
[next](#) [up](#) [previous](#) [contents](#) [index](#)

Next: [Multiple wavelength methods.](#) **Up:** [Protein Crystallography.](#) **Previous:** [Protein structure.](#)

Protein crystal structure determination.

The determination of a novel protein structure by X-ray crystallographic analysis involves the following steps. i) Protein purification and crystallisation of the native protein, ii) measurement of native diffraction data, iii) obtaining heavy atom derivatives, iv) measurement and analysis of derivative data, v) calculation of phases, vi) map interpretation and model building and vii) model refinement.

Crystallisation is often the rate limiting step in protein crystallography. Several methods of crystallisation are now well established but application of these methods is still very much trial and error. Crystallisation of a newly isolated protein can take weeks, months or years.

A prerequisite of the solution of the majority of protein crystal structures is to find an isomorphous heavy atom derivative of the native protein. Isomorphous, in this sense, means that the derivative protein crystal structure should be identical to the native protein structure except for the presence of one or more heavy atoms bound to protein molecules, i.e. the lattice, space group, cell dimensions and position and conformation of the protein molecule within the unit cell should be preserved. The most common method of heavy atom inclusion is to soak the native protein crystals in a solution containing a heavy atom compound. Other more recent methods involve the production of protein with modified amino acid residues, e.g. methionine containing proteins can be engineered where selenium replaces the sulphur of the methionine [50].

Once suitable native and derivative protein crystals become available, X-ray diffraction data are collected. One of several procedures may be adopted for data collection. Most single crystal diffraction data are measured using monochromatic X-rays from either a sealed tube generator, a rotating anode generator or from a synchrotron source. The availability of synchrotron radiation has led to the application of the less widely used Laue method of data collection to protein crystallography. It utilises the non-characteristic polychromatic radiation produced by a synchrotron. A detailed description of the Laue method may be found in [47].

Several data collection geometries may be used along with monochromatic radiation. A conventional diffractometer measures each reflection individually with a scintillation counter. A goniometer rotates the crystal so as to satisfy the Bragg condition for each reflection individually while the detector simultaneously records the diffracted X-ray intensity. Diffractometers are still widely used for small molecule work but have recently been superseded by area detectors which are able to measure

equivalent quantities of data in a much shorter time; an important factor when crystal samples are highly sensitive to the dosage of X-rays delivered during the experiment. This is particularly relevant for biological samples.

There are two area detector geometries which are at present widely used - rotation geometry and Weissenberg geometry. In both methods the Bragg condition is satisfied by rotating the sample crystal about a fixed axis. A series of diffraction images are measured whereby each image records all reflections satisfying the Bragg condition as the crystal is rotated through a specified angle $\Delta\phi_{max}$. A limited value of $\Delta\phi_{max}$ is necessary so as to avoid overlap of diffracted reflections on an image.

The essential difference is that a Weissenberg camera couples the sample rotation with a translation of the area detector. This helps avoid the problem of overlapping spots and uses the available space on the area detector more efficiently and can further reduce the overall time for data collection. The use of larger angular rotation ranges for each diffraction image in the Weissenberg geometry implies the accumulation of more background radiation per image, however large sample to detector distances, D , are also typically used which reduce the level of recorded background relative to the intensity of diffracted X-rays by a factor of $1/D^2$.

In a diffraction experiment one can only measure the intensities and diffraction angles of the diffracted beams. All information about the phases of the diffracted X-rays is lost. This phase information along with the amplitudes of the diffracted X-rays is essential for the solution of crystal structures and must be recovered.

There are four approaches which may be taken in recovering phase information in a diffraction experiment. The heavy atom method makes the assumption that if a significant contribution of the scattering from a structure is made by a heavy atom then the phases of the diffracted X-rays will be close to those phases which would be observed were only the heavy atoms present. The problem is thus reduced to finding the positions of the heavy atoms within the structure. This approach is in general not applicable to proteins since the heavy atom contribution to scattering is small with respect to the protein. The direct methods approach can make estimates about the reflection phases using assumptions about the internal structure of the crystal. Direct methods are routinely used for the solution of small structures and have only recently been applied to the solution of small proteins containing about 50 amino acid residues [102].

If a related or similar protein structure is already known then the method of Molecular Replacement (MR) [98] may be used. The idea is to find the rotation and translation which position the model structure in the unit cell so as to give the highest correlation between experimental diffraction measurements and those calculated from the model. This method relies on the existence of a known related structure and is therefore likely to become more and more applicable as the number of solved protein structures increases in the future.

The third and most prominent of the solutions to the phase problem in macromolecular crystallography is isomorphous replacement and related methods. In these methods phase information is retrieved by making isomorphous structural modifications to the native protein, usually by including a heavy atom or changing the scattering strength of a heavy atom already present and then measuring the diffraction amplitudes for the native protein and each of the modified cases. If the position of the additional heavy atom or the change in its scattering strength is known then the phase of each diffracted X-ray can be determined by solving a set of simultaneous phase equations. Methods which use such a strategy are Single Isomorphous Replacement (SIR), Multiple Isomorphous Replacement (MIR), Single Isomorphous Replacement with Anomalous Scattering (SIRAS) and within the last 15 years, the Multiple wavelength Anomalous Diffraction method (MAD).

With an experimental set of phases obtained from either direct methods or Isomorphous Replacement related methods one can calculate a 3-dimensional electron density map of the protein structure. This is not always readily interpretable as a single polypeptide chain and methods are usually employed to improve the density map using knowledge about the common characteristics of protein crystals. e.g. they nearly always contain between **30** and **70%** solvent, are made of individual amino acid residues with known structures and have predictable secondary structures e.g. α -helix or β -pleated sheet. Density interpretation and model building have been semi-automated with the recent development of powerful graphical computer hardware and software aimed specifically at macromolecular modelling, e.g. the program O [61]. After and during the main phase of model building, refinement of the model is carried out against the experimentally measured intensities. This stage may include the addition of ordered solvent molecules and if very high resolution X-ray data are available even the addition of hydrogen atoms, although this is rare for macromolecular structures.

[next](#) [up](#) [previous](#) [contents](#) [index](#)

Next: [Multiple wavelength methods.](#) **Up:** [Protein Crystallography.](#) **Previous:** [Protein structure.](#)

Gwyndaf Evans

Fri Oct 7 15:42:16 MET 1994