Structure studies of a protein: macromolecular crystallography (3D- structure) and Circular Dichroism (secondary structure)

X-ray crystallography - is a method of determining the arrangement of atoms within a crystal, in which a beam of X-rays strikes a crystal and diffracts into many specific directions. From the angles and intensities of these diffracted beams, can be produced a three-dimensional picture of the density of electrons within the crystal. From this electron density, the mean positions of the atoms in the crystal can be determined, as well as their chemical bonds, their disorder and various other information.

X-ray crystallography is a form of elastic scattering; the outgoing X-rays have the same energy, and thus same wavelength, as the incoming X-rays, only with altered direction. A regular array of scatterers produces a regular array of spherical waves. Although these waves cancel one another out in most directions through destructive interference, they add constructively in a few specific directions, determined by Bragg's law:

$$2d\sin\theta = n\lambda$$

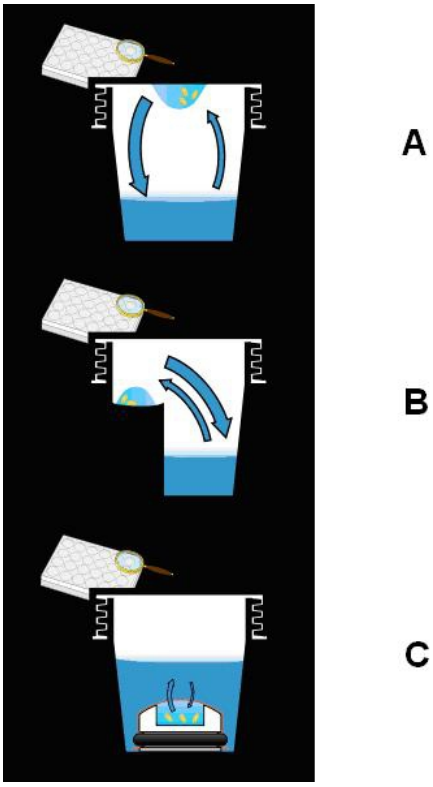
d is the spacing between diffracting planes, θ is the incident angle, *n* is any integer, and λ is the wavelength of the beam

These specific directions appear as spots on the diffraction pattern called *reflections*.

X-rays range in wavelength from 10 to 0.01 nanometers; a typical wavelength used for crystallography is 1 Å (0.1 nm), which is on the scale of covalent chemical bonds and the radius of a single atom. X-ray sources diffractometer (anode from Mo or Cu) and synchrotron.

Crystallization - Small-molecule and macromolecular crystallography differ in the range of possible techniques used to produce diffraction-quality crystals. Small molecules generally have few degrees of conformational freedom, and may be crystallized by a wide range of methods, such as chemical vapor deposition and recrystallization. By contrast, macromolecules generally have many degrees of freedom and their crystallization must be carried out to maintain a stable structure.

Three methods of preparing macrocrystals, A: Hanging drop. B: Sitting drop. C: Microdialysis



Protein crystals are almost always grown in solution. The most common approach is to lower the solubility of its component molecules very gradually; if this is done too quickly, the molecules will precipitate from solution, forming a useless dust or amorphous gel on the bottom of the container. Crystal growth in solution is characterized by two steps: *nucleation* of a microscopic crystallite (possibly having

only 100 molecules), followed by *growth* of that crystallite, ideally to a diffraction-quality crystal. The solution conditions that favor the first step (nucleation) are not always the same conditions that favor the second step (subsequent growth). The crystallographer's goal is to identify solution conditions that favor the development of a single, large crystal, since larger crystals offer improved resolution of the molecule. Consequently, the solution conditions should *disfavor* the first step (nucleation) but *favor* the second (growth), so that only one large crystal forms per droplet. If nucleation is favored too much, a shower of small crystallites will form in the droplet.

It is extremely difficult to predict good conditions for nucleation or growth of well-ordered crystals. In practice, favorable conditions are identified by *screening*; a very large batch of the molecules is prepared, and a wide variety of crystallization solutions are tested. Hundreds, even thousands, of solution conditions are generally tried before finding the successful one. The various conditions can use one or more physical mechanisms to lower the solubility of the molecule; for example, some may change the pH, some contain salts or chemicals that lower the dielectric constant of the solution, and still others contain large polymers such as polyethylene glycol that drive the molecule out of solution by entropic effects. It is also common to try several temperatures for encouraging crystallization, or to gradually lower the temperature so that the solution becomes supersaturated. These methods require large amounts of the target molecule, as they use high concentration of the molecule(s) to be crystallized. Due to the difficulty in obtaining such large quantities (milligrams) of crystallization grade protein, robots have been developed that are capable of accurately dispensing crystallization trial drops that are in the order of 100 nanoliters in volume. This means that 10-fold less protein is used per-experiment when compared to crystallization trials setup by hand (in the order of 1 microliter).

The growing crystals are generally held at a constant temperature and protected from shocks or vibrations that might disturb their crystallization. Impurities in the molecules or in the crystallization solutions are often inimical to crystallization. Conformational flexibility in the molecule also tends to make crystallization less likely, due to entropy. Crystals can be marred by twinning, which can occur when a unit cell can pack equally favorably in multiple orientations; although recent advances in computational methods may allow solving the structure of some twinned crystals.

Measurment and Data collection: In an X-ray diffraction measurement, a crystal is mounted on a goniometer (fig. 1) and gradually rotated while being bombarded with X-rays, producing a diffraction pattern of regularly spaced spots known as *reflections*. The two-dimensional images (fig. 2) taken at different rotations are converted into a three-dimensional model of the density of electrons within the crystal using the mathematical method of Fourier transforms, combined with chemical data known for the sample.

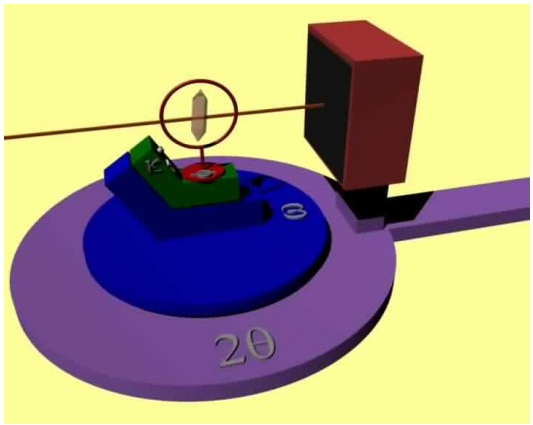


Fig. 1

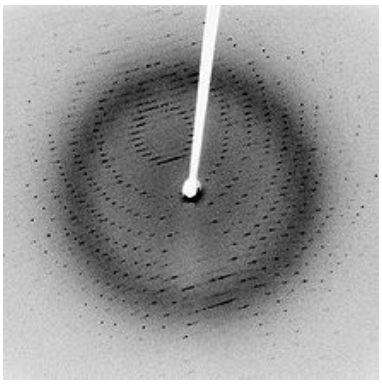


Fig. 2

The intensities of these reflections may be recorded with photographic film, an area detector or with a charge-coupled device (CCD) image sensor. To collect all the necessary information, the crystal must be rotated step-by-step through 180°. However, if the crystal has a higher symmetry, a smaller angular range such as 90° or 45° may be recorded.

Data analysis: Each spot corresponds to a different type of variation in the electron density; the crystallographer must determine *which* variation corresponds to *which* spot (*indexing*), the relative strengths of the spots in different images (*merging and scaling*) and how the variations should be combined to yield the total electron density (*phasing*). A byproduct of indexing is to determine the symmetry of the crystal, i.e., its *space group* (fig. 3).

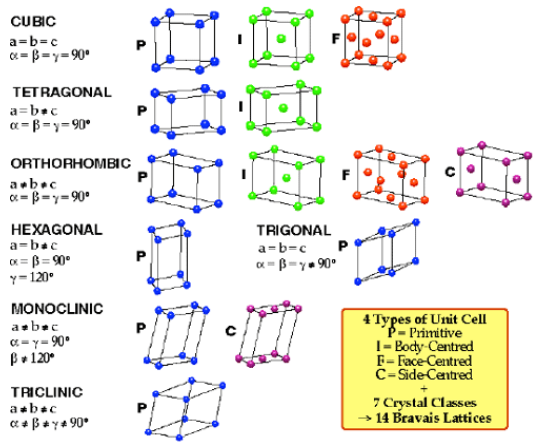
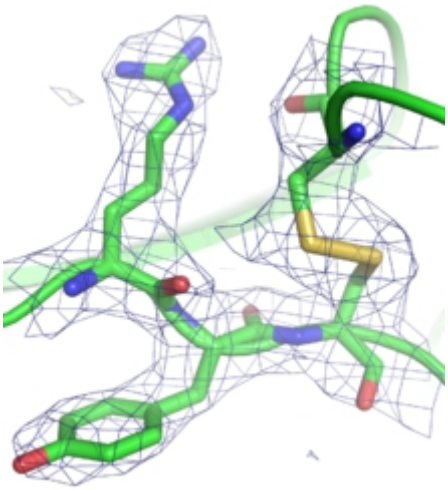


Fig. 3

Model building and phase refinement



The intensity of each diffraction 'spot' is recorded, and this intensity is proportional to the square of the *structure factor* amplitude. The structure factor is a complex number containing information relating to both the amplitude and phase of a wave. In order to obtain an interpretable *electron density map*, both amplitude and phase must be known. The phase cannot be directly recorded during a diffraction experiment: this is known as the phase problem. Initial phase estimates can be obtained in a variety of ways:

- ***Ab initio* phasing** or **direct methods** - This is usually the method of choice for small molecules (<1000 non-hydrogen atoms), and has been used successfully to solve the phase problems for small proteins.

- **Molecular replacement** - if a related structure is known, it can be used as a search model in molecular replacement to determine the orientation and position of the molecules within the unit cell.

- **Anomalous X-ray scattering** (*MAD or SAD phasing*) - the X-ray wavelength may be scanned past an absorption edge of an atom, which

changes the scattering in a known way. The most popular method of incorporating anomalous scattering atoms into proteins is to express the protein in a media rich in seleno-methionine, which contains selenium atoms.

- **Heavy atom methods** (multiple isomorphous replacement) - If electron-dense metal atoms can be introduced into the crystal, direct methods or Patterson-space methods can be used to determine their location and to obtain initial phases. Such heavy atoms can be introduced either by soaking the crystal in a heavy atom-containing solution, or by co-crystallization.

Having obtained initial phases, an initial model can be built. This model can be used to refine the phases, leading to an improved model, and so on. Given a model of some atomic positions, these positions and their respective Debye-Waller factors (or **B**-factors, accounting for the thermal motion of the atom) can be refined to fit the observed diffraction data, ideally yielding a better set of phases. A new model can then be fit to the new electron density map and a further round of refinement is carried out. This continues until the correlation between the diffraction data and the model is maximized. The agreement is measured by an *R*-factor defined as

$$R = \frac{\sum_{\text{all reflections}} |F_o - F_c|}{\sum_{\text{all reflections}} |F_o|}$$

A similar quality criterion is R_{free}, which is calculated from a subset (~10%) of reflections that were not included in the structure refinement. Both *R* factors depend on the resolution of the data. As a rule of thumb, R_{free} should be approximately the resolution in Ångströms divided by 10; thus, a data-set with 2 Å resolution should yield a final R_{free} ~ 0.2.

Circular dichroism (CD) refers to the differential absorption of left and right circularly polarized light (fig. 4). This phenomenon is exhibited in the absorption bands of optically active chiral molecules. CD spectroscopy has a wide range of applications in many different fields. Most notably, UV CD is used to investigate the secondary structure of proteins. UV/Vis CD is used to investigate charge-transfer transitions. Near-infrared CD is used to investigate geometric and electronic structure by probing metal d→d transitions. Vibrational circular dichroism, which uses light from the infrared energy region, is used for structural studies of small organic molecules, and most recently proteins and DNA.

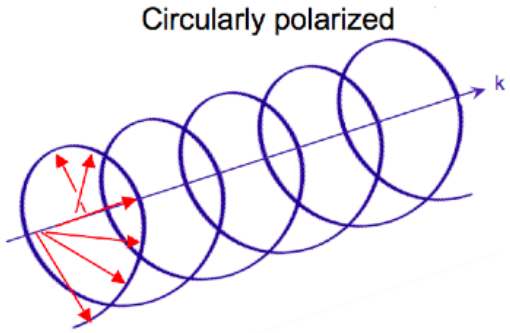
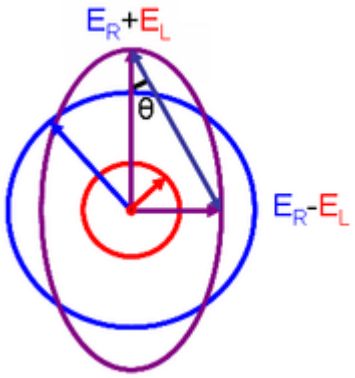


Fig. 4

Circular polarization of light -Electromagnetic radiation consists of an electric and magnetic field that oscillate perpendicular to one another and to the propagating direction. While linearly polarized light occurs when the electric field vector oscillates only in one plane and changes in magnitude, circularly polarized light occurs when the electric field vector rotates about its propagation direction and retains constant magnitude. Hence, it forms a helix in space while propagating. For left circularly polarized light (LCP) with propagation towards the observer, the electric vector rotates counterclockwise. For right circularly polarized light (RCP), the electric vector rotates clockwise.

When circularly polarized light passes through an absorbing optically active medium, the speeds between right and left polarizations differ ($c_L \neq c_R$) as well as their wavelength ($\lambda_L \neq \lambda_R$) and the extent to which they are absorbed ($\varepsilon_L \neq \varepsilon_R$). *Circular dichroism* is the difference $\Delta\varepsilon \equiv \varepsilon_L - \varepsilon_R$.

The electric field of a light beam causes a linear displacement of charge when interacting with a molecule (electric dipole), whereas the magnetic field of it causes a circulation of charge (magnetic dipole). These two motions combined cause an excitation of an electron in a helical motion, which includes translation and rotation and their associated operators.



Elliptical polarized light (purple) is composed of unequal contributions of right (blue) and left (red) circular polarized light.

This relationship is derived by defining the ellipticity of the polarization as:

$$\tan\theta = \frac{E_R - E_L}{E_R + E_L}$$

where

> E_R and E_L are the magnitudes of the electric field vectors of the right-circularly and left-circularly polarized light, respectively.

When E_R equals E_L (when there is no difference in the absorbance of right- and left-circular polarized light), θ is 0° and the light is linearly polarized. When either E_R or E_L is equal to zero (when there is complete absorbance of the circular polarized light in one direction), θ is 45° and the light is circularly polarized.

Although ΔA is usually measured, for historical reasons most measurements are reported in degrees of ellipticity. Molar circular dichroism and molar ellipticity, $[\theta]$, are readily interconverted by the equation:

$$[\theta] = 3298.2\,\Delta\varepsilon.$$

The CD may also be a function of temperature, concentration, and the chemical environment, including solvents. In this case the reported CD value must also specify these other relevant factors in order to be meaningful.

Additives, buffers and stabilizing compounds: Any compound which absorbs in the region of interest (250 - 190 nm) should be avoided. A buffer or detergent or other chemical should not be used unless it can be shown that the compound in question will not mask the protein signal. Therefore ensure that only the minimum concentration of additives are present in the protein solution.

Protein solution: The protein solution should contain only those chemicals necessary to maintain protein stability, and at the lowest concentrations possible**.** Avoid any chemical that is unnecessary for protein stability/solubility. The protein itself should be as pure as possible, any additional protein or peptide will contribute to the CD signal.

Contaminants: Unfolded protein, peptides, particulate matter (scattering particles), anything that adds significant noise (or artifical signal contributions) to the CD spectrum must be avoided.