X-ray crystallographic studies of protein–ligand interactions

R.A. Palmer and H. Niwa
School of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, U.K.

Abstract
X-ray crystallography enables details of covalent and non-covalent interactions to be analysed quantitatively in three dimensions, thus providing the basis for the understanding of binding of ligands to proteins as well as modes of action such as cell-surface binding. This article is concerned with current methods employed for the X-ray analysis of protein structures complexed with ligands. It deals mainly with ‘what can be done’ in current research, rather than providing details of ‘how to do it’. In recent years significant advances have been made in a variety of techniques: growing protein crystals from very small samples by scanning a wide range of conditions; X-ray intensity data collection and measurement through the use of charge-coupled devices and high-intensity, versatile synchrotron sources; cryo-crystallography which both stabilizes the crystals and provides improved data; methods for analysing and interpreting the structures, dependent, at least in part, on both structural and sequence databases; and improvements in hardware and software. To illustrate the type of results achievable two examples involving protein–sugar interactions are discussed: (i) SNAII (the lectin Sambucus nigra agglutinin-II from elder) N-terminal sugar-binding site where terminal sugar units in a glycosylation chain from a symmetry-related molecule bind and (ii) MLI (mistletoe lectin I) C-terminal sugar-binding site with lactose.

Introduction

The branch of X-ray crystallography, which has developed around the study of macromolecular structures, arose from the need in the 1940s and 1950s to extend the methods of small-molecule analysis. An important objective for determining the three-dimensional structure of a protein is to gain an understanding of its biological activity, which includes in many cases characterization of ligand-binding properties. X-ray crystallography enables details of covalent and non-covalent interactions to be analysed quantitatively in three dimensions, and is therefore ideally suited for this purpose, enabling further insights into other modes of action such as cell-surface binding.

Scheme 1 shows a flow diagram of the procedures used in the first stages of the analysis. Specific topics to be covered here include: methods for growing crystals; collection, measurement and use of the X-ray data; methods for analysing, refining and interpreting the structures; and finally examples from current research illustrating the binding of sugars to RIP II proteins (type 2 ribosome-inactivating proteins). Emphasis will be placed on what can be done using current research methods.

Key words: electron density, ligand binding, multiple isomorphous replacement, molecular replacement, refinement, X-ray crystallography.

Abbreviations used:
MIR, multiple isomorphous replacement; MR, molecular replacement; CCD, charge-coupled device; RIP II, type 2 ribosome-inactivating protein; MLI, mistletoe lectin I; SIRAS, single-isomorphous replacement with anomalous scattering; SNAII, lectin Sambucus nigra agglutinin-II from elder.

1To whom correspondence should be addressed (e-mail r.palmer@pgen.net).

Basic requirements

The viability of a proposed new protein structure determination depends in the first place on the following.

The availability of 5 mg or more of pure protein
An on-going supply of fresh material is usually required, especially if the need for new experiments arises, for example location of bound ligands or heavy atoms. Purity is essential for production of the best diffraction-quality crystals.

Knowledge of the complete amino acid sequence

Sequence data will be necessary in the later interpretation of electron density maps for model building. If the method of MR (molecular replacement) is to be used it will also be required for database location of homologous search structures; for MIR (multiple isomorphous replacement) it will help to pin-point possible heavy atom compounds to try. Any additional information on the characteristics of the protein will be invaluable, such as possible ligand-binding sites.

Growing crystals

Protein crystals usually form by nucleation at extremely high levels of supersaturation (100–1000%; small-molecule crystals nucleate at only a few percent of supersaturation). Work with proteins, however, usually means that the starting material is both expensive and sparse (20 mg is a large amount of protein). The requirement of high levels of supersaturation presents problems due to the formation of unwanted
amorphous precipitates. There is competition between crystals and precipitates at both nucleation and growth stages. High supersaturation can be achieved using a variety of precipitants, which include: (i) (NH₄)₂SO₄, (ii) polyethylene glycol, (iii) methylpentanediol and (iv) NaCl.

When starting from non-saturated solutions, supersaturation can be reached by varying parameters such as temperature or pH. Conditions for optimal nucleation are not the same as those for optimal growth. Nucleation can also lead to epitaxial growth (growth of crystals on other crystals), which is detrimental to crystal quality. Interface or wall effects and the shape and volume of drops when using hanging or sitting drops can affect nucleation or growth. This means that the geometry of crystallization chambers or drops can be important.

**Microbatch screening**

Initial experiments are conducted on microbatch plates with small sample volumes of around 1–3 µl and protein concentrations of around 5–10 mg/ml, allowing many different crystallization conditions to be screened. Each plate contains a number of small wells in the form of a matrix so that individual wells can be easily labelled for future reference. This is important, as crystals will only form in a very small number of wells, perhaps only one. Drops containing protein and precipitant solution are pipetted under a layer of silicon (polydimethylsiloxane) or paraffin oil (or a 1:1 mixture of the two). All reagents are present at a specific concentration and no significant concentration of the protein or the reagents can occur in the drop. Reagent-sampling kits are available commercially. Diffusion of water from the drop takes place through the oil, thus changing the concentration in the drop, hopefully towards the required crystallization conditions. Once established the localized well conditions can be refined to optimize crystal size and quality for X-ray diffraction using one of the techniques described below.

**Vapour-diffusion techniques**

The hanging-drop method is possibly the most popular version of this technique. It may be necessary, however, to try other variants of the method. Kits are available commercially for carrying out the different methods.

**Hanging-drop method**

Each drop is set up by being rapidly inverted over a prepared well where it hangs by surface tension. Wells are in the form of a matrix, typically 4 × 6. A great variety of plates are available, many of which are also adaptable for the sitting-drop technique. In the hanging-drop method the droplet (5–20 µl) containing the macromolecule, a buffer and a precipitating agent is equilibrated against a reservoir (1–25 ml) containing a solution of the same precipitant at a higher concentration than the droplet (for example, by a factor of 2). Equilibration proceeds by evaporation of the volatile species until the vapour pressure of the droplet equals that of the reservoir. Crystals form in the droplet.

**Sitting drop**

The principle here is essentially the same as that in the hanging-drop technique except that the drop, in this case, sits on a bridge over the precipitant reservoir.

**Vapour-diffusion rate of control**

A layer of oil can be used with hanging- or sitting-drop techniques to limit the rate of vapour diffusion. 200 µl of paraffin or silicon oil or a mixture of the two is applied over the reservoir solution. Varying the composition of the mixture provides additional control over the vapour-diffusion rate.

**Screening crystallization conditions**

The method allows a broad range of salts, polymers and organic solvents over a wide range of pH values to be sampled. When crystals are obtained a finer screening can be used to produce X-ray-diffraction quality crystals. Ready-to-use reagents formulated from salts, buffers and precipitants at various concentrations are available commercially.

**How to improve the crystals**

Where poor (or no) crystals have appeared initially, the following may be helpful. (i) Vary buffer, pH, precipitant,
Protein concentration, drop size, the method used and temperature (4, 22 and 37°C are commonly used). (ii) Try seeding: transfer finely crushed crystal particles into the wells using a cat’s whisker or hair. (iii) Try adding Cu²⁺, Zn²⁺, Ca²⁺, Co²⁺, EDTA, acetone, dioxane or phenol; for membrane proteins try n-octyl-β-D-glucopyranoside (up to critical micelle concentrations) and substrates, cofactors, inhibitors or binding sugars. (iv) If large crystals are obtained but are very sensitive to exposure to the X-ray beam freeze the crystals and book a session on a synchrotron facility. (v) Book a flight on a satellite rocket and try gravity-free conditions. This is a proven method but requires meticulous planning.

For further details of crystallization techniques see [1–3].

Protein complex crystals with ligands or heavy atoms
Complex crystals may be formed either by co-crystallization or soaking under controlled conditions. For co-crystallization the material is added into the protein solution prior to crystallization. Crystals grown in this way may suffer changes and are of no use in MIR but may be preferred for ligand binding as it is more likely to lead to full occupation of the binding sites. A wide range of heavy-atom reagents is available for MIR [4, 5]. These are mainly compounds of Hg, Au, Pt or U. The initial choice of possible compounds may be influenced by knowledge of the protein’s amino acid sequence [6].

Crystal mounting for X-ray data collection
Room temperature
For X-ray data collection at room temperature, protein crystals are usually mounted in sealed thin-walled glass capillary tubes in the presence of a drop of mother liquor to stabilize the crystal and prevent escape of water of crystallization, which is fatal. Crystal lifetime can be anything from a few minutes to several days under normal laboratory conditions. Exposure to the highly intense beam from a synchrotron source usually decreases crystal lifetime dramatically. This, however, is compensated by the increased intensity, which permits shorter individual exposures and consequent increase in the quantity of data recorded.

Cryo-crystallography
When the temperature of a crystal is lowered the thermal motion of the constituent atoms becomes less marked and the X-ray diffraction pattern can be improved both in intensity and resolution. Many proteins can be studied using cryo techniques in which the temperature is reduced to around 100 K. A popular and efficient cryo system is produced by Oxford Cryosystems (Cryostream Cooler; http://www.oxford-diffraction.com). Damage from freezing is prevented through the use of a cryoprotectant liquid which may be incorporated in the mother liquor after crystallization or, less commonly, as a component of the crystallization reagents. The procedure involves supporting the crystal in a film of cryoprotected mother liquor in a small fibre loop, which is rapidly cooled, free from ice, to around 100 K in a stream of nitrogen gas. The liquid surrounding the crystal must freeze as an amorphous glass to avoid crystal damage and diffraction from ordered ice crystals. Glycerol is most frequently used as a cryoprotectant. The required glycerol concentration must be carefully established, and unfortunately this may involve loss of several crystals in the early stages of the experiment. Once frozen, protein crystals are usually extremely stable, transportable and can be stored and kept ready for subsequent X-ray diffraction experiments.

Example conditions
For the low-temperature data collection of MLI [mistletoe (Viscum album) lectin I] with bound galactose sugar, a complex mixture consisting of 0.1 M glycine buffer at pH 3.4, with 0.9 M (NH₄)₂SO₄, 0.05 M galactose and 30% glycerol is prepared, soaking for a few minutes prior to flash freezing [7].

Recording X-ray diffraction from macromolecular crystals
X-ray sources
Determination of the structure requires measurement of the X-ray diffraction intensities from the crystal. Generally speaking, the more data collected, the better will be the quality of the final structure. For even moderately sized proteins this will involve possibly 20 000–50 000 measurements. This work can be carried out either in house using rotating-anode radiation or, more rapidly, in view of the enormously increased beam power, at a synchrotron installation (for information see http://lmb.biop.ox.ac.uk/www/synchr.html). Use of low-temperature data collection should enhance both crystal life time and the quality of the measurements. The use of synchrotron radiation also enables a greater variety of experiments to be carried out. These include SIRAS (single-isomorphous replacement with anomalous scattering), which provides good isomorphous replacement phasing from a native crystal and one derivative; time-resolved experiments; and MAD (multiple-wavelength anomalous dispersion), which requires the data to be collected, in the same session, at two different wavelengths from the same crystal. MAD is often used to solve the phase problem for proteins containing Se-mutated methionine residues (see for example [8]).

Diffractometers
Serial diffractometers
Traditional diffractometers incorporate a mechanical goniometer to orientate the crystal for each reflection and to rotate the counter to receive the scattered X-radiation from this single reflection. The process is very slow, but, arguably, the accuracy attainable is better than for most of the other methods. This however is offset by the need to use several
crystals. The Enraf Nonius CAD4 4-circle diffractometer, in recent years, has been one of the most popular machines available. Other diffractometers available include models from Siemens, Rigaku, Philips and Mac Science (now merged with Bruker-AXS).

**Area detectors**
The introduction of ‘electronic film’ area detectors and image plates, which have enjoyed rapid development in recent years, now enables protein data to be collected in several hours rather than days. Principal types of area detector are as follows. (i) Xenon gas/multiwire proportional counter (MWPC). (ii) FAST area detector (Enraf Nonius FAST), in which a fibre-optics screen is used to convert the X-ray signals into light photons which are intensified, integrated and digitized. (iii) Imaging plates made up of barium halide phosphor doped with Eu$^{3+}$, which is excited into the metastable Eu$^{3+}$ state by exposure to X-radiation. After exposure, the plate is scanned with a fine He–Ne laser beam, which causes these regions to emit violet light which is then detected with a photomultiplier system, integrated and digitized. After reading off the stored data the plate is cleaned by exposure to bright yellow light and the ‘film’ is then ready to be used to record the next image. Advantages of this method include the following. (a) Size; plates up to 30-cm diameter are available which enable data to be collected to a resolution of about 1.4 Å using Cu radiation and better with synchrotron radiation at smaller wavelengths. (b) Intensity range. A very wide range of intensities (approx. $10^5$) can be recorded compared with approx. $2 \times 10^2$ for X-ray film. (iv) CCDs (charge-coupled device) area detectors. (a) Phosphor screen-demagnifying fibre-optics-coupled CCDs. CCDs combine high sensitivity over a wide range of intensities with the capability of very fast readout times (10 s compared with 150 s for laser-read phosphor plates), low noise level (through cooling of the CCD chips to $-20$ to $-50$°C) and a capacity for storing large amounts of data. In operation the principle is similar to that used with imaging plates, and data processing is also very similar. Amongst the commercially available demagnifying fibre-optic-coupled CCD diffractometers are those produced by Bruker, ADSC, Nonius and MAR Research. (b) Electronically focused CCD’s (at ESRF, European Synchrotron Facility, Grenoble, France). This device is essentially a CCD incorporating electronic focusing in which the phosphor screen is imaged through an image intensifier and demagnified using CRT (cathode ray tube) technology. Magnetic shielding is required [9].

**Structure determination**
There are two commonly used methods for structure determination in macromolecular crystallography, MIR and MR.

**MIR**
A heavy-atom derivative may be made by incorporating one or more heavy atoms [3–6,18,19] into a protein crystal. Locating the heavy-atom positions is relatively straightforward and can lead, by employing changes measured in the diffraction pattern, to a first structure model. For the derivative crystal to be isomorphous (a) the heavy atoms must not disturb the protein structure significantly and (b) no significant changes may have taken place in the crystal parameters.

**Preparation and screening of heavy-atom derivatives**
Two methods are used for the preparation of heavy-atom derivatives: (i) soaking pre-grown native crystals in heavy-atom solutions and (ii) co-crystallization of the protein and heavy atoms together from solution. Method (ii) is most likely to lead to the breakdown of condition (b) (see the last section). The method is time-consuming and experimentally intensive. The use of synchrotron radiation, employing SIRAS or MAD (multi-wavelength anomalous dispersion), can reduce the overall number of crystals, derivatives and data collections.

**MR**
This method [10] requires (a) measured X-ray diffraction data for the protein crystal under investigation (target structure) and (b) co-ordinates of a good-quality structure (search structure) having similarities with the target structure. There are currently over 20 000 protein structures deposited in the Protein Data Bank (PDB at http://www.rcsb.org). In practice, a search structure, either from a crystallographic or an NMR study, having 40% homology with the target structure is worth trying. Sequence homology as low as 20% may be sufficient, especially if many of the sequence changes are conservative. Sequence homology may also be used in less obvious cases of structural similarity (for example proteins not belonging to the same or related families) but carries a greater risk of failure. It may also be possible to use a partial MR search model, which includes only the region of optimum correspondence, or proteins stripped to only main-chain atoms [11]. Partial sequence searches can be carried out using BLAST (basic local alignment search tool; http://www.ncbi.nlm.nih.gov/BLAST). Programs such as COMPOSER [12] can be used as a means of optimizing a model prior to the MR search. One of the major defects of MR is that of model bias. This describes the tendency of a model, which differs significantly, at least in parts, from the target structure, to retain its own structure too rigidly even when subjected to intensive refinement. The popular MR program AMoRe [13] uses rigid body refinement while X-PLOR and CNS [14–17] employ simulated annealing. Both strategies may produce MR solutions closer to a refineable model. The refinement process is also monitored using a free R-factor, which indicates if model bias is a problem. Other popular MR programs include BEAST, MOLREP and EPMR. Each program offers different protocols for applying the method.
Completing the structure and searching for ligands

The features of protein electron-density maps, which are generally far short of atomic resolution, may be summarized as follows: (i) main-chain density tends to be continuous and relatively strong; (ii) side-chain atoms tend to be less rigidly held in the structure than main-chain atoms, with correspondingly weak electron density; (iii) the most disordered regions tend to correspond to the solvent regions which may comprise some 35–70% of the structure and will consequently be associated with very weak electron density.

Extra ligands, both covalent and non-covalent, should be revealed once the protein model is sufficiently well refined. Refinement is a necessary step and must be carried out with extreme care, following a protocol such as that in Scheme 2. Basically, what is being done here is to obtain the best fit between the model and the data. This requires the model to be as complete as possible and ensures that the errors in the model parameters are minimized. Identification and modeling of the associated ligands is part of this process. Once identified the ligand atoms are added to the refinement. Ligand geometry and interactions with neighbouring parts of the structure are checked and analysed as for other parts of the structure through the atom co-ordinates. Further details of the above procedures may be found in [3,18,19].

Scheme 2 | Refinement protocol

Initially refinement should be carried out on the electron density (Fourier refinement). Once convergence has been reached least squares (e.g. RESTRAIN or SHELEX-97) or simulated annealing (X-PLOR) can be used in step IV. The cycle IV-II-III-IV is repeated until no further changes occur in the structure, i.e. convergence is again reached. Program names are in bold.

Examples of protein–ligand binding: RIP II B-chains

MLI from mistletoe is an AB toxin, belonging to the same RIP II family as ricin. It consists of two protein chains linked by an S-S bridge: the A-chain (252 residues, 28.6 kDa) enzymically arrests protein synthesis by depurinating a specific ribosomal adenine; the B-chain (263 residues, 31.2 kDa), like ricin, is a Gal-/GalNAc-specific lectin that triggers internalization of the toxin into a cell by binding cell-surface sugars. MLI is unusual in that it exists as a homodimer, which may have important consequences for its biological function [20]. The crystal structure of MLI with galactose has been determined at 100 K temperature to 3.0 Å resolution by molecular replacement [7]. The structure of MLI with lactose has been determined at 100 K at 2.5 Å resolution, the lactose binding for which is illustrated here (Figures 1 and 2) for comparison with SNAII (lectin Sambucus nigra agglutinin-II from elder; see below).

A number of RIPS and their related proteins have been extracted from elder (S. nigra). The structure of the Gal-/GalNAc-specific SNAII has been determined at 2.3 Å. It is identical to the B-chain lectin domain of a co-existing RIP II (258 residues, 30 KDa) in the same tissue and

Figure 1 | The SNAII N-terminal sugar-binding site where a terminal xylose unit in the glycosylation chain at Asn-63 from a symmetry-related molecule binds

Table 2: Refinement protocol. Initially refinement should be carried out on the electron density (Fourier refinement).

*Once convergence has been reached least-squares (e.g. RESTRAIN or SHELEX-97) or simulated annealing (X-PLOR) can be used in step IV. The cycle IV-II-III-IV is repeated until no further changes occur in the structure, i.e. convergence is again reached. Program names are in bold capitals.
Figure 2 | (Top panel) The MLI N-terminal sugar-binding site with lactose and (bottom panel) the MLI C-terminal with bound lactose.

refinement co-ordinates of the SNAII model include sugar units shown in bold in an oligosaccharide, which is also found in MLI [21] (see Scheme 3).

Differences in fine sugar-specificity amongst MLB (MLI B-chain), RTB (ricin B-chain) and SNAII: MLB has specificity for Gal but little for GalNAc; SNAII has higher specificity for GalNAc than Gal; RTB has specificity for both Gal and GalNAc to a similar extent in its high-affinity site but its lower-affinity site has specificity only for Gal. The differences in sugar specificity between the two sites in MLB and also in SNAII are not yet known.

Figure 1 shows the SNAII N-terminal sugar-binding site where a terminal xylose unit in the glycosylation chain at Asn-63 from a symmetry-related molecule binds. Figure 2 (top panel) shows the MLI N-terminal sugar-binding site with lactose and Figure 2 (bottom panel) shows the MLI C-terminal sugar-binding site also with bound lactose. In these figures the contours represent difference electron density, calculated without the co-ordinates of the sugar molecules. For SNAII N-terminal sugar the contour level is 4σ, for MLI N-terminal 3σ and for MLI C-terminal 1.75σ. This means that the MLI C-terminal density is quite weak compared with that for both of the N-terminal binding sites. This may indicate that the MLI C-terminal site is only partially occupied. Figures 1 and 2 were produced with SETOR [22].

We thank the School of Crystallography, Birkbeck College, London, U.K., for provision of laboratory and computing facilities. Data were collected at CLCC Daresbury Laboratory, Cheshire, U.K. We are also indebted to Edel Sweeney, Colin Reynolds, Alex Tonevitsky, Steve Saward, Jose’ Saldanha and Willie Peumans for their continued help and interest. On request, useful additional material, including suppliers and databases, can be obtained from the author.

References
4 Blundell, T.L. and Jenkins, J.A. (1977) Chem. Soc. Rev. 6, 139

Received 18 May 2003