Protein crystallography

X-ray crystallography is essentially a form of very high resolution microscopy. It enables us to visualize protein structures at the atomic level and enhances our understanding of protein function. Specifically we can study how proteins interact with other molecules, how they undergo conformational changes, and how they perform catalysis in the case of enzymes. Armed with this information we can design novel drugs that target a particular protein, or rationally engineer an enzyme for a specific industrial process.

In all forms of microscopy, the amount of detail or the resolution is limited by the wavelength of the electro-magnetic radiation used. With light microscopy, where the shortest wavelength is about 300 nm, one can see individual cells and sub-cellular organelles. With electron microscopy, where the wavelength may be below 10 nm, one can see detailed cellular architecture and the shapes of large protein molecules. In order to see proteins in atomic detail, we need to work with electro-magnetic radiation with a wavelength of around 0.1 nm or 1 Å, in other words we need to use X-rays.

In light microscopy, the subject is irradiated with light and causes the incident radiation to be diffracted in all directions. The diffracted beams are then collected, focused and magnified by the lenses in the microscope to give an enlarged image of the object. The situation with electron microscopy is similar only in this case the diffracted beams are focused using magnets. Unfortunately it is not possible to physically focus an X-ray diffraction pattern, so it has to be done mathematically and this is where the computers come in. The diffraction pattern is recorded using detector which used to be X-ray sensitive film, but nowadays is usually an image plate or a charge-coupled device (CCD).

The diffraction from a single molecule would be too weak to be measurable. So we use an ordered three-dimensional array of molecules, in other words a crystal, to magnify the signal. Even a small protein crystal might contain a billion molecules. If the internal order of the crystal is poor, then the X-rays will not be diffracted to high angles or high resolution and the data will not yield a detailed structure. If the crystal is well ordered, then diffraction will be measurable at high angles or high resolution and a detailed structure should result. The X-rays are diffracted by the electrons in the structure and consequently the result of an X-ray experiment is a 3-dimensional map showing the distribution of electrons in the structure.

A crystal behaves like a three-dimensional diffraction grating, which gives rise to both constructive and destructive interference effects in the diffraction pattern, such that it appears on the detector as a series of discrete spots which are known as reflections. Each reflection contains information on all atoms in the structure and conversely each atom contributes to the intensity of each reflection. As with all forms of electro-magnetic radiation, X-rays have wave properties, in other words they have both amplitude and a phase. In order to recombine a diffraction pattern, both of these parameters are required for each reflection. Unfortunately, only the amplitudes can be recorded experimentally – all phase information is lost. This is known as "the phase problem". When crystallographers say they have solved a structure, it means that they have solved "the phase problem". In other words they have obtained phase information sufficient to enable an interpretable electron density map to be calculated.

Protein preparation

Firstly protein needs to obtain a pure sample of target study. This can by either isolating it from its source, or by cloning its gene into a high expression system. The sample then needs be assessed for suitability according to the following criteria:

1. pure and homogeneous – can test this by various electrophoretic methods and mass spectrometry.
2. protein soluble and folded? – if protein estimations suggest that a lot of protein is being lost, then it may be due to precipitation. The degree of ordered secondary structure can be tested with circular dichroism – if this is very low then the protein may be misfolded. This may occur if the protein is being produced faster than it can fold and may result in the formation of insoluble inclusion bodies. Attenuating the induction can alleviate this problem e.g. using a lower temperature.
3. Sample monodisperse – in other words is the sample free from aggregation? This can be monitored using a dynamic light scattering (DLS) device.
4. protein still active – check with activity assays
5. Sample stability: Occasionally good protein crystals will form overnight at room temperature, but usually it may take several days to one or two weeks before suitable crystals can grow. Therefore, ideally the sample needs to remain stable over that period

If the sample fails one or more of the above criteria, it may be worthwhile returning to the expression and purification protocols and trying something different, such as the addition of ligands known to interact with the protein, or adding extra purification steps. In extreme cases it may be worthwhile switching to a different expression system altogether or working with a mutated or truncated construct. It may be possible to refold protein successfully using chaotropic reagents such as urea. Aggregated or polydisperse samples may be made monodisperse by simply changing pH or adding some salt. However, without DLS, this is very difficult to assess.

Crystallization

Before beginning trials the sample needs to be concentrated and transferred to dilute buffer containing little or no salt if the protein is stable and active under these conditions. This can easily be achieved using centrifugal concentrators. In order to screen a reasonable number of conditions we need at least 200 µL of protein at 10 mg/mL. If this is not the case then need to scale up the expression and purification to make it required concentrations.

If a similar protein has already been crystallized then it is definitely worth trying the conditions used to grow crystals of this protein. In any case if you have enough material one would normally subject it to one or more sparse matrix screens. To date the total number of different conditions in our repertoire of screens comes to about 400.

We normally use these tissue culture trays to set up crystallizations with up to 24 different conditions per tray. The method used is **hanging drop vapour diffusion** – it has the advantage of being the least expensive on protein. The set up is as follows:

The well is prepared first and usually contains 1ml of a buffered precipitant solution such as polyethylene glycol or ammonium sulfate – or even a mixture of PEG and salt. Sometimes additives are also included such as detergents or metal ions which may enhance the crystallization. Then 1 µL of the concentrated protein sample is pipetted onto a siliconized cover slip, followed by 1 µL of the well solution. The coverslip is then inverted over the well and sealed using a bead of vacuum grease. This is then left undisturbed for at least 24 hours to equilibrate. At the start of the experiment, the precipitant concentration in the drop is half that of the well. Equilibration then takes place via the vapour phase. Given the relatively large volume of the well, its concentration effectively remains the same. The drop however loses water vapour to the well until the precipitant concentration equals that of the well. Hopefully, if the conditions have been favorable, at some point during this process the protein has become supersaturated and been driven out of solution in the form of crystals. All too often however these trials result in precipitate or the formation of salt crystals, or nothing happens at all and the drops remain clear. It is recorded to be the success rate at this stage is less than 0.1%.

If no promising leads are found then there are several possible courses of action. Various types of buffers and conditions can be changed which may affect crystallization. Working at work at a different temperature – temperature can have a profound effect on protein solubility. Temperatures of 4° C and 18° C are typically used. If we have already been round this cycle more than once, it may be time to go back to the purification and expression and try something different, such as working with a fragment of our target protein.

Essentially the procedures need to refine all variables and possibly introduce some new ones in order to achieve our goal, which is large, single crystals (see below). Things to try at this stage include varying the concentrations of all components in the crystallization, slight pH changes, using additives, switching to similar buffers or precipitants, or even using different crystallization methods (e.g. dialysis). Occasionally good crystals will form overnight, but more typically they will take from several days to several weeks to grow.

Testing crystals

Once the crystals were obtained, these crystals can be tested with X-rays. In some cases we may wish to modify the crystals in some way prior to testing such as oxidizing or reducing, or binding a ligand. We may also wish to bind heavy atoms. The crystal then needs to be mounted in some way either in a capillary at room temperature or flash-cooled to 100 K in a loop - nowadays most data collection is done using this method - and then attached to a device called a goniometer head which enables the sample to be accurately positioned in the X-ray beam by means of a number of adjustment screws. For cryogenic data collection, a cold nitrogen gas stream keeps the crystal at 100 K throughout the experiment. Focused X-rays emerge from a narrow tube called a collimator and strike the crystal to produce a diffraction pattern. This is recorded on the X-ray detector. If the proper crystals are formed without distortions clean sharp spots will be obtained, one lattice of spots indicating you have a single crystal (see below), and no evidence of salt or ice crystals which give rise to very strong spots or rings. Ideally the crystal should diffract to better than 4 Å resolution. If any if these criteria are not satisfied, then it's time to try another crystal. If all the crystals are the same then you may need to go back to the crystallization stage.

X-ray data collection

Firstly we need to ascertain the crystal symmetry, the unit cell parameters, the crystal orientation and the resolution limit. Armed with this information we derive a data collection strategy which will maximize both the resolution and completeness of the data set. The method we use is to rotate the crystal through a small angle, typically 1 degree, and record the X-ray diffraction pattern. If the diffraction pattern is very crowded, then the rotation angle should be reduced so that each spot can be resolved on the image. This is repeated until the crystal has moved through at least 30 degrees and sometimes as much as 180 degrees depending on the crystal symmetry. The lower the symmetry, then more data are required. A typical medium resolution data set may take up to 3 days using an 'in house' X-ray source. For high resolution data collection you need to go to a synchrotron where the X-ray intensity is greater and therefore data collection times are shorter - sometimes as fast as 10 minutes! This is the point where we become heavily dependent on computers - every spot on each image needs to be measured. This is a formidable task. Some of our nitrogenase data sets contain around 300 images with over 5000 spots per image.

Structure solution

In order to visualize our structure we need to solve the phase problem, in other words we need to obtain some phase information. For protein structure determinations we can do it in one of several ways:

If we already have the coordinates of a similar protein we can try to solve the structure using a process called Molecular Replacement which involves taking this model and rotating and translating it into our new crystal system until we get a good match to our experimental data. If we are successful then we can calculate the amplitudes and phases from this solution which can then be combined with our data to produce an electron density map. This is effected using a Fourier transform and is equivalent to focusing the diffraction pattern in other forms of microscopy.

If we have no starting model, then we can use Isomorphous Replacement methods whereby one or more heavy atoms are introduced into specific sites within the unit cell without perturbing the crystal lattice. This is another trial-and-error procedure and often it is not apparent whether it has worked until more X-ray data have been collected. Heavy atoms are electron dense and give rise to measurable differences in the intensities of the spots in the diffraction pattern. By measuring these differences for each reflection, it is possible to derive some estimate of the phase angle using vector summation methods. In practice data from one or more heavy atom derivatives is required to get good enough phases - hence Multiple Isomorphous Replacement (MIR). Again we use a Fourier transform to calculate a map.

In some cases we can make use of the anomalous scattering behavior of certain atoms at or near their X-ray absorption edges to gain useful phase information. Many of the atoms used in isomorphous replacement are also useful in this respect. This additional information can enhance the structure solution. Multiwavelength Anomalous Dispersion (MAD) is an elegant and often very effective method that relies entirely on the measurement of the anomalous differences produced by one or more anomalously scattering atoms in the crystal. In practice three or more

consecutive data sets are recorded from the same crystal at different wavelengths around the X-ray absorption edge of the anomalous scatterer. As this method requires a tuneable X-ray source, it can only be performed at a synchrotron. The resultant phase information can often produce very high quality electron density maps, thereby simplifying the subsequent interpretation. Selenium is a particularly good anomalous scatterer and it can be incorporated into proteins by over-expressing them in strains of *E. coli* that are auxotrophic for methionine. The host cells are then grown on minimal media supplemented with amino acids using selenomethionine in place of methionine.

Model building

This is the process where the electron density map is interpreted in terms of a set of atomic coordinates. This is more straightforward in the molecular replacement case because we already have a coordinate set to work with. In the case of isomorphous replacement we simply have the map. It is essentially a 3-dimensional jigsaw puzzle with the pieces being the amino acid residues. The normal procedure is to fit a protein backbone first then if the resolution permits, we insert the sequence. The amount of detail that is visible is dependent on the resolution and the quality of the phases. Shown below is a high resolution electron density map with atomic coordinates superposed. Often regions of high flexibility are not visible at all due to static disorder, where the structure varies from one molecule to the next within the crystal, or dynamic disorder, where the region is mobile within the crystal. The latter type of disorder is eradicated in cryogenic data collection.

Refinement

Once we have a preliminary model we can refine it against our data. This will have the effect of improving the phases which results in clearer maps and therefore better models. We would typically go round this cycle several times until we get little or no further improvements. At this stage we would expect a crystallographic R-factor of below 25%. This is a measure of the agreement between the model and the data – the lower the value the better the model. Nevertheless, the final model must make chemical sense and there must be no large regions of electron density unaccounted for.

Microscopic view of protein crystals   X-ray diffraction image of protein crystal
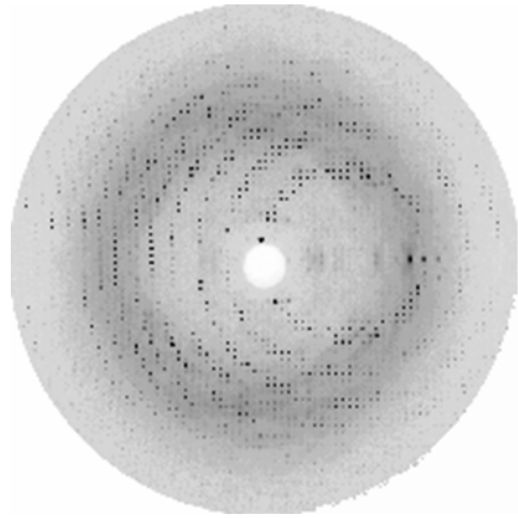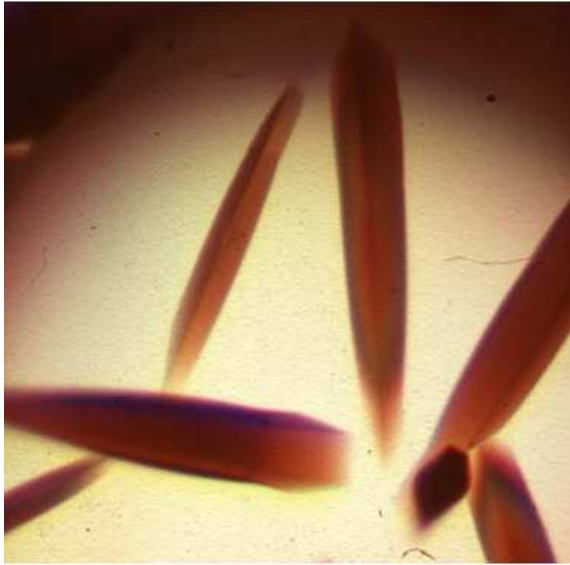
Electron density cloud obtained from   final protein sturcture

Diffraction data