LABORATORY OF BIOPHYSICS FOR ADVANCED

Experimental exercises for III year of the First cycle studies

Field: “Applications of physics in biology and medicine”

Specialization: “Molecular Biophysics”

X-RAY CRYSTALLOGRAPHY

SMALL MOLECULES STRUCTURE DETERMINATION

(ex. 37)

HUMAN CAPITAL
NATIONAL COHESION STRATEGY

EUROPEAN UNION
EUROPEAN
SOCIAL FUND

Project co-financed by European Union within European Social Fund

Content

1. Principles of X-ray crystallography

Study of the objects with electromagnetic radiation, requires radiation to have a wavelengths similar to dimensions of the object. Atoms dimensions and bond lengths are typically in the range of 1 to 3,5Å. 1Å is dimension of hydrogen atom, 2,4 – 3,5Å are lengths for hydrogen bonds, 1,3 - 1,6Å is a typical length range of C-C bonds in proteins [1]. To research objects of that size X-rays radiation has to be used.

But, the result if crystallographic experiment, is not picture of atoms, it is electron density map. This is because, the electromagnetic radiation interact with matter through its electromagnetic field. The intensity of scattered radiation is proportional to charge to mass ratio. As the electrons are few thousands times lighter than protons and atomic nuclei they interact much stronger with radiation. Moreover the velocity of electrons in the atoms is many times higher than the velocity of changes of electromagnetic field od the X-ray, that what is observed are not single electrons, but time-averaged distribution of electron density in molecule. However, as electrons are tightly localised around the nuclei and bonds, the electron density map, gives good picture of molecule itself.

Diffraction on single molecule is extremely week, thus difficult to detect, and measure above the noise level (the scattering of water and air molecules). Crystal contains large number of molecules ordered in space, so scattered radiation is coherent in phase. Thus the constructive interference occurs, the waves amplitudes can add up and reflection intensity rise to the measurable level – crystal acts as an amplifier.

Of course, if the waves adds up in some direction, due to interference, it has to cancel out in many other directions, thus the diffraction pattern of crystal is not continuous, in fact this is an array of spots.

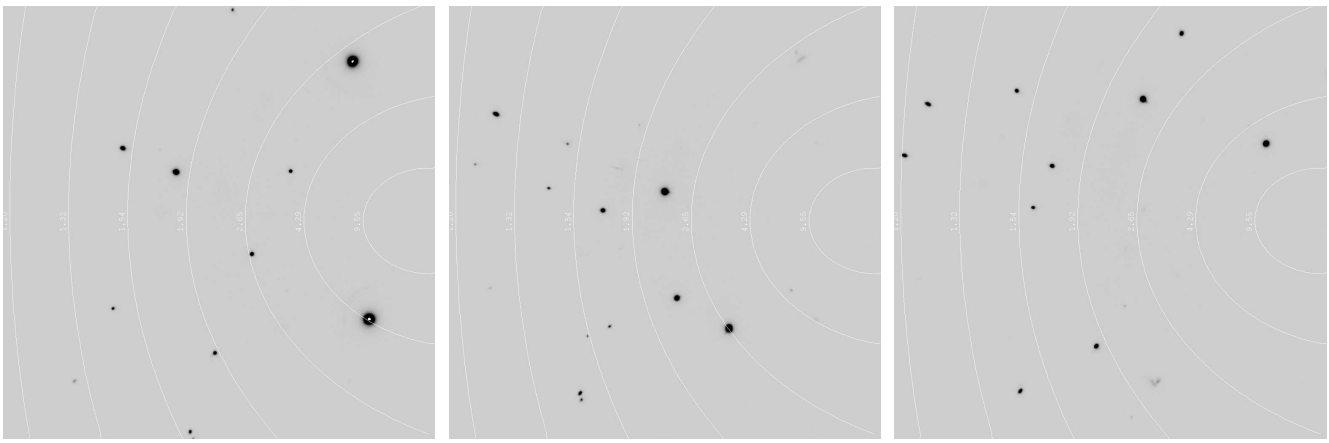


Fig. 1: Diffraction image of sucrose crystal

1.1. Diffraction on crystals

When a wave scatter on the electrons in the crystal scattered waves interfere with each other. Dependently on the relative distances of the electrons and the angle of incidence the waves can add up, cancel up, or something in between. Which of this will happen, depends of the total distance between source and detector. If the pathlenght of diffracted waves differ by multiple of wavelength, the waves will be in phase, and they will add up, and the as the result of interference the wave will amplify; if the pathlenght differ by multiple of wavelength plus half wavelength, they will cancel up. The condition for constructive interference, can be obtain quite easily, think of diffracted waves, like if they were diffracted from the plane passing through atoms, this plane is called Bragg's plane, and it behaves like mirror reflecting radiation.

When a parallel beam is reflected from the mirror, the incidence angle equals reflection angle, the same happens for Bragg's planes. So if the incident beam is in phase, the reflected beam is also in phase, independently where they hit the plane. Scheme given at Fig. 2 explains why.

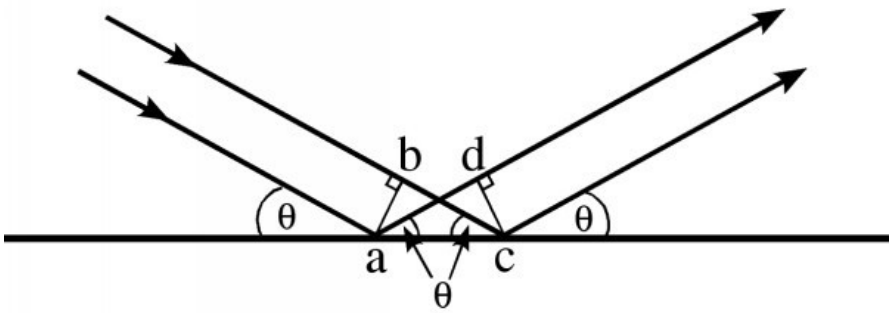


Fig. 2: Diffraction on the plane[2]

The incident beam is in phase, that means that in points a and b incident waves are in phase. Distance bc is equal to distance ad, because triangles abc and acd are congruent. Thus at points c and d waves have the same phase, and the reflected beam is in phase.

If waves scattered on on plane have to have the same pathlenght to be in phase, waves scattered on different planes have to have different pathlenght, and the difference in pathlenghts has to be equal to multiple wavelength. Bragg's law gives the relationship between spacing of Bragg's planes to give constructive interference.

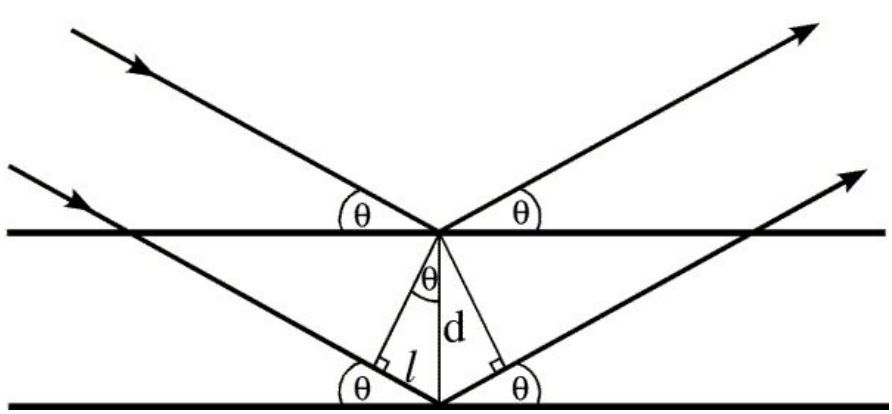


Fig. 3: Diffraction on the Bragg's planes[2]

Difference of pathlenght of waves diffracted on different planes is equal to $2l$ (fig. 3) and it depends of the incident angle , $l = d\sin\theta$, thus tge Bragg's law is given as:

$$n\lambda = 2d\sin\theta \qquad (1)$$

According to Bragg's law, the higher the angle of diffraction, the smaller distance between Bragg's planes has to be, to keep the pathway difference equal wavelength. That means, that the higher angle, the smallest details can be “seen” in diffraction experiment.

This inverse proportionality causes that diffraction data is usually analysed in reciprocal space. The bigger distance between objects in reciprocal space the closest they are in real space, the diffraction is reflected at higher angles, and more sensitive to smaller details.

As mentioned above, the crystal act as an amplifier thanks to interference of scattered radiation. For constructive interference to happen, scattering on all unit cells has to be in phase, thus the Bragg's planes has to go through the same points in every unit cell of the crystal.

If the objects on Bragg's plane scatter in phase, objects placed of planes scatter out of phase, and the phase shift is proportional to the distance from the Bragg plane. Thus a single diffraction experiment allows to calculate relative distances of objects from Bragg's planes. If all the objects are on the Bragg's planes, the diffraction image is one dot. If half of the

objects is on the Bragg's plane, and the other half on the parallel planes exactly in half of the distance between Bragg's planes, this to sets of object will scatter out of phase, and phase shift will be 180°, so there will be destructive interference. The diffraction image will not change. If the other set of objects will be in any other distance that $d/2$ the phase shift will be somewhere between zero and 180°, th e diffracted wave will add contribution to interference. On the diffracion image the intensity if the spot will change.

The relationship between diffraction image and the object on which the radiation scattered is given by Fourier transform. Assuming that the electron density is a mathematical function, the diffraction image is a Fourier transform of electron density. Most of the mathematical function have their inverse function, lib sinus and arc sinus, the Fourier transform is not an exception. That means, that electron density is an inverse Fourier transform. So, just calculate?[1]

Unfortunately its not that simple. Tu compute electron density amplitudes and phases of diffracted wave has to be known. But during experiment, the number of photons scattering on the detector are measured. This number gives intensity of scattered wave, and intensity is proportional to amplitude of the wavelength, but the is no experimental method allowing to measure phases of scattered waves. This is so called The Phase Problem[2]

1.2. The Phase Problem

Understanding, why phases “disappear” during diffraction experiment, the wave theory if light us not enough, the quantum-mechanic theory is necessary. In quantum mechanic, the probability that a photon will be reflected in certain direction is given by a square of amplitude of the scattered wave.

$\Psi(x,t)=A\cdot e^{2\pi i(\nu t-x/\lambda)}$ - wave function of photon

$P_{a<x<b}=\int_a^b \Psi(x,t)\cdot\Psi^*(x,t)dx=\int_a^b |\Psi(x,t)|^2 dx$ - probability that photon happened to be in the area of $x\in(a,b)$

$|\Psi(x,t)|^2=A\cdot e^{2\pi i(\nu t-x/\lambda)}\cdot A\cdot e^{-2\pi i(\nu t-x/\lambda)}=A^2$

That's why, there is only intensity of reflected waves on the diffraction image.

As the phases can not be measured directly, they has to be obtained from indirect measurements.

1.3. Solutions of The Phase Problem

There are three methods for solving The Phase Problem: guessing the phases, interpretation of Patterson map, direct methods. These methods allow to obtain only approximate solution.

1.3.1. “Guessing” phases

Solving The Phase Problem by guessing the phases is historically first method. For relatively small molecules the initial phases can be estimated basing on same prior knowledge of crystal symmetry and molecule features. The crystallographer need to know size and type od unit cell, how many molecules fill unit cell, what symmetries unrelated to crystal structure molecule can have. This knowledge combine with general knowledge of bonds lengths and angles can provide initial estimations of phases. Using this guessed phases a diffraction pattern is computed, and compared with experimental data. Then in recursive process phases are adjusted to recreate experimental data.

1 Kevin Cowtan's page: *The Interactive Structure Factor Tutorial* provides a nice look on how inverse Fourier transform works http://www.ysbl.york.ac.uk/~cowtan/sfapplet/sfintro.html

2 Why phases are so important? Look at another page of Kevin Cowtan: *Kevin Cowtan's Book of Fourier* http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html

1.3.2. Patterson's Function

Patterson's function is an inverse Fourier transform of intensities of measured peaks.

The Patterson's function, can be visualised as a Patterson's map. This map, is somehow similar to electron density map, but its maxima are not at the positions of atoms, but in positions corresponding to relative distances between atoms. If, in the unit cell, there is an atom at position $\vec{x}_1$ and another atom in position $\vec{x}_2$, on the Patterson's map there will be two peaks in positions $\vec{x}_1 - \vec{x}_2$ and $\vec{x}_2 - \vec{x}_1$. The heights of peaks on the Patterson's map is proportional to product of peaks on the electron density map. If there is N atoms in the unit cell, there should be N peaks of electron density. On the Patterson's map, each of this N atoms is linked by a vector with every N atom (with itself too), that gives N^2 vectors. N of these vector connects atoms with themselves, that will generate a large peak at position zero. The rest of the N^2-N peak will surround that peak.

The figure 4 shows a Patterson's map for a unit cell with one molecule. A Patterson's map is sum of images of the molecule, with each atom placed in sequence in the origin of coordinate system. Because for each vector there is one in the opposite direction, the same Patterson map is also a sum of inverted images of the molecule, as shown in the bottom representation.

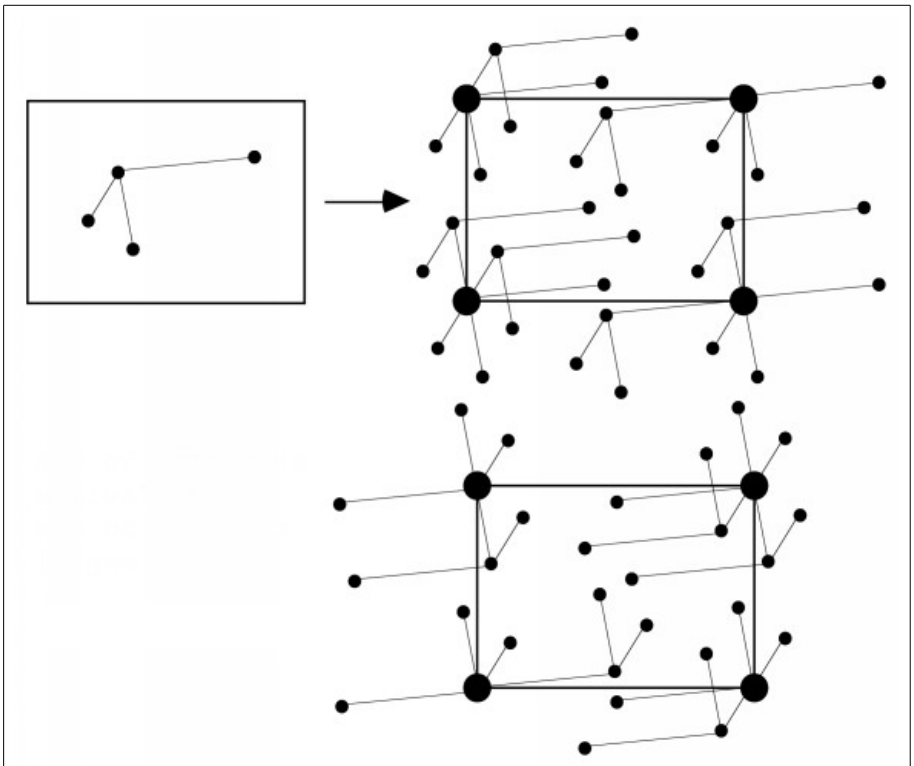


Fig. 4: Patterson's map for unit cell with one molecule inside

For sufficiently small molecules, the positions of atoms, that would give rise to peaks on Patterson's map, can be determined. This operation is called deconvoluting the Patterson's function. But it quickly becomes impossible to deconvolute a Patterson for larger molecules. If there is N atoms in a unit cell and the resolution of the data is high enough, there will be N separate electron density peaks in an electron density map, but in Patterson map there will be N^2 vectors. N of these vectors create big origin peak, but there is still N^2 - N vectors to assign to atoms.

1.3.3. Direct methods

Direct methods are based on presence of heavy atom, means atom with large atomic number, thus large number of electrons. The natural occurrence of heavy atoms can be used, or an isomorphous derivative can be created.

If there is few heavy atoms in the structure in comparison tu number of light atoms, the Patterson's function become easier to solve. The positions of heavy atoms are easy to determine. When this potions are know, the electron density function can be span into Fourier series. As the heavy atoms gives main contribution to phases, in first approximation they can be just

equal.

1.4. Refining and validation of structure

A structer obtained by one of above descibed mothod as an approximate structure, with not well defined details.

Fo assessment of structure quality the factors of discrepancy R are calculated. They are defined as average difference between structure factors computed from model F^{calc} and experimental structure factors $F^{\exp}$ according to the following expresion:

$$R=\frac{\sum_{hkl}\left||F^{\exp}_{hkl}|+|F^{calc}_{hkl}|\right|}{\sum_{hkl}|F^{\exp}_{hkl}|}$$

The smaller R is the better model fit the data. The minimisation of R is achieved by using methods based on Fourier series or least square fit.

2. Crystals

In the ideal crystal molecules are regularly ordered in space creating three-dimensional net, called crystallographic lattice. Unit cell is primary structural element of crystal repeating in space and creating that crystallographic lattice. Unit cell is parallelepiped characterised by three edges of lengths: a, b, c and three angles between those edges: α, β, γ. Lengths and directions of edges of unit cell are defined as three vectors: $\vec{a}, \vec{b}, \vec{c}$, these vectors span a crystallographic real space. Position of every point in that space can be described by the translation vector: $\vec{r}\,' = \vec{r} + n\vec{a} + m\vec{b} + l\vec{c}$. Points of coordinates (na, mb, lc) are called nods of the lattice (n,m,l are any integers).

Sets of parallel Bragg's plane are described by Miller indices: h, k,l. Values of Miller indices are given by the points of intersections of axis coordinate system with this one of the set of planes, which is the closes to the origin of coordinate system and goes through the nodes of lattice(Fig. 5)

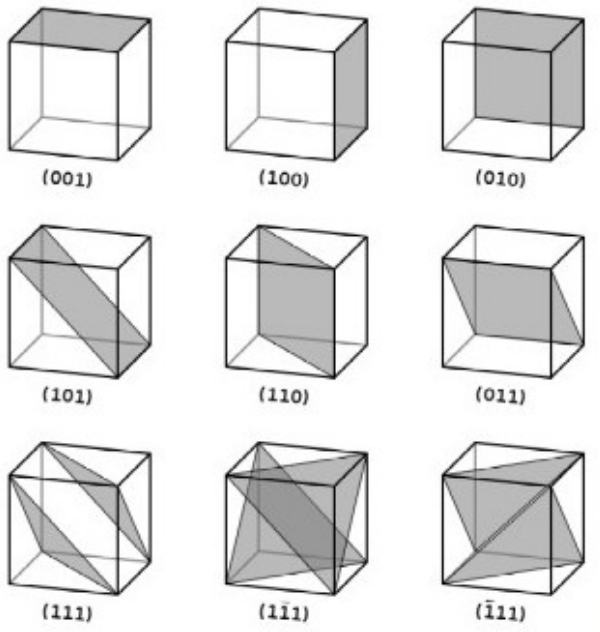


Fig. 5: Miller indices for basic crystallographic planes in cubic unit cell[3]

Crystallographic lattice in reciprocal space is also defined by edges of the unit cel: $\vec{a}^*, \vec{b}^*, \vec{c}^*$ and angles between this

edges: $\alpha^*, \beta^*, \gamma^*$. The relationship between vectors spanning real and reciprocal space are as follows:

$$\vec{a}^* \cdot \vec{a} = \vec{b}^* \cdot \vec{b} = \vec{c}^* \cdot \vec{c} = 1$$

$$\vec{a}^* \cdot \vec{b} = \vec{a}^* \cdot \vec{c} = \vec{b}^* \cdot \vec{c} = 0$$

This means that a^* is perpendicular to b and to c and it's length equals $1/a$.

Point described in a real space with vector $\vec{r}' = n\vec{a} + m\vec{b} + L\vec{c}$, is described with vector $\vec{H} = h\vec{a}^* + k\vec{b}^* + L\vec{c}^*$ in a reciprocal space. The symmetry of real space is preserved in reciprocal space, not only geometry but also the intensity of scattered waves.

Relations between edges lengths and angles of unit cell defines seven crystal classes(Fig. 6): cubic, tetragonal, orthorhombic, hexagonal, trigonal, monoclinic, triclinic.

Primitive lattice, build of prmitove cells (P) have nods only in the apexes of cell. Nonprimitive lattices can have more nods: on the faces (*face-centered,* F), on two opposite faces (side-*centered,* C) or in centre of the unit cell volume *(body-centered,* I). Seven crystal classes together with four type of unit cells creates fourteen Bravais lattices(Fig. 6).

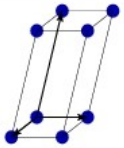
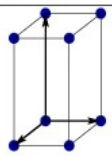
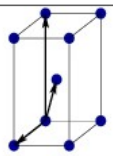
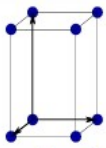
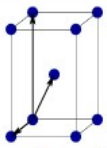
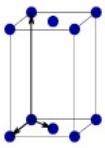
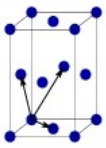
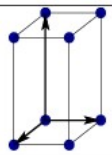
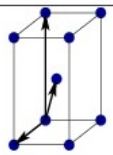
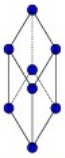
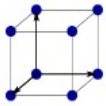
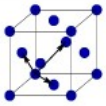
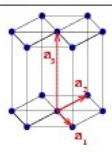
Bravais lattice	Parameters	Simple (P)	Volume centered (I)	Base centered (C)	Face centered (F)
Triclinic	$a_1 \neq a_2 \neq a_3$ $\alpha_{12} \neq \alpha_{23} \neq \alpha_{31}$				
Monoclinic	$a_1 \neq a_2 \neq a_3$ $\alpha_{23} = \alpha_{31} = 90°$ $\alpha_{12} \neq 90°$				
Orthorhombic	$a_1 \neq a_2 \neq a_3$ $\alpha_{12} = \alpha_{23} = \alpha_{31} = 90°$				
Tetragonal	$a_1 = a_2 \neq a_3$ $\alpha_{12} = \alpha_{23} = \alpha_{31} = 90°$				
Trigonal	$a_1 = a_2 = a_3$ $\alpha_{12} = \alpha_{23} = \alpha_{31} < 120°$				
Cubic	$a_1 = a_2 = a_3$ $\alpha_{12} = \alpha_{23} = \alpha_{31} = 90°$				
Hexagonal	$a_1 = a_2 \neq a_3$ $\alpha_{12} = 120°$ $\alpha_{23} = \alpha_{31} = 90°$				

Fig. 6: Bravais lattices

In molecular crystals unit cell are often quite big and contain more than one molecule, that means that there is another element of symmetry, smaller than the unit cell, it's called motive. motive copied in space according to symmetry operations

given by crystal group of symmetry fills the unit cell.

3. Diffractometers

There are various types of diffractometers, from so called pocket diffractometers to X-rays beam created in synchrotrons. Every diffractometer has to contain X-ray source, detector, to register diffraction images, goniometer, to allow positioning of the crystal in the X-ray beam, and cryo-system, to keep crystal in safe temperature of 100K. There is a large variety of ways, this parts can be done. This brief introduction is not designed to be a lecture on X-ray diffraction equipment. The main parts are presented for the SuperNova diffractometer by Oxford Diffraction, an instrument which works in Department of Biophysics.

3.1. Source

The source is a microfocus sealed tube X-ray generator with unmovable copper anode.

The X-ray source is a vacuum tube containing cathode and anode. The cathode is heated by high intensity current, the heat causes cathode to emits electrons, which are accelerated by strong electric field towards anode. The field is generated by high voltage(30-150kV) power source connected across cathode and anode. Electrons collide with anode, carrying enough energy to strike out electrons from atomic inner shells. Electrons coming back to its shells emit few characteristic wavelenghts. In protein crystallography the Kα line is used, of the wavelength of 1,5418Å

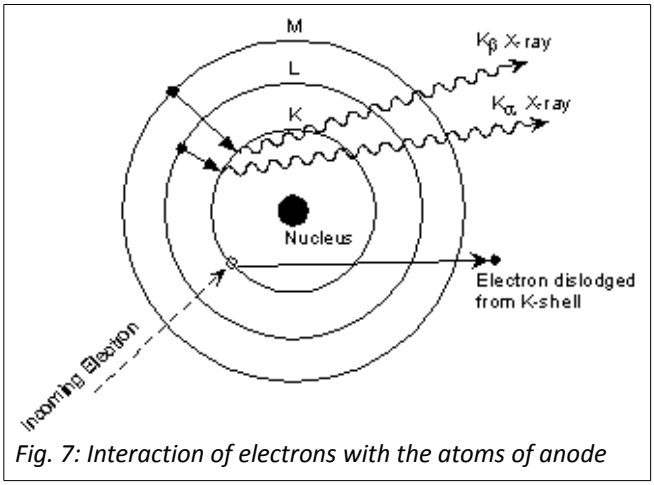


Fig. 7: Interaction of electrons with the atoms of anode

Only about 1% of energy absorbed by anode is emitted in the form of radiation. The rest change into heat, and cooling anything in the vacuum is not easy task. There are two solutions that are used: rotating anode and micro focusing of electron beam. When anode rotates, the electron beam collides with different part of anode, not allowing to overheat any part of that. In case of microfocusing, the electron beam is precisely collimated, and collide with very small area of anode, of diameter of tens of micrometers, this allow reduce intensity of electron beam and heating of anode. Another advantage is better collimation of X-ray beam emitted from anode, reducing power consumption and prolonging X-ray source lifetime.

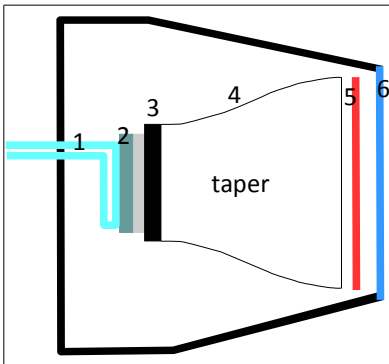
The X-ray beam is additionally collimated by optical system, given finally beam of diameter of tens to few hundreds micrometers.

3.2. Detector

For the X-ray detection a scintillation counter combined with CCD camera is used.

Scintillator is substance, which hit by quantum of ionisation radiation produces a photon of visible radiation. Behind the scintillator there is taper, an optical element, which “scales” image from scintillator into CCD matrix, which is smaller. To decrease noise level from CCD matrix this is cooled to -40°C . This system also works in vacuum. The berylium window in front

of scintillator selas the vacuum compartment, beryllium is chosen because is well transparent for the X-ray radiation

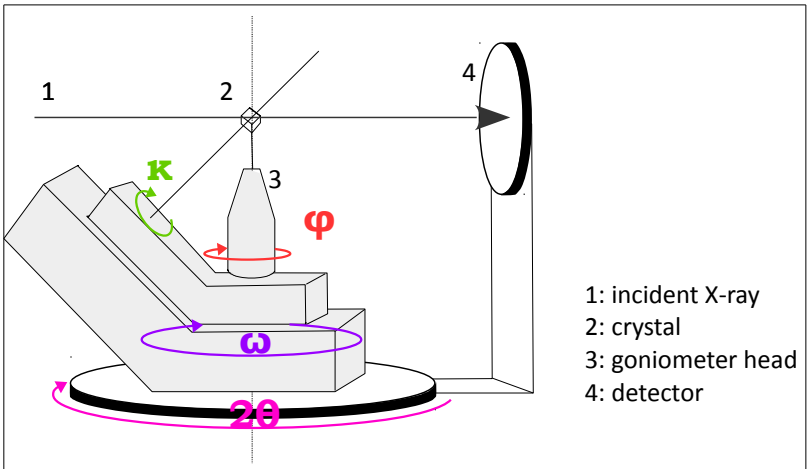


1: cooling of Peltier's element
2: Peltier's element
3: CCD matrix
4: taper
5: scintillator
6: beryllium window

Fig. 8: Scheme of detector

3.3. Goniometer

Protein crystal is mounted on the head of the goniometer. This devices allow to align crystal precisely in relation to X-ray beam and detector, and rotate crystal into almost every position, to collect waves scatter into whole sphere. The goniometer showed in Fig. 9 is so called 4-circle kappa goniometer, there are three axes of rotation for the crystal (the angles of rotation φ, κ i ω are shown on the scheme) and one axis of rotation for the detector 2θ. Goniometer is equipped with precise steeper motor which allows adjustment of crystal position with accuracy of about 10µm.



1: incident X-ray
2: crystal
3: goniometer head
4: detector

Fig. 9: 4-circle kappa goniometer

Through whole the experiment crystal is in the jet of gaseous nitrogen at the temperature 100K

4. Measurement of crystals diffraction patterns

Collecting full data set, means measure all of the scattered waves, it may demand collecting data from the full sphere(when crystal has now rotational symmetry) but usually its just a fraction of that, because of rotational symmetry of the crystal. The diffraction pattern repeats every 180, 90 or 60° dependently of the group of symmetry of crystal lattice, so its enough to collect that fraction of full sphere with reasonable margins.

Therefore, for designing reasonable data collection, first the inner symmetry of crystal has to be determined. Pre-experiment consist of collecting at least two images, often spaced by 90°, this is enough for initial assessment of group of symmetry, shape and dimension of unit cell. Based on this data, with the help of software designed for that purpose the full experiment is planned, and subsequently data is collected.

Collecting diffraction images is the last experimental step in protein solving, all the subsequent steps are computational.

5. Solving structure

For structure solving the program Autochem and Olex2 will be used.

6. Bibliography

[1] Bernahard Rupp, Crystallography 101, http://www.ruppweb.org/Xray/101index.html
[2] University of Cambridge, Cambridge Institute for Medical Research (CIMR), Protein Crystallography Course, http://www-structmed.cimr.cam.ac.uk/course.html
[3] Yuan Ming Huang, Solid State Physics: Miller indices, http://www.lcst-cn.org/Solid%20State%20Physics/Ch16.html
[4] Meaurice Van Meerssche, Janine Feneau-Dupont: Krystalografia i chemia strukturalna, PWN, Warszawa 1984.
[5] C. Giacovazzo: Fundamentals of crystallography, Oxford : Oxford Univ. Press, 2002.

7. Further reading

- C. Giacovazzo: Fundamentals of crystallography, Oxford : Oxford Univ. Press, 2002.

8. Topics for preliminary test

- Interacting of light and matter – diffraction, interference
- Crystals features, symmetry of crystals.
- What can be seen on diffraction images?
- The Phase Problem and methods of solving it.
- Refinement and validation of structure.

9. Performing the exercise

The aim of this exercise is to collect data and solve structure of small molecule, e.g. sucrose.

The crystal will be provided. By visual assessment the best crystal will be chosen. The full dataset will be collected and structure solved with program Autochem or Olex2.