplementary diffraction experiments to accomplish the reconstruction. Eventually, a model can be built from these information.

1.2 Diffraction Image

1.2.1 Protein Crystallization

We want the crystal grower to produce fairly large protein crystals, so the molecules have sufficient long range order that an X-ray beam will be diffracted into a pattern of reflections. This is based on the fact that protein crystal still maintain the same structure and functionality, so we can analyze it this way.

1.2.2 X-ray Diffraction Setup

We setup the device as following in order to get the image.

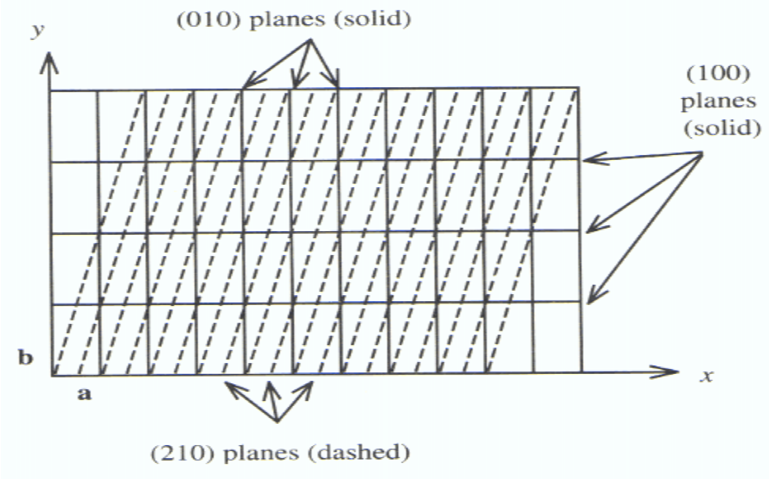
1. source
2. crystal
3. receiver screen

1.2.3 Vapor Diffusion

Start off by putting a drop containing protein in stabilizing buffers, which is connected to a reservoir. The reservoir usually contains the same chemicals minus the protein with overall higher concentration so that water will evaporate from the drop. When the conditions are right, the system will produce a gradual increase the concentration in protein, so that a few crystals may form.

1.2.4 Crystal Diffraction

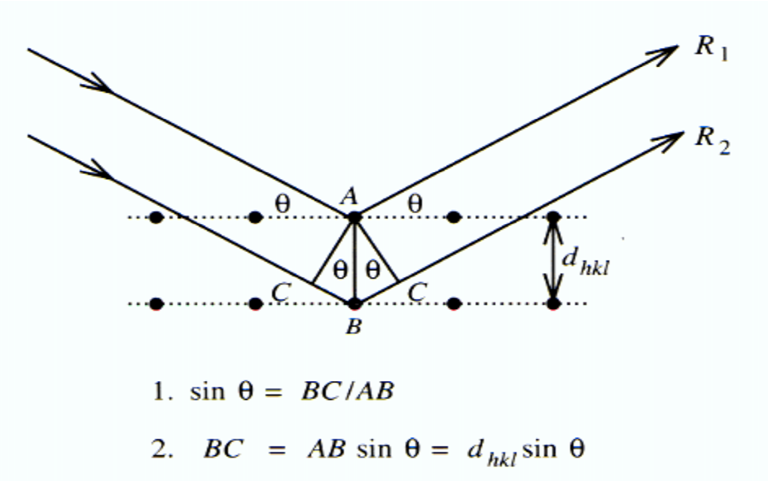
A crystal contains of repetition of a unit cell. Normally diffraction can be thought of either caused by molecule atoms or by the electron cloud surrounding the atoms in the unit cell. Through certain grouping of repetitive lattice points, we can form many sets of parallel planes. We use Lattice Indices to represent how many planes are there in each direction. (hkl) represents the number of planes crossing the unit lattice cell along the (a, b, c) directions respectively.



1.2.5 Bragg's Law

$$2d_{hkl}sin\Theta = n\lambda \tag{1}$$

This essentially says that the difference in distance by two rays reflected by two parallel planes is equivalent to the multiple of X-ray's wave length. To notice that n is known as the order of diffraction, but we can refer to first order diffraction from planes of spacing d/n.



1.3 The Structure Determination Process

The spacing of reflections on the X-ray is related to reciprocal lattice spacings. These in turn are the inverse of real lattice spacing. Symmetries of the real lattice are also present in the reciprocal lattice, giving rise to equivalent reflections. These symmetries aid the crystallographer in determining the unit of the real lattice. Let's walk through the phases. First, starting off by collecting X-ray diffraction data. Second, turn it into electron density maps, which involves some Fourier analysis over the complex domain. Third, the crystallographer needs to fit the chemical structure of the molecule into the density map to get an idea what the structure is like.

1.3.1 Fourier Inversion

The electron density ρ can be recovered, if we know the structure factors F_{hkl}, which are in part obtained by experiments. To be precise, the absolute value of F_{hkl} is the square root of the measured reflection intensity.

$$\rho(x,y,z) = \frac{1}{V}\sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)} \tag{2}$$

However, in the real situation, we can only record the optical intensity, which is the square of the amplitude of sine wave and have lost the phase information. So,we need to guess the phase information to regenerate the crystal structure. The technique of isomorphous replacement with heavy atoms is used to improve the precision of the structure determination.

1.3.2 Fitting the Chemical Structure

From the density of electro clouds to real chemical structure, the electro clouds density resolution should be high enough. A lot of automated tools are use for fitting. For

example, Medical Axis Assessment, and thinning the axis.

2 Nuclear Magnetic Resonance Spectroscopy

Nuclear Magnetic Resonance spectroscopy is based on the principle that the energy levels of atomic nuclei are discrete and radiation can boost the nuclei from low energy status to higher status. But once the magnetic excitation disappear, the nuclei will back to the lower energy and at the same time will emit the spectrum. We record the spectrum and from the spectrum, we can determine the proximities of the protein atoms. The technique has low sensitivity and the data obtained tends to be noisy. The spectrum test includes two steps:

1. Prepare system and issue pulse

2. Detect spectrum

2.1 Chemical Shift and Scalar Coupling

The magnetic field and the excitation radiation frequency can be varied. Chemical shift is defined as nuclear shielding or as applied magnetic field. Chemical shift is a function of the nucleus and its environment. It can be measured relative to a reference compound. Transfer of magnetization between different nuclei can cause couplings, which leads to split responses.

2.2 Multidimensional Nuclear Overhauser Effect

Multi-dimensioned NMR is used in the real situation. Many other improved methods based on the NMR are also frequently used. One of them is NOE. Nuclear Overhauser Effect (NOE) Spectroscopy (NOESY) uses dipolar interaction of spins for proton corrector. NOE correlates all protons that are close enough. Thus, it provides information about pairs of atoms that are close in the tertiary structure. It eventually becomes very helpful in structure determination. COSY and TOCSY have the same functionality as well.

2.3 Resonance Assignment Problem

After we get the spectrum information,we need to fit them in the chemical structure. In order to find out which atoms are interacting, we can try isotope labelling or combinatorial search.

2.4 Tool to interpret NMR

Distance Geometry can be used to recovered the position of two nuclei. Given n atoms $a_1, a_2 \dots a_N$ at unknown position $p_1, p_2 \dots p_N$ and interatomic distances d_{ij} so that:

$$d_{ij} = \|p_i - p_j\| \tag{3}$$

Through the following steps, we can recover the positions P_i. We assume the centroid of the atoms is at the origin. Start off by defining a matrix G,which has rank 3. Then let W be its eigenvectors and λ be its eigenvalues.

$$G = [g_{ij}] = [\overline{P_i}.\overline{P_j}] \tag{4}$$

$$g_{ij} = \frac{1}{2}(d_{i_0}^2 + d_{j_0}^2 - d_{ij}^2) \tag{5}$$

Then rewrite g_{ij}

$$g_{ij} = \sum_{k=1}^{3} \overline{P_i}.\overline{P_j} = \sum_{k=1}^{3} w_{ik} w_{jk} \lambda_k$$

The positions then can be recovered from W and λ

$$\overline{P_{ik}} = \sqrt{\lambda_k} w_{ik} \tag{6}$$

2.5 Practical Issues

In reality, the distances are missing and only ranges are known. The problem becomes NP complete if missing distances or small ranges are given. However,probability distribution on distances may be given, we can get intervals on missing distances via the triangle inequality. Then missing distances can be sampled with these bounds. Alternative approaches can be: distance geometry, geometric build-up, multi-dimensional scaling, CNS partial metrization, alternating projections, graph reductions, and global optimization. DGII, DGEOM,XPLOR, and CNS are available packages.

References

Rhodes, Gale Crystallography Made Crystal Clear 1993
Wuthrich, Kurt NMR of Proteins and Nucleic Acids 1986
http://www.stanford.edu/class/cs237.htm
http://www.shu.ac.uk/schools/sci/chem/tutorials/molspec/nmr1.htm
Hermann,T. Gunther,P. and Withrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J. Mol. Biol. 319,209-227