View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Theseus

Bachelor's thesis (AMK)

Biotechnology and Food Technology

Biotechnology

2011

Jesse Mattsson

DOCUMENTATION OF PROTEIN CRYSTALLIZATION TECHNIQUES AND CHARACTERIZATION OF CRYSTALS

TURUN AMMATTIKORKEAKOULU
TURKU UNIVERSITY OF APPLIED SCIENCES

OPINNÄYTETYÖ (AMK) | TIIVISTELMÄ

Turun ammattikorkeakoulu

Bio- ja elintarviketekniikka | Biotekniikka

2011 | Sivumäärä 68

Ohjaajat Kari Haajanen, lehtori, Tassos Papageorgiou, FT

Jesse Mattsson

PROTEIINIEN KITEYTTÄMISTEKNIIKOIDEN DOKUMENTOINTI JA KITEIDEN KARAKTERISOINTI

Röntgenkristallografia on yleinen tekniikka, jolla voidaan selvittää proteiinien ja proteiini-kompleksien kolmiulotteinen rakenne. Se perustuu röntgensäteiden sirontaan proteiineista muodostuvien kiteiden sisältämien atomien elektroneista.

Periaatteena on ensin kasvattaa sopiva proteiinikide, johon kohdistetaan intensiivinen röntgensäde. Säteiden sironta mitataan ja saatua dataa käytetään molekyylin rakennetutkimuksessa. Tuotettuja malleja voidaan käyttää apuna esim. lääkkeiden suunnittelussa ja paikkakohtaisessa mutageneesissä.

Tämä opinnäyte tehtiin Turun biotekniikan keskuksessa (BTK), joka on Turun yliopiston ja Åbo akademin yhteinen erillislaitos. BTK tarjoaa korkean tason tekniikkaa ja osaamista akateemisille ja teollisen alan tutkijoille.

Työn tarkoituksena oli dokumentoida biotekniikan keskuksen käyttämät tekniikat röntgenkristallografiassa ja kirjoittaa niiden pohjalta protokollat henkilökunnalle.

ASIASANAT:

Röntgenkristallografia, diffraktio, proteiini, rakenne, kide

BACHELOR´S THESIS | ABSTRACT
TURKU UNIVERSITY OF APPLIED SCIENCES

Biotechology and Food Technology | Biotechnology

2011| Total number of pages 68

Instructors Kari Haajanen, M.Sc, Tassos Papageorgiou, Ph.D

Jesse Mattsson

DOCUMENTATION OF PROTEIN CRYSTALLIZATION TECHNIQUES AND CHARACTERIZATION OF CRYSTALS

X-ray crystallography is a method for determining the detailed three-dimensional structure of proteins and protein complexes. It is based on the diffraction of X-rays from protein crystals.

The principle is to grow suitable protein crystals and shoot them with an intense X-ray beam. The resulting scattering of X-rays is recorded and the resulting data can be used in molecular modeling. The models can be used for example in structure-based drug design and site-directed mutagenesis.

This thesis project was conducted in the Turku Center for Biotechnology (CBT), which is a joint department of the University of Turku and Åbo Akademi University. It provides high-end technologies and expertise to academic and industrial researchers.

The purpose of this thesis was to document the methods used in X-ray crystallography in CBT and to produce easy-to-read protocols for the staff.

KEYWORDS:

X-ray crystallography, diffraction, protein, structure, crystal

TABLE OF CONTENTS

APPENDIX

FIGURES

TABLES

1 INTRODUCTION

To understand how proteins function at the molecular level, it is necessary to determine their three-dimensional structure.

X-ray crystallography is the most common method for determining the detailed three-dimensional structure of proteins and protein complexes. It has been used to determine the majority (about 87%) of macromolecular structures to date. The method is based on the diffraction of X-rays from a crystal consisting of an orderly, repeating pattern of many identical molecules.

The protocols in this thesis were made using single-crystal X-ray crystallography, which has three basic steps. The first step is to produce a high-quality, imperfection-free crystal of purified protein. Next the crystal needs to be placed in a beam of X-rays and the resulting intensities and directions of diffracted beams are measured while the crystal is rotated.

Lastly the resulting diffraction patterns are combined with computers to produce a model of the atom arrangement within the crystal. Most finished models are uploaded to the Protein Data Bank (PDB), which is freely available online (http://www.rcsb.org).

2 PROTEINS

Proteins are the most abundant biological macromolecules in nature and are related to almost all functions in living organisms. They are made of amino acids that are linked covalently into linear sequences. These covalent bonds, which are called peptide bonds, are formed between the carboxyl and amino groups of adjacent amino acids. Chains that have <100 amino acids are called peptides, while longer chains which may have thousands of amino acids are called polypeptides. Proteins are formed from one or several polypeptides. (1)

All proteins are formed from the same set of 20 different amino acids. These 20 amino acids allow for a wide range of different combinations. For example, a 10 amino acid sequence has over 10 trillion different combinations. (1)

Peptides have the ability to fold, which allows for many different spatial conformations. Proteins are structurally rigid with only one preferable (native) conformation. However, some proteins can change from one conformation to another to perform their specific tasks in the cell. (1)

2.1 Protein structure

As mentioned earlier, the function of a protein is linked to its conformation. The folding is determined by various non-covalent interactions such as Van Der Waals forces, hydrogen bonding, and ionic interactions. Since proteins are typically very large and complex molecules, their structure is usually divided into four levels: primary, secondary, tertiary, and quaternary. (1)

The primary structure is the amino acid sequence in the polypeptide chain which also shows the locations of possible disulfide bonds. The primary structure can be illustrated by writing the amino acid sequence using the standard three-letter abbreviations for each amino acid as seen in Figure 1. (1)

The secondary structure refers to recurring, stable structural patterns of amino acids in localized regions of the protein. The most common and stable secondary structures are the α-helix and the β-sheet. The α-helix is a right-handed clockwise spiral in which each peptide bond is in *trans*-conformation. The β-sheet consists of polypeptide chains that run parallel or antiparallel with adjacent chains of which all are in trans-conformation. (1)

The tertiary structure illustrates all aspects of the three-dimensional folding of a polypeptide i.e. it is the three-dimensional structure of a protein.

The quaternary structure is used when a protein is composed of two or more polypeptide chains, called subunits, which form a single protein complex. Small proteins do not have a quaternary structure and function as monomers. (1)

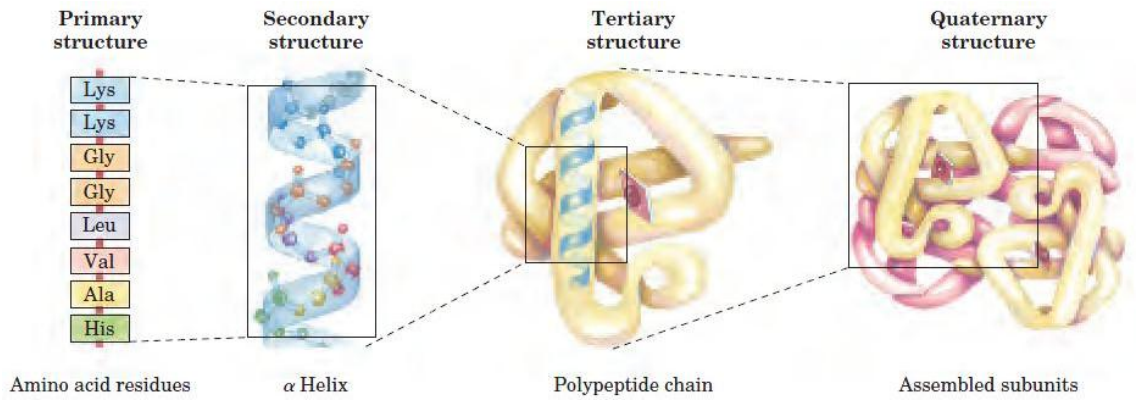

Figure 1. Levels of protein structure (1)

2.2 Protein functions

Proteins are responsible for several different functions in living cells. Differences in their functions result from the difference in their amino acid composition and sequences, which also determine the tertiary structure of the protein. (1)

The main reason proteins have such a large array of functions is that they have binding sites which allow them to bind specific molecules. This binding ability is caused by the tertiary structure of the protein. Typical protein functions include: (1,2)

- Enzymes
  These are proteins that catalyze chemical reactions i.e. increase the rate of chemical reactions both inside the living cells and outside them. This is the biggest and most important group of proteins. Enzymes are responsible for all metabolic reactions in the living cells. Reactions catalyzed by enzymes can be up to 10^{17} times faster. (1,2)

- Hormones
  These proteins are responsible for the regulation of many processes in organisms. They are chemical messengers that transport signals from one cell to another. Probably the most known hormone is insulin. (1,2)

- Transport proteins
  These proteins are responsible for transporting chemical compounds and ions. (1,2)

- Receptors
  Proteins responsible for signal detection and translation. (1,2)

- Antibodies
  Proteins used by the immune system to identify and neutralize foreign objects, such as bacteria, inside the body. (1,2)

- Structural proteins
  These proteins maintain the structures of biological components, e.g. cells and tissues. (1,2)

- Motor proteins
  Proteins that convert chemical energy into mechanical energy which is crucial for cellular motility in single-cell organisms. They also generate the forces needed for muscles. (1,2)

3 CRYSTALS

Even though single molecules scatter X-rays, the scattering in this case is very weak. Most of the X-rays pass through the molecule without diffraction and the scattered X-rays are too weak to be measured. For this reason, it is more advantageous to use crystallized forms of molecules that contain a high number of ordered molecules. Each molecule diffracts identically and the scattered X-rays from each molecule are added to produce detectable X-ray beams. (3)

Crystals present matter in its most highly ordered form. The high degree of order is what allows the imaging of molecules, since all the cell units that form the crystal are highly identical. Real crystals are mosaics of several arrays in regular alignment with each other based on symmetry rules. Figure 2 illustrates a molecule in a crystalline array. (3)

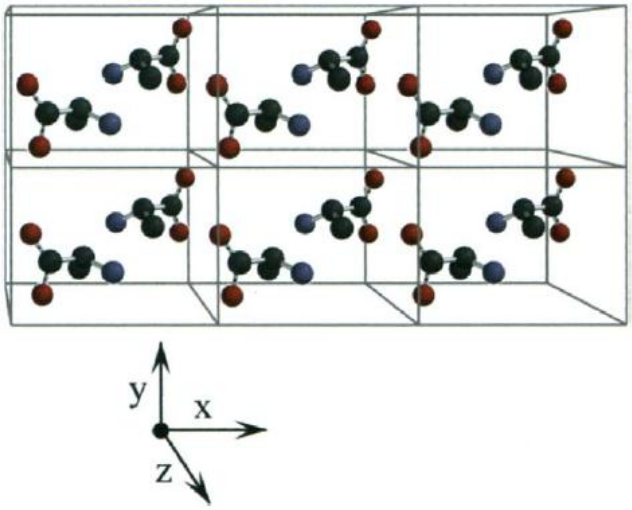

Figure 2. Unit cells in a crystalline lattice (3)

The lines in the figure divide the crystal into a three-dimensional lattice of repeating unit cells. The unit cell is the smallest repeating unit (a “shoebox”) that is repeated millions of times to create the whole crystal. Thus the crystal can be thought of as an orderly stack of unit cells. (3)

Most protein molecules are not altered by crystallization and they retain their function in crystalline state. For example, crystalline enzymes are still able to convert substrates to the product. In addition, models created from solution and crystal structures produce similar results. (3)

Crystals can be both organic and inorganic. Inorganic substances can be crystallized by preparing a saturated solution of the substance and then cooling it or by precipitation using organic solvents. Crystals produced in this manner can be several centimeters in their shortest dimension. Proteins however are denatured by organic solvents and heat and therefore they require gentler handling. (3)

3.1 Crystal size, integrity and form

Unlike with inorganic crystal, protein crystals cannot generally be grown to be larger than 1 millimeter in their shortest dimension. and in most cases they are between 0.1 and 0.3 mm. Compared to inorganic crystals, protein crystals are very fragile due to weak non-covalent bonds and higher water content inside

them. The lightest pressure or drying will destroy the protein crystals and thus they have to be grown and handled carefully. (3)

Proteins can crystallize in more than one form, even when they have almost identical molecular structures. The growth conditions have a large impact on this. Different forms may cause variation in diffraction quality in which case the form with the best diffraction should be chosen, because it determines the quality of the final crystallographic model. Figure 3 illustrates a few different crystal forms.

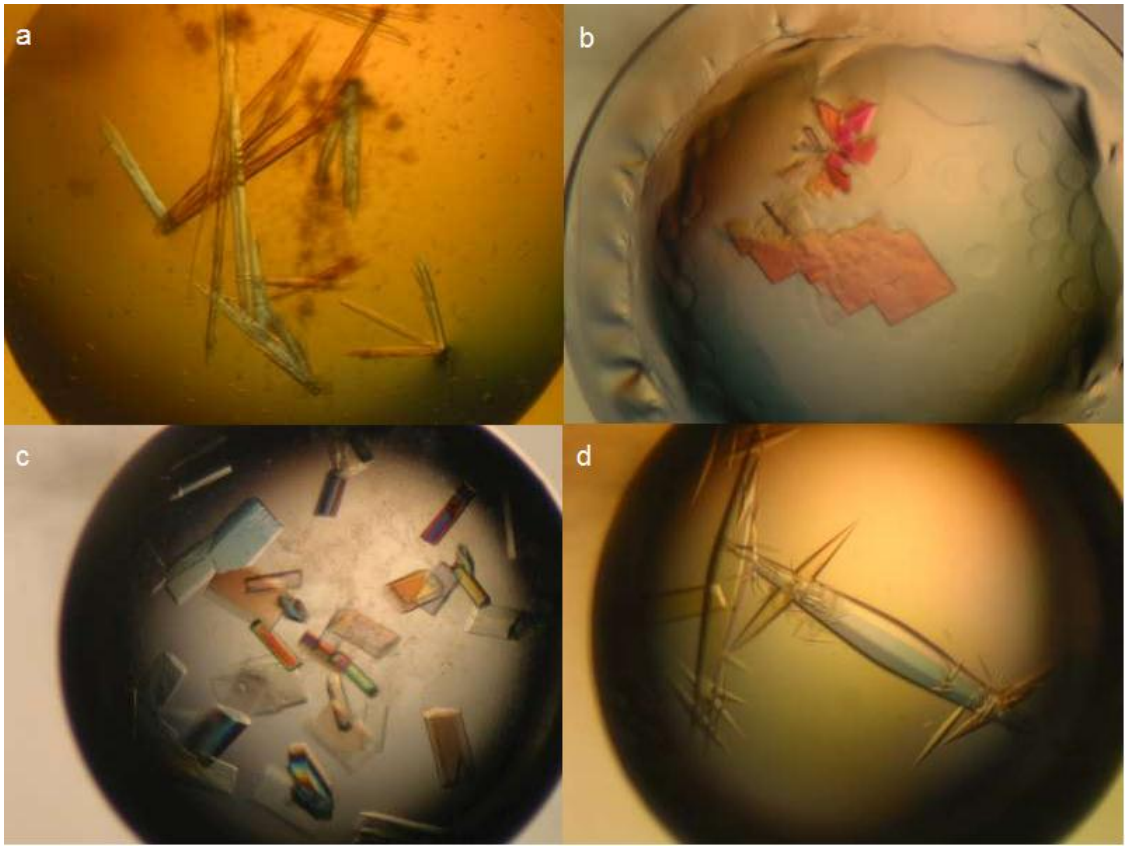

Figure 3. Different crystal types: a) needles b) plates c) boulders d) rods

3.2 Growing crystals

Production of good crystals is the most time consuming phase of X-ray crystallography since some crystals are not suitable for X-ray diffraction and getting a protein to produce crystals is not always easy. Since data collection is completely dependent on crystal quality, crystals need to be screened in a trial-

and-error process in large numbers of differing conditions until optimal conditions for crystal growth are found. (4)

Crystals are grown by controlled precipitation. The crystallization process begins by producing super-saturated solutions containing the protein to be crystallized. The super-saturation causes the molecules in the solution to form small aggregates that act as nuclei for subsequent crystal growth. Crystals are often of better quality when the formation process is slow. For this reason crystals are often grown below ambient temperatures (i.e. room temperature)**.** Lower temperatures also have the effect of preserving proteins from degradation. (4)

Crystals also need to be of certain size for good diffraction, so the degree of super-saturation can be adjusted just below the one required for nucleation. This will inhibit the formation of many nuclei, which in turn inhibits the formation of many very small crystals, which could not be used with X-rays**.** (4) In case only small crystals are grown, the ones with good quality can be used as seeds to grow larger crystals.

In practice, a drop containing the protein to be crystallized is added to another drop containing precipitant (crystallizing agent and additives). The resultant drop is equilibrated against a reservoir, which contains a much larger volume of precipitant, in a sealed, airtight area. This causes water to move from the protein drop to the reservoir by vapor diffusion until the system reaches equilibrium, where the transfer of water stops and the protein solution is kept in the optimal precipitant concentration. The precipitant concentration does not change because the vapor pressure in the drop and reservoir are equal**.** (4,3)

The protocols in this thesis were made using the sitting and hanging drop methods, both of which are based on the vapor diffusion method detailed above. Figure 4 illustrates both methods. The sitting drop method is being used for finding the optimal crystal growth conditions by creating 192 preset condition(s) and checking which condition produces the best crystals. The optimal condition was then used with larger volumes of protein solution and

precipitant with the hanging drop method. In seeding the procedure is the same except that small crystals are added to each drop (preferably after they have been mechanically crushed).

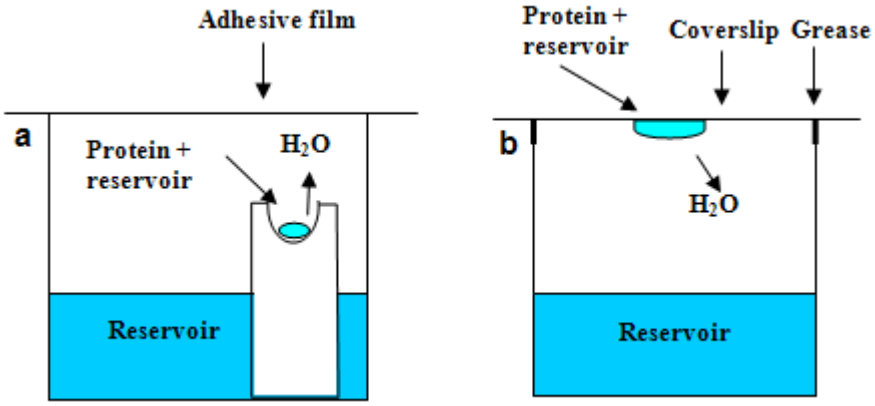

Figure 4. a) Sitting drop method b) hanging drop method

Crystal formation is influenced by several factors. Crystallization trials vary in the used precipitant and its concentration, buffer pH's, temperatures and possible additives, like anions or cations. Figure 5 details how only two variables, in this case the concentrations of the protein and precipitant, affect the crystal formation.

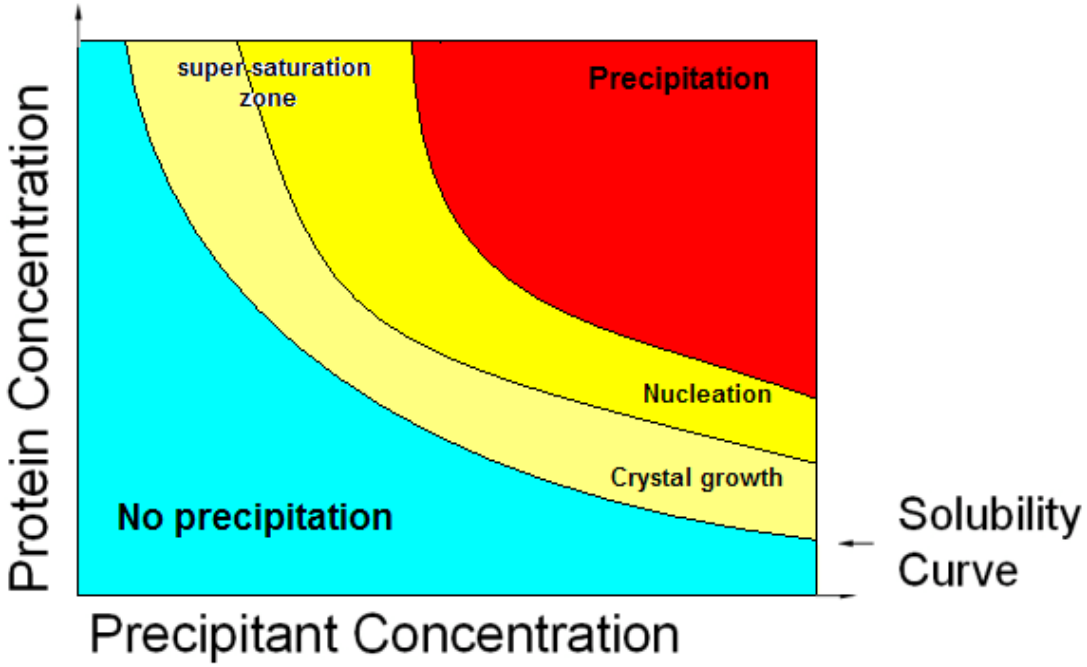

Figure 5. Protein crystallization phase diagram

3.3 Mounting crystals

Crystal mounting is done either by scooping the crystal out of the drop with a specialized loop or transferring them to a capillary tube. First the crystals are examined with a microscope and good crystals are selected, preferably the ones with well defined edges and no cracks. Then the crystal is either transferred to the capillary or scooped into the loop in the presence of a cryoprotectant. Lastly the crystal is mounted to the goniometer for X-ray shooting. The cryoprotectant is added to prevent ice formation on the protein when it is put into a liquid nitrogen stream, which protects the protein from radiation damage.

4 DATA COLLECTION

When we see an object, rays of visible light hit the object and are diffracted by it. The diffracted rays then enter the eye through the lens that reconstructs an image of the object and focuses it onto the retina. Optical microscopes work largely the same way except that the lens is spherical which causes the image of the object to focus at a short distance within the microscope's tube. The image is then magnified by a second lens and is visible through the eyepiece. (3)

For an object to be visible, it has to diffract light. Therefore the wavelength of the light must be smaller or equal than the object. Visible light has the wavelength of 400-700 nanometers. However bonded carbon atoms are approximately only 0.15 nm apart from each other so visible light cannot produce an image of such small objects. X-rays are in the corresponding radiation wavelength range so they can be used for diffraction. (3). Figure 6 illustrates different kinds of electromagnetic radiation and their wavelengths.

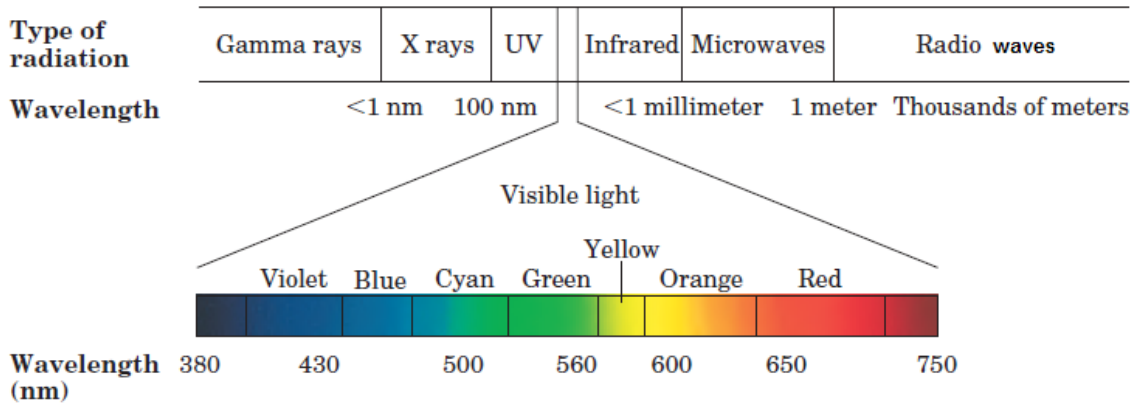

Figure 6. Electromagnetic radiation spectrum (1)

However, it is not possible to produce focused images of molecules with X-rays, since X-rays have high energies no lenses can be constructed. We therefore measure the intensities and directions of scattered rays in order to produce a computer-generated image of the original object. (3)

The distances between atoms are so small that they are usually expressed in ångströms (Å). An Ångström is a unit of length, which equals to 0.1 nm.

Figure 7 illustrates the basics of collecting X-ray diffraction data. An X-ray beam is shot through the crystal placed between the radiation source and an X-ray detector. The detector then “records” the resulting reflections coming from the crystal.

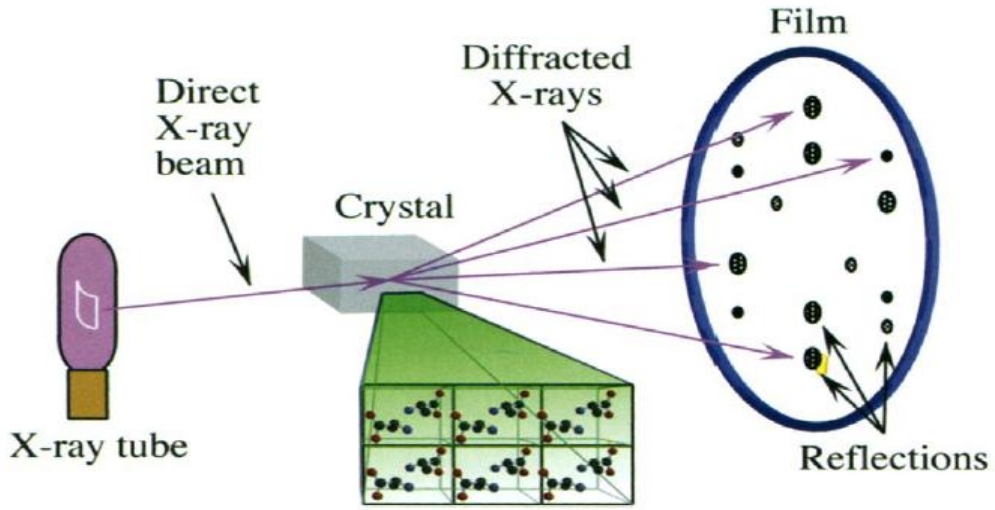

Figure 7. Data collection (3)

4.1 Radiation damage

X-rays are usually described in terms of the energy they carry. The energies range from less than one thousand electron volts (keV) to more than 100 keV. X-rays in the 10-0.10 nm (0.12 to 12 keV) range are called “soft” X-rays and the 0.10-0.01 nm (12 to 120 keV) range is called “hard” X-rays due to their penetrative abilities. (5)

Radiation damage to the crystal is caused by the absorption of photons from the X-ray beam by the crystal. When an atom absorbs X-ray photons, the absorbed energy ionizes the atom. Each photoelectron has enough energy to produce several ionization events that can result in the formation of free radicals. Since free radicals are atoms or ions with unpaired electrons on their open shell configuration, they are highly reactive. Intense X-rays also generate heat. (6)

Even after the radiation exposure has stopped, if radiation damage has begun, it will continue at a steady rate due to the free radicals. This cannot be avoided due to the ionizing nature of X-ray radiation. Therefore, it is beneficial to do the measurements in a prompt manner and without interruptions. (6)

Radiation damage can be reduced by cooling the crystals in liquid nitrogen, which has the temperature of -173 °C or 100 K. This causes no damage to the crystals and may even enhance the crystal order. The crystals need to be cooled very quickly, in the presence of cryoprotectant like glycerol or glucose, to avoid the formation of ice crystals, to below 190 K. In this temperature, water is in a glassy, stable state. (3, 6)

However, radiation damage can still occur despite cryocooling. Radiation can cause loss of measured reflection intensities and may cause artefacts to appear in solved structures. (6). Artefacts are systematic errors leading to incorrect observations. For example an artefact might be bond lengths appearing too short due to libration (libration is movement where an object repeatedly rotates back and forth) or inaccurate determination of hydrogen atom positions. Despite this, cryocooling is still a very effective method to reduce radiation damage. (7)

As said earlier in the crystal mounting part, crystals can be prepared for flash cooling by dipping them in a cryoprotected mother liquor for a few seconds. After this they are dipped in liquid nitrogen and if the freezing is successful, the crystal is placed onto the goniometer, which holds it in a steady stream of cold nitrogen gas for data collection. Reduction in radiation damage increases the data one can collect from a single crystal and thus reduces the amount of crystals needed for comprehensive results

X-rays can damage surface tissues and especially eyes so care should be taken when working with them.

4.2 X-ray generators

X-rays can be generated by directing a beam of electrons to a metal target, usually copper or molybdenum, in a vacuum with electrons produced by a heated wire and accelerated by an electric field. The resulting high-energy electrons then collide with the metal target and displace electrons from low energy orbitals. This causes another electron from a higher orbital to drop down into the resulting vacancy, which then emits its excess energy as an X-ray photon. (3,8)

The resulting radiation should be single-wavelength, because radiation with two wavelengths gives two sets of reflections which overlap. (3)

The most common X-ray generators are X-ray sealed tubes, rotating anode tubes and particle storage rings (synchrotrons). X-ray tubes (Figure 8a) operate by shooting electrons to a water cooled target metal, which is called the anode. The output is limited by the rate at which heat is dissipated from the anode by the circulating water. (3,8)

The rotating anode (Figure 8b) is similar to the X-ray tube, but the target metal is a rotating metal disk. The rotation allows for higher output because the heat build up is distributed through a larger volume of metal. The radiation from rotating anodes is several times stronger than that of fixed anodes. (8). The protocols in this thesis were made with a rotating anode source.

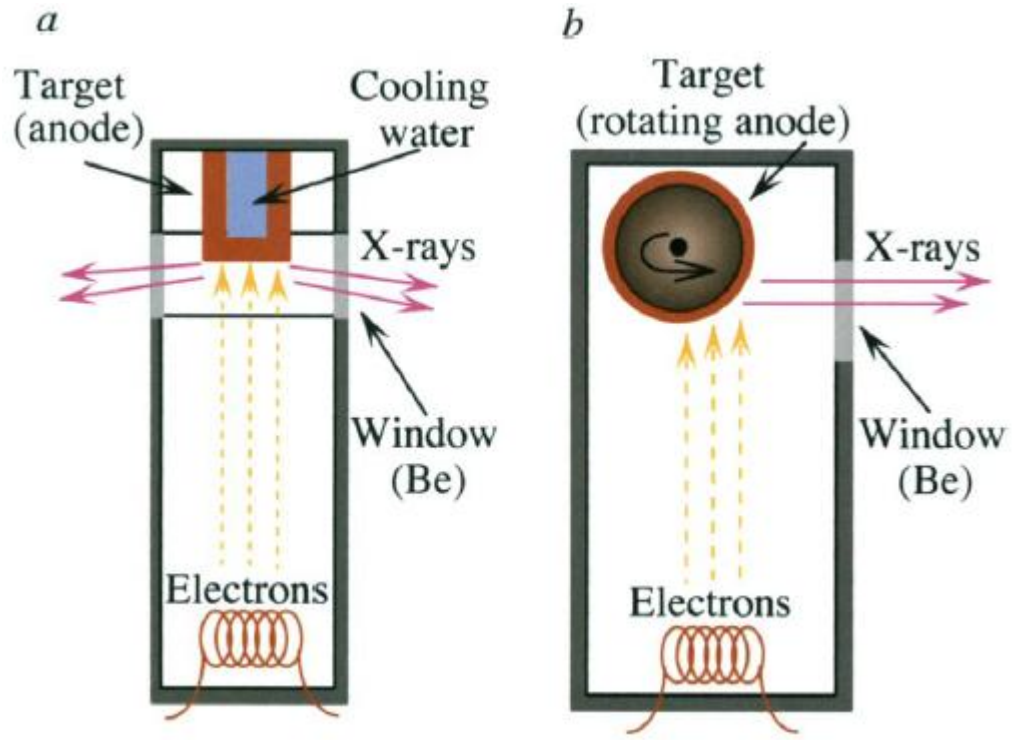

Figure 8. a) X-ray tube b) rotating anode tube (3)

Synchrotrons are cyclic particle accelerators, in which the magnetic and electric fields are carefully synchronized with the used particle beam. They are by far more powerful than any other X-ray source. In rotating anodes the energy is measured in thousands of electron volts, keV, whereas in synchrotrons it is in the range of billions of electron volts, GeV. The electrons travel on a circular tracks at nearly the speed of light. (3,8)

The main advantage of synchrotrons is that data that requires several hours of exposure in rotating anodes can be obtained in a matter of minutes in synchrotrons. The synchrotrons can also shoot X-rays in selectable wavelengths, which can be helpful later on in structure determination. (4)

4.3 X-ray Detectors

When a crystal is mounted and exposed to a beam of X-rays, the resulting reflections can be observed on a screen behind the crystal. X-ray detectors respond to the intensity of diffracted beams which is proportional to the number of photons it contains. (8)

The simplest X-ray detector is the photographic film, but it has been replaced by image plates and charge-coupled devices (CCD). Image plates have the capacity to record reflections with widely varying intensities. They store diffraction images to phosphor-coated plastic sheets in which atoms are promoted to an excited state when they absorb X-ray photons. Next the plate is scanned with light and the intensity of fluorescence is measured. The measured intensity is proportional to the intensity of the absorbed X-rays. Afterwards the plate is flooded with light to clear the plate. The image plate can be reused indefinitely. (3,8)

One disadvantage of image plates is that the exposure is interrupted frequently due to the data scanning and clearing, though this can be reduced by using more than one image plate. CCDs are arrays of μm-scale pixels of which each accumulates charge in direct proportion to the amount of light that strikes them. At the end of data collection cycle, the pixel charges are read out one line at a time by transferring rows of pixels sequentially into a serial readout row at one edge of the CCD. The charges in the readout row are then serially transferred to an amplifier at the end of the row. Since all the data is read at the end, CCDs have no pauses similar to image plates. (3,8)

CCDs are better suited for data collection of well diffracting crystals and crystal screening since they are noticeably faster than image plates. Image plates on the other hand allow data collection for weakly diffracting crystals via long exposures and can be used the same as CCDs for everything else. (3). Both methods are viable in small laboratory setups.

4.4 Diffraction

Diffraction refers to the phenomena which occur when a wave encounters an obstacle. Figure 9 gives some help in seeing what information can be gained from diffractions.

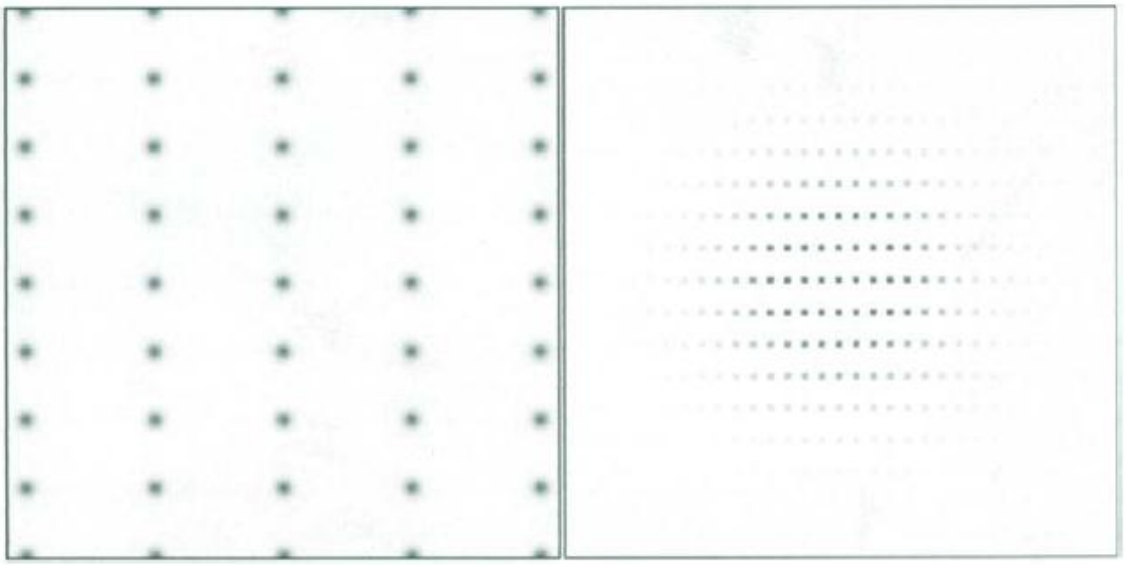

Figure 9. Lattice of spheres and its diffraction pattern (3)

Diffraction patterns consist of reflections in an orderly array recorded on detectors. “The spacing of reflections varies with the spacing of the spheres in their array. Observe, that although the lattice spacing of the crystal is smaller vertically, the diffraction spacing is smaller horizontally. There is a simple inverse relationship between the spacing of unit cells in the crystalline lattice, called the real lattice, and the spacing of reflections in the lattice on the film, which, because of its inverse relationship to the real lattice, is called the reciprocal lattice”. (Rhodes, 2000, p.16)

Because the lattice spacings are inversely proportional, it is possible to calculate the dimensions of the unit cell in the crystal from the spacings of the reciprocal lattice. Also, note the intensities of the reflections in Figure 9. The pattern of varied intensities is that of the average sphere in real lattice, because all the spheres contribute to the pattern. The intensity pattern is a superposition of the many identical diffraction patterns of all the spheres in real lattice. (3)

Crystals however are not two-dimensional, like the pattern in Figure 7. When crystals are rotated in an X-ray beam, the resulting diffraction pattern is different for each angle of rotation. Basically each two-dimensional diffraction pattern from a crystal is a cross section cut of the three-dimensional lattice, meaning that only one plane from the three-dimensional lattice is seen. Figure 10

illustrates a diffraction pattern while showing all the possible reflections at other crystal orientations. (3)

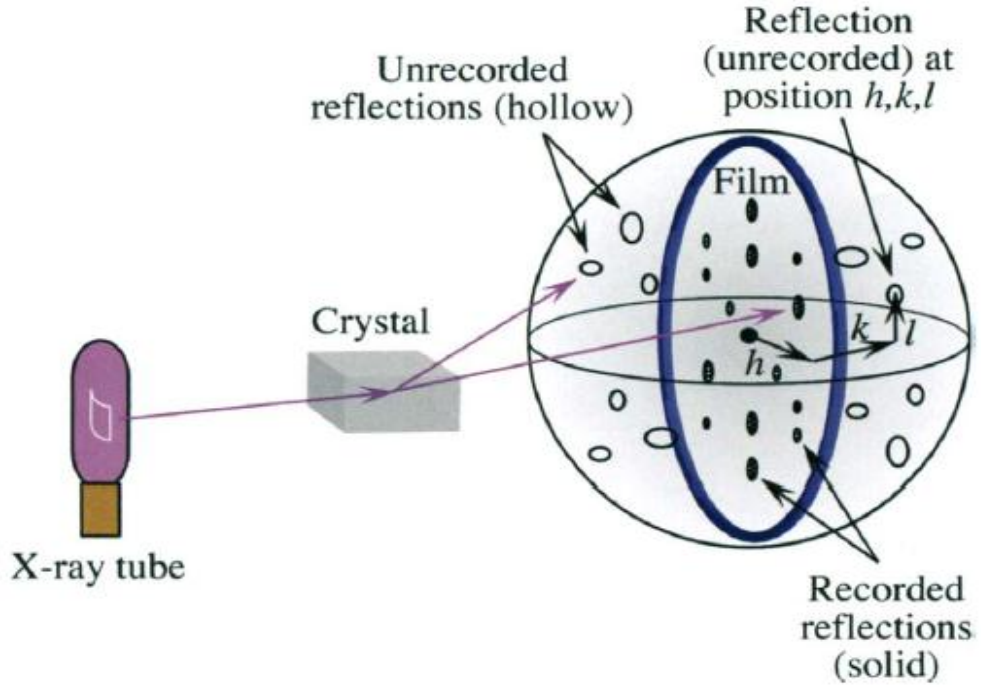

Figure 10. Data collection showing all the possible reflections (3)

After data collection, the collected data is basically a list of intensities for each point in the 3-D reciprocal lattice and this information is the basis for protein structure determination. (3). Figure 11 shows an example of a diffraction pattern from a crystal.

From the diffraction image we can ascertain that the diffraction is of acceptable resolution for structure determination. Several image display programs can determine the resolution of particular spots. Generally the resolution has to be 3 Å or smaller, but 3 Å is accurate enough to detect amino acid side chains in the electron density map. Higher resolution (smaller number) provides more details in the structures and is always preferable in structure determinations (9)

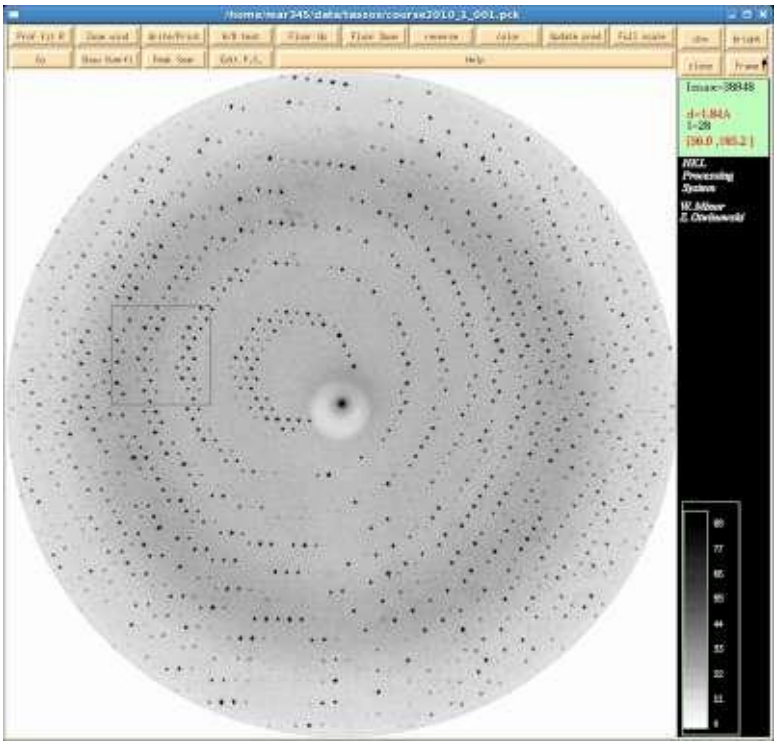

Figure 11. Diffraction pattern from a protein crystal

Notice the hole in the middle of the diffraction pattern. It is caused by a beam stop which is a small piece of metal. The beam stop is located just behind the crystal in data collection to block excessive radiation from the direct, non-scattered beam to protect the detector from overload and damage.

4.5 Generation of protein structures from the diffraction data

The construction of a protein model begins by combining all the recorded reflections using a mathematical operation called reverse Fourier transform. The recorded reflections are produced by X-rays so the reflections are treated as waves. The transformation is made by using the values of the total scattered wave, which is a combination of the amplitude and phase of the scattered X-rays. The amplitude is proportional to the measured intensities and can thus be derived from the measured intensities. However, the phase of each reflection is lost in the measurements, because X-ray detectors measure only the intensity. This loss of phase in measurements is called the “phase problem”. Each

reflection has a phase so the phase problem has to be solved for each one. (10)

There are several ways to determine the phase. One method is isomorphous replacement known also as the heavy-atom method, where a protein crystal is soaked in a heavy-atom solution. This will cause one or more heavy atoms to bind into the protein molecule. The purpose is to create a derivative protein crystal without changing the protein conformation or unit cell dimensions so that the resulting crystal is isomorphous with the native crystal. (9)

When the diffractions of the two crystals are compared to each other, any differences between them will be caused by the heavy atoms and thus the positions of the heavy atoms can be deduced. After the locations of the heavy atoms are known, the phases can be determined. (9)

Another common method is molecular replacement, where the phases from similar crystals, whose structure is known, are used. First the known model structure has to be placed into the unit cell in the same orientation and position as the new molecule. Once the model is placed in, the phases from it may be used with the new amplitudes to produce the new structure. (9)

Next, using the estimated phases and measured amplitudes, the electron density map can be calculated after which the protein structure can be constructed. The structure is then refined by calculating new amplitudes from the structure which are compared to measured values. Refinement is continued until a good fit is attained. (10)

This is a drastically simplified depiction of the phasing process. In reality, phasing is much more complex, but complete depiction of the process is beyond the scope of this thesis.

4.6 Other crystallography methods

The same principles behind X-ray crystallography apply also to methods using different forms of radiation. Neutron and electron crystallography are similar to X-ray crystallography and can be used to give complementary information.

Like X-rays, which is electromagnetic radiation, also subatomic particles like electrons and neutron have a wavelength due to their motion and their diffraction can thus be described with a Fourier transform. Where X-rays are reflected by the electrons with surround atoms, neutron are diffracted by atomic nuclei instead and electrons are influenced by both the positively charged atomic nuclei and the negatively charged surrounding electrons. (3)

Instead of an electron-density map, neutron crystallography produces a map of nuclear mass distribution. As with X-rays, neutron diffraction loses the phase information during measurements. Neutrons can be produced by nuclear fission in a reactor. The main advantage of neutron diffraction over X-rays is that neutrons can determine the positions and thermal movement of hydrogen atoms accurately. Neutrons can also be used to identify hydrogen isotopes e.g. deuterium. The main drawback is that the neutron method requires either very large crystals or long exposure times for small crystals due to low flux of currently available neutron beams. Unlike X-rays, neutrons are non-destructive and do not destroy crystals in exposure. (11)

Electrons are produced by transmission electron microscopes. Unlike neutron or X-rays, it is possible to focus electrons with electron lenses, and so the phase information can be determined experimentally with electron crystallography. One advantage of electron diffraction is the ability to collect diffraction from nanometer scale crystals and the method has high resolution but it suffers from weak diffraction due to the thin samples used. Electron microscopes can also be used to collect electron diffraction data from two-dimensional arrays of molecules so it is useful for some membrane proteins. The radiation damage caused by electron beams is such that the effective resolution is often lower than the method would allow. (12)

Nuclear magnetic resonance crystallography can also be used in protein modeling for proteins in solution. It utilizes NMR spectroscopy to determine the structure of objects on the atomic scale. It is mainly used for amorphous and small materials which are not compatible with X-ray, neutron or electron diffraction. (3)

5 PROTEIN DATA BANK (PDB)

The protein data bank is a public repository for three dimensional structures of large biological molecules which include proteins and nucleid acids. The data, which are submitted by structural biologists/protein crystallographers around the world, are freely available to everyone and can be accessed with an Internet browser. The PDB is overseen by the Worldwide Protein Data Bank organization. (13)

Today it is the major resource in areas of structural biology, such as structural genomics. Many major scientific journals and some funding agencies, now require scientists to submit their structure data to the PDB. (13)

PDB database is updated weekly on Wednesdays. See Table 1 for holdings breakdown:

Table 1. PDB holdings breakdown as of 12.05.2011 (13)

Experimental Method	Proteins	Nucleic Acids	Protein/NA Complexes	Other	Total
X-RAY	59425	1276	2865	18	63584
NMR	7749	944	171	7	8871
ELECTRON MICROSCOPY	250	22	94	0	366
HYBRID	29	3	1	1	34
other	132	4	5	13	154
Total	67585	2249	3136	39	**73009**

52,991 structures in the PDB have a structure factor file.
6,169 structures have an NMR restraint file.
45 structures in the PDB have a chemical shifts file.

The files in PDB come in three formats; PDB file format, mmCIF (macromolecular Crystallographic Information file) format and the XML version of the mmCIF format called PDBML. The structure files can be downloaded in any of these three formats.

Below is a short description of PDB file contents. Some line types do not apply to all models and may be missing. (14,3)

The contents of the file, in order of appearance, are

- HEADER lines, containing the file name and date.
- COMPND lines, containing the name of the protein.
- SOURCE lines, giving the organism from which the protein was obtained.
- AUTHOR lines, listing the persons who placed this data in the Protein Data Bank.
- REVDAT lines, listing all revision dates for data on this protein.
- JRNL lines, giving the journal reference to the lead article about this model.
- REMARK lines, which contain 1) references to journal articles about the structure of this protein, and 2) general information about the contents of this file.
- DBREF lines, providing cross-reference links between PDB sequences (what appears in SEQRES record) and a corresponding database sequence.
- SPRSDE lines, which list older coordinate files of this same structure.
- SEQRES lines, which give the amino-acid sequence of the protein.
- FTNOTE lines, which contain footnotes

- HET and FORMUL lines, which list the cofactors, prosthetic groups, inhibitors or other nonprotein substances present in the structure.
- HELIX, SHEET, TURN, CISPEP and SITE lines, which list the elements of secondary structure in the protein, residues involved in *cis*-peptide bonds and residues in the active site of the protein
- CRYST1, ORIG, and SCALE lines, which contain some general information about the protein crystals from which this structure was obtained X-ray crystallography.
- ATOM and HETATM lines, which contain the atomic coordinate data needed to display the structure of the protein. Notice that no hydrogen atoms are listed in the file. Most protein crystals do not diffract well enough to allow hydrogen atoms to be resolved. Positions of hydrogen atoms must be inferred from the positions of other atoms. HETATM contain the same information as ATOM lines for nonprotein molecules.
- CONECT lines, which list bonds between nonprotein atoms in the file.
- MASTER and END lines, which mark the end of the file.

The structure files can be viewed with several different programs. The PDB website has a large list of different programs at http://www.rcsb.org/pdb/static.do?p=software/software_links/molecular_graphics.html.

6 REFERENCES

(1) D.Nelson and M.Cox. (2004) Lehninger Principles of Biochemistry, Fourth Edition. W. H. Freeman, NY, USA

(2) http://proteincrystallography.org/protein/

(3) Gale Rhodes. (2000) Crystallography made crystal clear: A guide for users of macromolecular models, second edition. Academic press, USA

(4) Dulce H. F. Souza, Heloisa S. Selistre-de-Araujo, Richard C. Garratt. (2000) Determination of the three-dimensional structure of toxins by protein crystallography. *Toxicon*, *Volume 38, Issue 10, Pages 1307-1353*

(5) NASA's solar flare theory educational webpage – what are hard X-rays? http://hesperia.gsfc.nasa.gov/sftheory/xray.htm

(6) Elspeth F. Garman. (2010) Radiation damage in macromolecular crystallography: what is it and why should we care? Acta Crystallographica D; Biological Crystallography, volume 64, p. 339-351.

(7) Peter Müller. (2006) Crystal Structure Refinement: A Crystallographer's Guide to SHELXL. Oxford University Press, NY, USA

(8) David Blow. (2002) Outline of crystallography for biologists. Oxford university press, NY, USA

(9) M. Smyth and J. Martin. (2000) X-Ray crystallography. Journal of clinical pathology: Molecular pathology, volume 53, issue 1, February 2000, p. 8-14.

(10) Teeri Tuula. (1988) Protein structure determination. TEKES, Hki, FI

(11) M.Blakeley, P.Langan, N. Niimura, A. Podjarny. (2008) Neutron crystallography; opportunities, challenges, and limitations. Current opinion in structural Biology. Volume 18, issue 5, October 2008, p. 593-600.

(12) Martin Egli. (2010) Diffraction Techniques in Structural Biology; Overview for unit 7 "Biophysical Analysis of Nucleic Acids". Current Protocols in Nucleic Acid Chemistry, June 2010, chapter: Unit 7.13.

(13) Protein data bank homepage
http://pdbbeta.rcsb.org/pdb/static.do?p=general_information/about_pdb/index.html

(14) Protein data bank current holdings webpage
http://www.wwpdb.org/documentation/format32/v3.2.html

The protocols in this thesis are confidential and therefore are not included in the public version