Electron crystallography of three dimensional protein crystals

Georgieva, D.

Electron Crystallography of Three Dimensional Protein Crystals

Dilyana Georgieva

Electron Crystallography of Three Dimensional Protein Crystals

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. P. F. van der Heiden,
volgens besluit van het College voor Promoties,
te verdedigen op donderdag 11 december 2008
klokke 10.00 uur

door

Dilyana Georgieva

Geboren te Sofia, Bulgarije
in 1980

Promotiecommissie

Promotor:	Prof. dr. J.P. Abrahams
Overige leden:	Prof. dr. J. Brouwer
	Prof. dr. H. W. Zandbergen
	Prof. dr. M. Noteborn
	Prof. dr. S. Hovmöler
	Dr. R. de Graaff
	Dr. N. Pannu
	Dr. S. Nikolopoulus
	Dr. M. Kuil

This thesis is part of the research program Stichting voor Fundamenteel Onderzoek der Materie (FOM), financially supported by Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO).

Contents

Chapter 1

Introduction

1.1 Outline

One powerful method for structure determination is diffraction. Single-crystal X-ray crystallography is a well established technique for structure determination even of beam sensitive materials, in the solid state. A prerequisite for using X-ray diffraction is the availability of macro-crystals, which at the same time is the rate limiting step of this technique. Nucleation and crystallization of biological macromolecules remain poorly understood. Currently the only way of finding the right crystallization conditions is by trial and error. Often sub-micron crystals are obtained initially and these crystals need to be grown further into macro-crystals. Such an optimization is not trivial and sometimes even impossible. Using electron sources relaxes the requirement for macro-sized crystals. The higher scattering cross-section of electrons compared to X-rays allows diffraction information to be obtained from nano-crystals. If electron diffraction can be shown to work for a 3D protein crystal, substantial progress in structural biology may be achieved.

Electron diffraction has been used for structure determination since the 1960s. Due to limitations related to sample preparation methods and collection of diffraction data of sufficient quality (resolution) as well as the lack of algorithms for analyzing data acquired from different nano-crystals (the high beam sensitivity of protein crystals does not allow a full 3D data collection from one single nano-crystal), the method did not grow in importance. Current advances in cryo-electron microscopy [1] allow bio-molecules to be successfully vitrified and studied in their frozen-hydrated state. In the area of data recording two important techniques have been introduced: CCD cameras [2] and image plates which made it possible to quickly and accurately collect data.

This thesis describes an investigation of the potential of electron diffraction for studying three dimensional sub-micro-crystals of proteins and pharmaceuticals. **Chapter 1** gives a general introduction to the theory of protein crystallization and describes some current trends in this field. Emphasis is given on nano-crystallization

and heterogeneous crystallization. A short overview is given of the techniques for studying the mechanism of heterogeneous crystallization used to grow protein nano-crystals for electron diffraction studies. Some basic concepts of the electron diffraction theory as well as the different TEM diffraction techniques relevant to understanding the work presented, are also discussed.

A prerequisite for using electron diffraction for structural studies is the predictable availability of tiny crystals. In **Chapter 2** a method for growing such crystals using heterogeneous nucleation is demonstrated. The heterogeneous nucleant (in this case hair fibers) was serendipitously selected. Four different proteins (lysozyme, glucose isomerase, a Fab fragment and potato protease inhibitor) were shown to nucleate preferentially on the selected substrate and sub-micron crystals were grown. Further studies on the mechanism of heterogeneous nucleation using lysozyme as a test protein and different imaging techniques such as atomic force and fluorescence microscopy are also discussed.

Sub-micron crystals of potato protease inhibitor and lysozyme were subject of electron diffraction studies. A detailed description of the diffraction experiments is presented in **Chapter 3**. A special focus is given on the sample preparation procedure and in particular the vitrification and the cryo-preservation of the crystals. The preliminary results showed that the heterogeneously grown nano-crystals are well ordered and suitable for electron diffraction. The high beam sensitivity of the protein nano-crystals appeared to be the rate limiting step in the data collection, not allowing orientation of the crystals (a technique used in electron diffraction studies of inorganic crystals) or a 3D data collection of a single crystal (a technique used in X-ray protein crystallography). This suggested that new approaches for data collection and data analysis needed to be developed.

Optimization of the diffraction data collection, as is described in **Chapter 4** allowed high diffraction resolution (up to 2.1Å) to be obtained from vitrified lysozyme crystals. An algorithm for unit cell determination of randomly oriented diffraction patterns of different crystals is presented. The method was used for the analysis of the electron diffraction data acquired from lysozyme nano-crystals.

The methods for collecting and analyzing electron diffraction data from lysozyme crystals were also confirmed in the case of penicillin-type nano-crystals. The motivation behind these studies and the results obtained are discussed in **Chapter 5**. In this case crystalline powder samples were subjected to electron diffraction studies. Resolutions up to 0.8Å were obtained from oxacillin crystals and up to 1Å from

penicillin G crystals. The unit cell parameters found from the analysis of electron diffraction data with the algorithm presented in the previous chapter were consistent with the unit cell parameters obtained by X-ray crystallography on the same two types of penicillin.

The thesis is finalized with conclusions and future perspectives described in **Chapter 6**. A summary (in both English and Dutch) is presented.

1.2 Structural analysis

Partly publishes as: Georgieva, D.G., Kuil, M.E. and Abrahams, J.P. (2006). Protein nanocrystallization. *Springer series in Biophysics, Advanced Techniques in Biophysics, Springer-Verlag Berlin Heidelberg*.

The study of 3D structures of materials covers the recognition, nature and functionality of materials. Many properties of matter can be understood from its crystallographic analysis. The crystallographic structural information gives us not only the ability to understand the specific functions of different materials but also to design new structures with altered properties or to compensate for existing defects. A particular example in life science is the development of structure-based drug design (also known as rational drug design). The classical method of drug discovery is based on the trial-and-error testing of chemical substances on living organisms. The obtained effects are then related to the applied treatment. In contrast, the rational drug design approach involves the use of three-dimensional structural information about bio-macromolecules. For example, some approaches attempt to stop the functioning of a pathologic pathway by causing a key molecule to stop functioning. In such cases drug molecules may be designed that bind to the active region, inhibiting this key molecule. Structural information about inorganic materials is widely used also in the synthesis and search for novel materials with superior properties such as for instance high temperature superconductors.

Diffraction techniques, NMR spectroscopy and electron microscopy are the three main methods used for structure determination. Most of the recent innovations within NMR spectroscopy have been made in the field of protein NMR which has made this technique very important for structural biology. One common goal of these innovations is to obtain a resolution similar to X-ray crystallography. Advances in cryo-electron microscopy and the development of high resolution electron microscopes allow direct imaging of inorganic structures and macromolecules. However, the highest resolution and therefore the most detailed structural information is still obtained using diffraction techniques. The diffraction phenomena refer to the propagation of waves when they encounter obstacles in their path. While diffraction always occurs when propagating waves (such as sound waves, water waves, electromagnetic waves) interact with an obstacle, its effect is more prominent when the wavelength is of the order of the size of

the diffracting objects. The complex patterns resulting from the intensity of diffracted waves are a result of interference between different waves that are travelling via different paths.

1.3 Protein nucleation and crystallization

A prerequisite for using crystallographic techniques for structure determination is the availability of well diffracting crystals. Biological macromolecules follow the same thermodynamic rules as inorganic or organic small molecules concerning supersaturation, nucleation and crystal growth. However, protein macromolecules are organized in tertiary and quaternary structures. The intramolecular interactions responsible for their tertiary structure, the intermolecular interactions involved in the crystal contacts, and the interactions necessary to solubilize them in a solvent are similar. These different interactions may compete with one another in solution. It has to be considered as well that the biological properties of the macromolecules may be conserved but their physico-chemical properties, such as their net charge depend on the crystallization conditions (pH, ionic strength etc.). A charged biological macromolecule requires counterions to maintain the electroneutrality of the solution. In this case it is considered as a protein salt with its own physico-chemical properties.

To crystallize a biological macromolecule, its solution must have reached supersaturation which is the driving force for crystal growth. The under- and supersaturated states are defined by the solubility of the macromolecules (see Figure 1.1). When the concentrations of the crystallization agent and the macromolecules correspond to the solubility condition, the saturated macromolecule solution is in equilibrium with the crystallized macromolecules. Below the solubility curve the solution is undersaturated and the system is thermodynamically stable. In this case, phase transition (crystallization) will not occur. Above the solubility curve, the concentration of the biological macromolecules is higher than the concentration at equilibrium. This corresponds to the supersaturation zone. A supersaturated macromolecular solution contains an excess of macromolecules which will appear as a solid phase until the macromolecular concentration reaches the solubility value in the solution. The higher the supersaturation, the faster this solid phase appears. However, at very high supersaturation precipitation, no crystallization occurs. At this stage insoluble macromolecules rapidly separate from the solution in an amorphous state. A

schematic representation of the solubility diagram showing the different zones is shown in Figure 1.1.

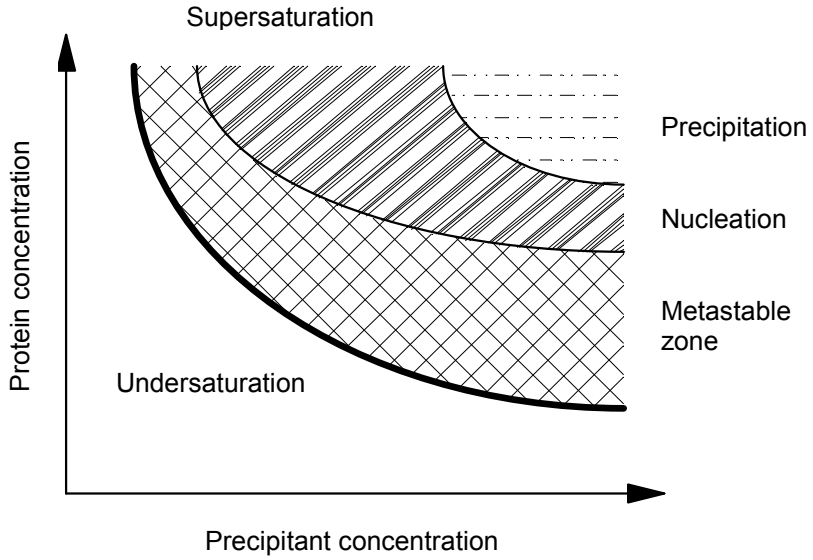


Figure 1.1 Two dimensional solubility diagram showing the different zones of the supersaturation domain. The thick solid line represents the solubility boundary. (From Decruix & Giege, 1999)

Unlike in small molecule crystallography, nucleation is usually the rate limiting step in protein crystallization. When a solution is supersaturated, the solid phase forms more or less rapidly depending on the conditions: concentration of solute, crystallization agent etc. Unfortunately, in practice it is very difficult to predict when or where an unknown protein will nucleate and crystallize. Finding the right crystallization conditions is to a great extent still a process of trial and error. Two different types of nucleation can be distinguished. If the nuclei form in the bulk of the solution, the nucleation is called homogeneous. On the other hand, if the nuclei preferentially form on substrates such as the wall of the crystallizer, solid particles etc, it is called heterogeneous. From a theoretical point of view nucleation is considered as an addition of monomers to clusters consisting of a few monomers called i-mers. When the system is in steady state, the rate of formation of such a cluster is equal to its rate of disappearance.

The small clusters turn into stable nuclei only if they contain more than a critical number of monomers.

Since the solute concentration is the same everywhere in the bulk, nucleation occurs if there are energy fluctuations, somewhere in the solution, around the mean value imposed by the supersaturation. To create a nucleus it is necessary to create a volume and a surface. The activation free energy for homogeneous nucleation can be expressed as follows [3]:

$$\Delta G = -ik_bT \ln \beta + A_1\gamma_1$$

where i is the number of molecules in the nucleus, A_1 the area of the nucleus, γ_1 interfacial free energy with respect to the solution, k_b the Boltzmann constant, T the absolute temperature, and β = C/Cs where C and Cs are the actual and the saturation concentration. The first term represents the energy to create the volume. The second term is the excess energy to create the surface. If it is assumed that the nucleus is a sphere then:

$$\Delta G = -\frac{4}{3}\pi r^3 k_bT \ln \beta + 4\pi r^2 \gamma_1$$

At equilibrium, the nucleus has the critical radius r* [3]:

$$r^* = \frac{2\gamma_1 V}{k_bT \ln \beta}$$

where V is the volume of a molecule. Therefore ΔG* can be expressed also as:

$$\Delta G^* = -\frac{16\pi V^2 \gamma_1^3}{3(k_bT \ln \beta)^2}$$

and it can be also written as:

$$\Delta G^* = \frac{1}{3}(4\pi r^{*2}\ \gamma_1)$$

The critical activation free energy for creating the nucleus with critical radius r* is one third of the energy required for creating its surface. At the critical r*, the nucleus is in a very labile equilibrium. If it gains one molecule, so that r > r* it grows. If it loses one molecule, so that r < r* then it spontaneously dissolves.

Heterogeneous nucleation often occurs prior to homogeneous nucleation especially when supersaturation is low.

However, this implies that the solute molecules have some affinity for the substrate onto which they stick. To simplify the matter, the nucleus can be considered cap-shaped (see Figure 1.2), making a contact angle α with the substrate [3].

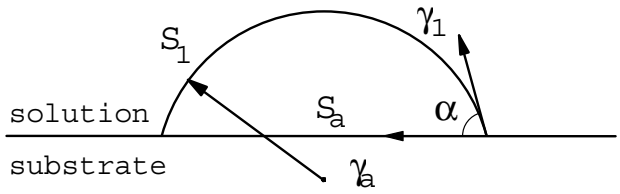


Figure 1.2 Cap-shaped nucleus formed by heterogeneous nucleation on a substrate. (From Decruix & Giege, 1999)

Three surface free energies are involved in heterogeneous nucleation: γ_1 between the nucleus and the solution, γ_a between the nucleus and the substrate, and γ_0 between the substrate and the solution. They are related by Young's equation:

$$\gamma_0 = \gamma_a + \gamma_1 \cos \alpha$$

If S_1 is the area of the nucleus and S_a the area of the interface between the nucleus and the substrate, the activation free energy for heterogeneous nucleation is:

$$\Delta G_{het} = -ik_b T \ln \beta + S_1 \gamma_1 + S_a \gamma_a - S_a \gamma_0$$

Taking Young's equation into account ΔG_{het} becomes:

$$\Delta G_{het} = -\frac{4\pi r^3}{3V} \frac{2 - 3\cos\alpha + \cos^3 \alpha}{4} k_b T \ln \beta$$
$$+ 4\pi r^2 \frac{1 - \cos\alpha}{2} \gamma_1 - 4\pi r^2 \frac{1 - \cos^2 \alpha}{4} \gamma_1 \cos\alpha$$

At equilibrium, the radius of the critical nucleus is:

$$r^*_{het} = \frac{2\gamma_1 V}{k_b T \ln \beta}$$

The critical radius of the nucleus formed by heterogeneous nucleation is the same as for homogeneous nucleation. However, the cap-shaped nucleus contains fewer molecules than does the full sphere of the same radius.

ΔG_{het} can be also expressed as:

$$\Delta G^*_{het} = \Delta G^* (\frac{1}{2} - \frac{3}{4}\cos\alpha + \frac{1}{4}\cos^3\alpha)$$

If α =180º, the substrate doesn't have any effect on nucleation. For α=90º, $\Delta G^*_{het} = \Delta G^* / 2$

If α tends towards zero, then ΔG^*_{het} tends also to zero. That means that the nucleation rate drastically increases whit a decreasing contact angle α: the higher the affinity of the molecule for the nucleating substrate, the higher the nucleation rate.

1.4 The role of heterogeneous substrates in the process of protein nucleation and crystallization

The supersaturated protein solution is thermodynamically unstable. The process of nucleation and crystal growth depend on various factors. As long as the pH, the temperature, the nature of solvent constituents are constant, the solubility remains unchanged, but the nucleation zone as well as the precipitation zone may be shifted depending on the crystallization technique. For example, in the metastable zone the critical supersaturation is not yet reached and therefore spontaneous nucleation cannot occur unless it is induced by heterogeneous substrates. In general, additives play an important role in protein crystallization. Heterogeneous substrates are usually regarded as additives when they are purposefully added to the solution in order to obtain a desired effect (inhibition of nucleation, habit change of crystals). However, impurities of foreign substances may also exist in the solution originating from other sources (the solvent, crystallization agent etc). Often it is difficult to make a clear distinction between additives and impurities. Heterogeneous crystallization which is induced by a properly chosen additive may allow better control of nucleation and growth. However, it has to be considered as well that heterogeneous substrates may also induce a modification of the crystal morphology and in some special cases they may induce a

polymorph transition. A typical example when impurities or additives influence crystal growth is the crystallization of proteins at different pH or in the presence of different salts. It cannot be ignored though that the crystallization agents, in general salts, may be incorporated in crystal structures and therefore the different forms obtained cannot always be regarded as polymorphs: polymorphs are forms of the same chemical composition with different crystal structures. Often those different forms of the same protein are referred to as different phases of the same compound.

However, foreign substrates not always induce or facilitate protein nucleation and crystallization. Impurities absorb on the terraces between the growth steps, along the steps or in the growth sites. Depending on the energy of the bonds between the impurity and the adsorption sites, adsorption is a reversible event. When growth proceeds, there is a competition between the rate of molecule incorporation and the rate of impurity adsorption and desorption. Accordingly, impurities obscure the crystallization process so that nucleation and growth rates are sometimes drastically slowed down as shown by Vekilov [4].

1.5 Protein nano-crystallization

For several reasons studying protein nano-crystallization is a pertinent issue, especially when we consider crystallization of proteins which are physically confined within a very small volume. There is a practical reason also for studying protein crystallization in small confined volumes. As crystallization conditions can be found through a trial and error only, current practice requires simultaneous testing of many different conditions. The obvious idea that minimizing the volume of single tests, maximizes the number of different conditions that can be screened with a given quantity of protein prompted the development of high-throughput nano-crystallization systems [5, 6].

Although nano-crystallization is quickly becoming a mainstream method, the crystallization step remains the major bottleneck in the structure production process [7]. This is illustrated by recent data from a large structural genomics initiative, indicating that the least successful step in going from sequence to structure is the one from purified protein to crystals (see Figure 1.3). The overall trend illustrated in Figure 1.3 is not very different from a report predating the widespread use of nano-crystallization [8]. Microheterogeinity of the proteins may be the prime cause of this bottleneck.

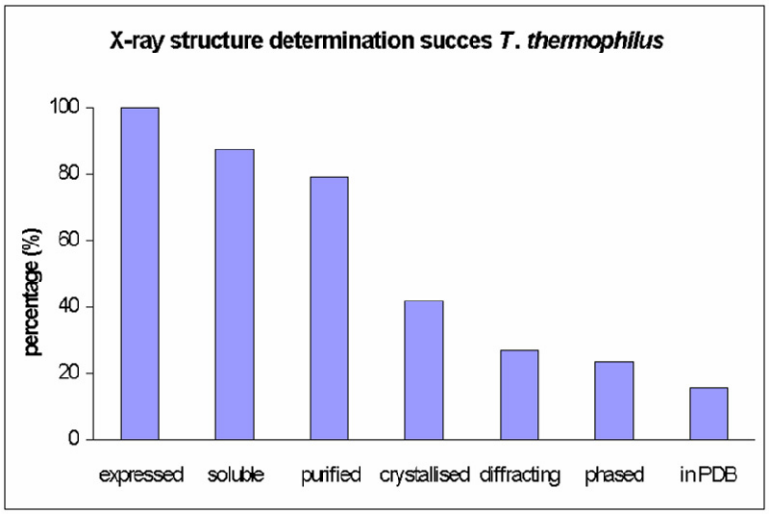


Figure 1.3 The success rate of high-throughput crystallization. The overall success of the different stages in the high-throughput approach used by RIKEN consortium is shown. The numerical data were presented at the ICCBM10 conference in Beijing by S. Yokoyoama and represent the throughput obtained using expression in Thermus thermophilus.

Constructing genetic variants and developing more advanced means of protein production and purification may increase the success rate. Nevertheless, advances in nano-crystallization should also accompany this as nano-crystallization favours throughput whilst substantially reducing demands on large-scale production and purification platforms.

One might think that the protein concentration determines the level of supersaturation regardless of the volume. However, this probably is not the case, considering that in tiny droplets the surface tension forces become relevant and below a certain volume even predominant. Inside a small nano-droplet the pressure can be substantially higher than the ambient pressure, it may be calculated using the Young-Laplace equation [9, 10]. However, these effects are less likely to influence protein crystallization in the microlitre range. The pressure difference between the inside of a water droplet of 100 μm radius and the gas phase for a surface tension of 72 mN/m is only equal to 1.44 kPa (kN/m^2). Lorber and co-workers [11] studied the influence of external hydrostatic pressure on the nucleation and growth of lysozyme nano-crystals and reported that increasing the pressure from 0.1 MPa (atmospheric pressure) to 250 MPa

leads to reduction of the size and number of lysozyme crystals. Moreover a transition to urchinlike particles made of crystalline needles progressively occurs.

These considerations are obviously irrelevant when the protein is confined within a lipid membrane and thus do not apply for proteins dissolved in the cytoplasm of living cells. The pressure inside a living cell is well regulated and partially determined by the presence of surrounding tissue. In plant cells the turgor of intracellular pressure can reach several atmospheres at most [12].

For practical purposes it is more important that the homologous nucleation rate in protein crystallization is theoretically determined by the level of superstauration, and it is independent on the volume of the mother liquor. However, if at a certain level of supersaturation, it takes on average a whole day to form a stable nucleus that grows into a macroscopic protein crystal in say 1μL, then it would take 50 days on average for a similar event to occur in a volume of 20 nL. If the nucleation rate per unit volume is constant, reduction of the crystallization volume therefore results in a reduced chance of finding crystals. In other words, one has to increase the level of supersaturation in nano-liter crystallization trials in order to observe rare nucleation events. The relation between the crystallization volume in sub-microliter volumes and the observed number of crystals is shown in Figure 1.4 and indicates that there is a dependence on the droplet volume [13].

The relation appears to be linear but it doesn't go through the origin, indicating that a basic assumption of the nucleation theory is not satisfied. This suggests that heterogeneous nucleation plays an important role in low volumes. Vekilov and Galkin reported that despite precautions, heterogeneous nucleation is always observed in their experiments and led to a non-zero intercept of the linear dependence of N (mean number of observed crystals) as a function of the induction time ΔT, in a volume of 700nL [14, 15].

Although the probability of finding a crystal is very low, a nucleus can always be formed owing to a spontaneous (homogeneous) nucleation event because of density fluctuations [16]. At this point two types of heterogeneous nucleation can be distinguished: heterogeneous nucleation that depends on nuclei that float in the bulk solution and heterogeneous nucleation that is somehow related to the surface of the mother liquor. In the first case homogeneous and heterogeneous crystallization cannot be distinguished by changing the crystallization volume.

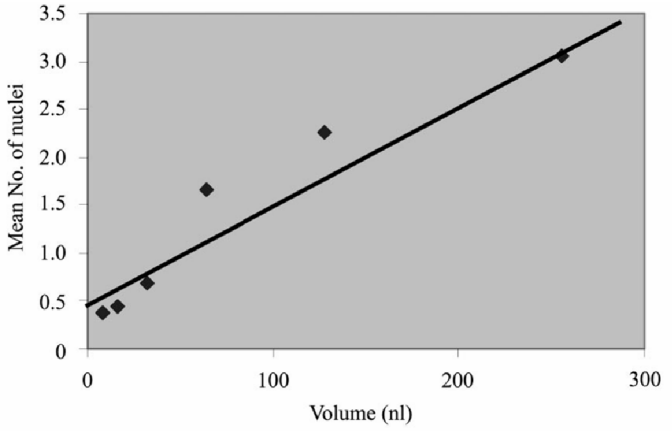


Figure 1.4 Heterogeneous nucleation in sub-microliter volumes. The average number of tetragonal crystals per droplet detected 24h after mixing as a function of the volume of the droplet. Each data point is the count obtained from 16 droplets. In the smaller droplets needle-like crystals showed a higher relative abundance. (From Bodenstaff *et al*., 2002)

In the latter case reduction of the crystallization volume would increase the relative contribution of heteogeneous nucleation. On the basis of the experimental results it can be argued that a certain (very low) volume may exist below which heterogeneous nucleation will be the dominant nucleation mechanism [17]. The early stages of crystallization have been probed using fluorescence energy transfer [18] but the mechanism of nucleation (homogeneous or heterogeneous) remains poorly understood. Most of the atomic force microscopy on protein crystals focused on crystal growth [19] with the exception of the work from the Vekilov group [20]. To induce nucleation or to reduce the induction time of crystallization, different engineered and natural seeding materials have been tested, but they turned to be successful only for certain proteins. This indicates that probably there is no "universal nucleating surface", finding a suitable substrate is another process of trial and error in the quest for crystals [21, 22].

Another factor which significantly influences protein crystallization in nano-volumes is the process of mixing of liquids. Two consecutive processes dominate the mixing of liquids: fusion of the liquid boundaries and diffusion of the components. Although both processes are similar in bulk fluid and microfluidics, their outcome is significantly different; unexpected mixing results in microfluidics have been reported. Two liquid streams can flow alongside each other in a tube a few micrometers wide over a period

of time without mixing, almost as if the were separated by glass [23]. The fusion step is not only the first step, but is also the rate-limiting step in the mixing process. Macroscopically stirring can speed up fusion, as turbulence increases the interfacial area between the liquids, but in small channels it is almost impossible to produce such a turbulent flow. Mechanical forces such as in shaking or thermal forces inducing convective flows are less effective and more difficult to apply to very small volumes. A recent approach to accelerate mixing of small volumes is electro-osmosis, where components are displaced by electric fields. For the aim of protein nano-crystallization mixing of protein droplets on a small scale can best be achieved during the dispensing phase. For example, in the Microdrop robot system droplets are shot from the nano-dispenser with a linear velocity of 3-5m/s and they drive in the bulk solution without splashing as there is not enough energy for splashes to be formed - droplets with a diameterless than 100µm have more surface energy than kinetic energy at the speeds generated (see http://www.microdrop.de).

When two droplets meet on a solid surface they usually fuse if the liquids are miscible. Little is known about what happens next in the subsequent diffusion of dissolved components (e.g. proteins) in high concentration within the fused nano-volumes. Techniques like dynamic light scattering and fluorescence correlation spectroscopy probe diffusion in very small volumes of a few micrometers [24]. Although usually homogeneous mixing is aimed for, the lack of homogeneous mixing can sometimes also be an advantage as is demonstrated in free-interface diffusion methods used in protein crystallization without evaporation [25, 26].

1.6 Fluorescence microscopy

The fluorescence is based on the phenomenon that certain materials emit light energy detectable as visible light when irradiated with light of specific wavelength. The sample can be fluorescing in its natural form like chlorophyll and some minerals, linked to fluorescing chemicals called fluorophores (such as green fluorescent protein, fluorescein etc.) Fluorescence occurs when a molecule, atom or nanostructure relaxes to its ground state after being excited. Usually the absorbed photon has a high energy (and a short wavelength), and the emitted light has a lower energy (and a lower wavelength). This depends also on the absorbance curve and the Stoke shift (the difference in wavelength or frequency between the maxima of the absorption and emission spectra). When a system absorbs a photon, it gains energy and enters an

excited state. One way for the system to relax is to emit a photon, thus losing energy (another method would be the loss of heat energy). Fluorophores can lose their ability to fluoresce as they are illuminated in a process called photobleaching.

To become detectable, the emitted light is separated from the much brighter excitation using a filter which filters out the excitation photons but transmits the emitted light which is of lower energy and has a longer wavelength. The fluorescing areas can be observed in the microscope and stand out against a dark background with high contrast.

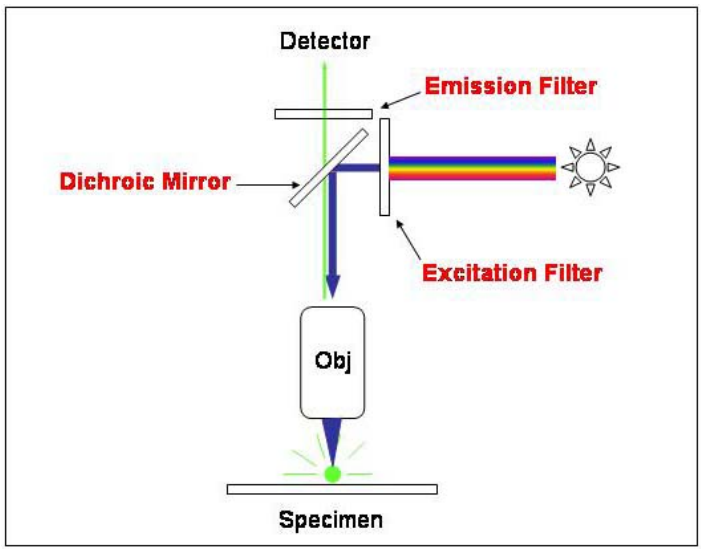


Figure 1.5 Schematic representation of a fluorescent microscope (from the science education resource center at Carleton college).

Typical components of a fluorescence microscope are the light source, the excitation filter, the diachroic mirror and the emission filter (see Figure 1.5). The filters and the mirror are chosen to match the spectral excitation and emission characteristics of the fluorophore used to label the specimen. Therefore, a single type of fluorophore (color) is imaged at a time. Multi-color images of several fluorophores can be composed by combining several single-color images. Most fluorescence microscopes observe epifluorescence i.e. the excitation and observation of the fluorescence are from above the specimen.

In conventional fluorescence microscopy the whole specimen is illuminated and all parts of the specimen throughout the optical path will be excited. In contrast, confocal microscopy uses point illumination and a pinhole in the optically conjugate plane in front of the detector to eliminate out-of-focus information. Only the light within the focal plane can be detected. The practical effect of this is that the image comes from a thin section of the sample (there is a small depth of field). By scanning many thin sections, one can build up a three dimensional image of the sample.

1.7 Atomic Force Microscopy

Unlike traditional microscopes, the atomic force microscopy (AFM) does not rely on electromagnetic radiation or an electron beam to create an image. An AFM uses opto-mechanical imaging to measure the three dimensional topography as well as the physical properties of a surface with a sharpened probe. The sharpened probe is positioned close enough to the surface so that it can interact with the force field associated with the surface. Then the probe is scanned across the surface such that the forces between the probe and the surface remain constant. An image of the surface is then reconstructed by monitoring the precise motion of the probe during the scan. The force sensor in an atomic force microscope is typically constructed from a light lever. In the light lever, the output from a laser is focused on the back of a cantilever and reflected into a photo-detector with four sections. When the probe at the end of the cantilever interacts with the surface, the cantilever bends and the light part changes causing the direction of light in the photo-detector section to change. The electronic output of the light lever force sensor is related to the force between the probe and the sample. There are many variants of AFM, depending on the nature and the way in which the force between the mobile probe and the surface of the sample is measured.

1.8 Transmission Electron Microscopy

1.8.1 Interactions of electrons with matter

Electrons are one type of "ionizing radiation" or in other words radiation that is capable of removing one of the tightly bound inner-shell electrons from the attractive field of the nucleus.

The advantage of using ionizing radiation is that it produces a wide range of secondary signals from the specimen (schematically represented in Figure 1.6).

Elastically scattered electrons participate in the formation of the TEM image and the electron diffraction pattern and are used to obtain structural information.

Inelastic scattering generates a whole range of signals. Many of those signals are used for analytical analysis, giving chemical information about the specimen. This information is used in analytical techniques such as X-ray energy dispersive spectrometry (XEDS) and electron energy loss spectroscopy (EELS). The most important signals are the characteristic X-rays, the inelastically scattered electrons themselves and the secondary electrons. The X-rays characteristic emission is used to define the elemental composition of the specimen as well as to quantify the amount of each element. X-rays are generated when a high energy beam electron penetrates through the outer electron shell and interacts with the inner core (shell) electrons. The inelastically scattered electrons contribute to the formation of the energy-loss spectra. The secondary electrons are considered in relation to SEM, where they are used to form the images which are very sensitive to surface topography.

1.8.2 Elastic scattering and diffraction

Electrons going through a thin specimen are either scattered or not scattered, and either lose energy or don't lose energy. Elastic scattering can occur in one of two ways, both of which involve Coulomb forces. As shown in Figure 1.7 the electron may interact with the electron cloud, resulting in a small angular deviation. Alternatively, if the electron penetrates the electron cloud and approaches the nucleus, it will be strongly attracted and maybe scattered through a larger angle.

Many electron-electron interactions are inelastic. The nuclear interaction may result in the generation of bremsstrahlung X-rays (if the electron is decelerated by the Coulomb field of the nucleus, it emits an X-ray). Since the electron can suffer any amount of deceleration depending on the strength of the interaction, these X-rays can have any energy up to the beam energy, or the nuclear interaction may even result in the displacement of the atom from its site in the crystal, both of which involve some energy loss for the electron. The higher the angle of scattering of an electron emerging from the specimen, the greater the chance that it will have undergone an inelastic event during its passage through the specimen.

Electron diffraction is by far one the most important scattering phenomena in the TEM. Diffraction can be used to determinate the spacing of planes in crystals. The interplanar spacings in different crystal structures are characteristic of their space group and unit cell. As a result we can distinguish between different crystal forms by observing and measuring diffraction patterns. The positions of the diffracted beams of electrons are determined by the size and shape of the unit cell, and the intensities of the diffracted beams are governed by the distribution, number and type of atoms in the specimen.

For an amorphous specimen, the atoms are almost randomly arranged. A random arrangement would result in a similar plot as Figure 1.8 but there are certain interatomic spacings that tend to occur in an amorphous structure.

As a result the amplitude and the intensity of diffraction is stronger at some angles than others which is displayed as rings. If the specimen is crystalline, the intensity of the diffracted beams is a maximum at specific angles because the interplanar spacings are fixed (see Figure 1.8).

Bragg showed that waves reflected off adjacent scattering centres must have a path difference equal to an integral number of wavelengths if they are to remain in phase.

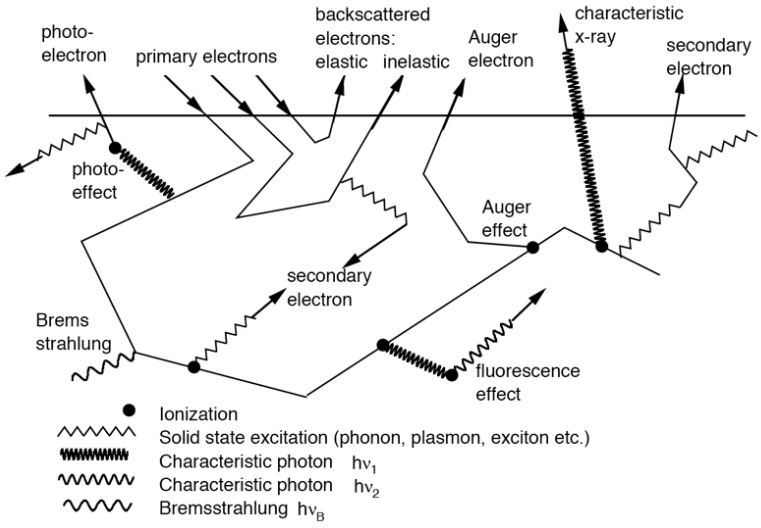


Figure 1.6 Different kinds of electron scattering.

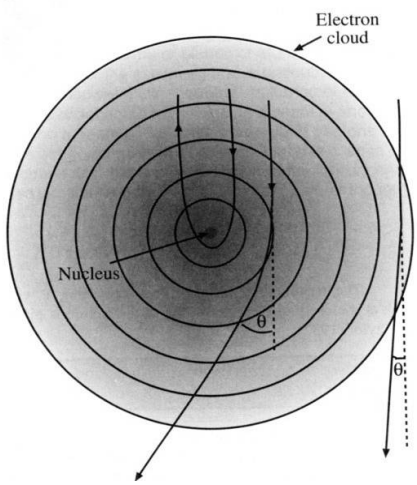


Figure 1.7 Coulombic interaction within the electron cloud results in low angle (θ) scatter while Coulombic attraction by the nucleus causes high θ scatter and perhaps complete backscatter. (From Williams & Carter, 1996)

The path difference between electron waves reflected from the upper and lower planes in Figure 1.9 is (AB+BC). Thus, if the "reflecting" *hkl* planes are spaced a distance *d* apart and the wave is incident and reflected at an angle θ, both AB and BC are equal to *d* sin θ and the total path difference is 2*d* sin θ. The Bragg's law is nλ=2*d* sin θ (see Figure 1.9)

Electrons, which come from the condenser system of the TEM (above the specimen), are scattered by the sample. Electrons scattered in the same direction are focused in the back focal plane, and as a result a diffraction pattern is formed there. Electrons coming from the same point of the object are focused in the image plane. In the TEM, the first intermediate image is magnified by further lenses (projective system).

The objective lens forms a diffraction pattern in the back focal plane with electrons scattered by the sample and combines them to generate an image in the image plane (intermediate image). Thus, diffraction pattern and image are simultaneously present in the TEM. It depends on the intermediate lens which of them appears in the plane of the second intermediate image and magnified by the projective lens on the viewing screen. Switching from real space (image) to reciprocal space (diffraction pattern) is easily achieved by changing the strength of the intermediate lens.

In image mode, an objective aperture can be inserted in the back focal plane to select one or more beams that contribute to the final image (BF, DF, HRTEM). In selected area electron diffraction (SAED), an aperture in the plane of the first intermediate image defines the region of which the diffraction is obtained. A schematic representation of the formation of an image and a diffraction pattern is given in Figure 1.10.

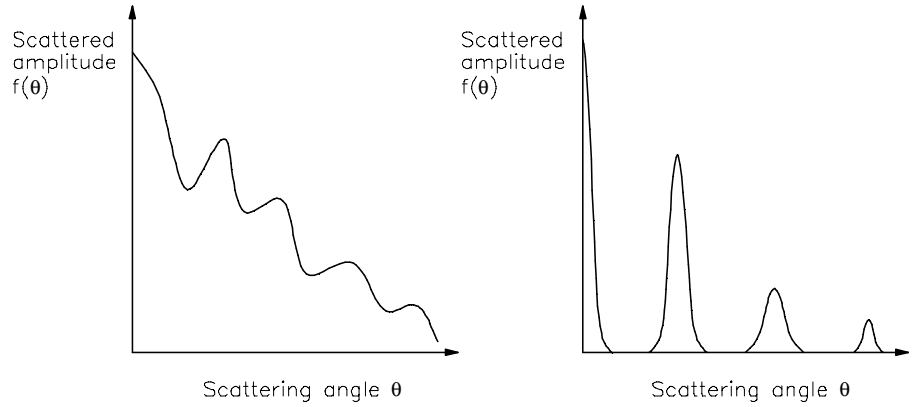


Figure 1.8 Change of f(θ) with θ for an amorphous material (left graph) and crystalline specimen (right graph). (From Williams & Carter, 1996)

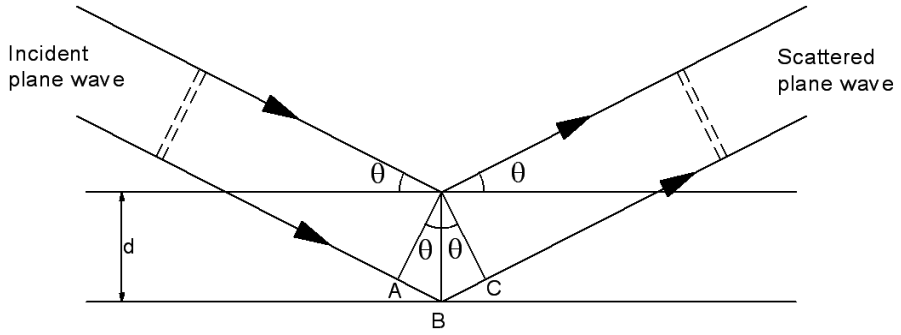


Figure 1.9 The Bragg description of diffraction in terms of the reflection of a plane wave incident at an angle θ to atomic planes of spacing *d*. (From Williams & Carter, 1996)

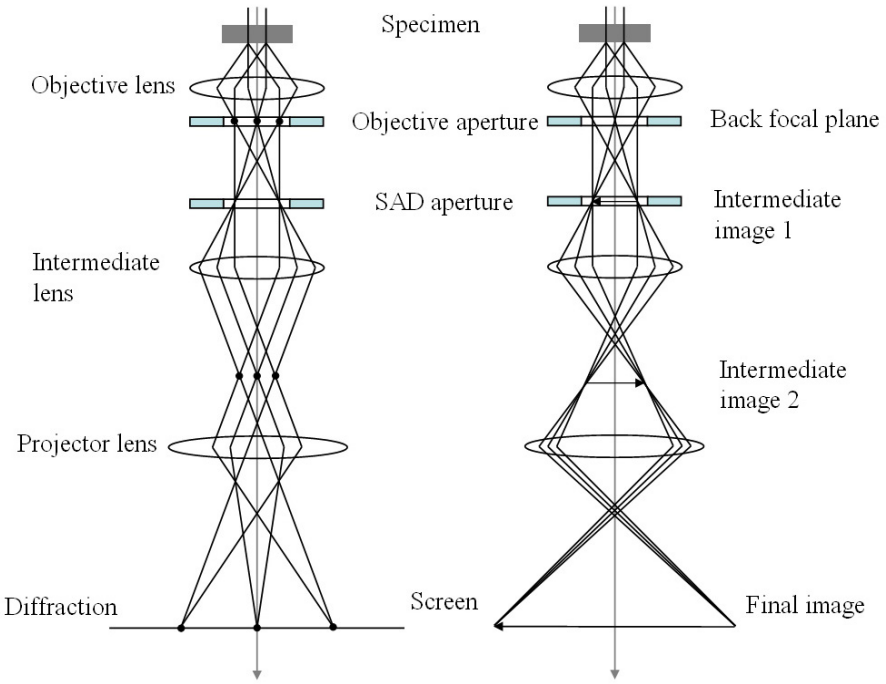


Figure 1.10 Formation of image and diffraction pattern in a TEM. (From Williams & Carter, 1996)

1.8.3 Formation of diffraction patterns in the TEM

The diffraction pattern contains electrons from the whole area of the specimen illuminated. However, in most of the cases only diffraction from a certain area is required. There are two ways to reduce the illuminated area of the specimen contributing to the diffraction pattern. One approach is to use a selected area aperture inserted above the specimen. In this case only electrons which pass through it will hit the specimen. A different approach to restrict the area is by making the beam smaller. The first approach allows semi-parallel illumination leading to the formation of spots. Converging the beam destroys any parallelism, and the spots on the pattern are not defined but spread into discs.

However, it is not possible to insert an aperture at the specimen plane (because the specimen is there). Therefore, an aperture is inserted in a plane conjugate with the specimen in one of the image planes, then it creates a virtual aperture at the plane of the specimen. Any electron that hits the specimen outside the area defined by the virtual aperture will hit the real diaphragm when it travels to the image plane and it will be thus excluded from contributing to the diffraction pattern.

It is also possible to obtain micro- or nanodiffraction patterns in TEMs. In this case first a fine probe by using a small condenser aperture is created and then the probe is focused onto the specimen (similar to convergent beam techniques). However, since a much smaller condenser aperture is used in this case (10 or 30μm) compared to convergent beam diffraction (where a condenser aperture of 150 μm is used), the illumination in micro- or nanodiffraction mode is more parallel.

References

1. Dubochet, J., Adrian, M., Chang J., De Homo, J.C., McDowall, W. and Shultz, P. (1988). *Quart. Rev. Biophys.* **21**, 129.
2. Faruqi, A.R. and Subramaniam, S. (2000). *Quart. Rev. Biophys.* **33**, 1.
3. Decruix, A. and Giege, R. (1999). Crystallization of nucleic acids and proteins. *A practical approach, Oxford University Press.*
4. Vekilov, P.G. (1993). *Prog. Cryst. Growth*, **26,** 25.
5. Stevens, R.C. (2000). *Curr. Opin. Struct. Biolog.* **10**, 558.
6. Rupp, B. (2003). *The TB Structural Genomics Consortium Crystallization Facility. Acc. Chem. Res*. **36**, 173.
7. Bundell, T.L. and Patel, S. (2004). *Curr. Opin. Pharmacol*. **4**, 490.
8. Chayen, N.E. and Saridakis, E. (2002). *Acta Cryst.* D**58**, 921.
9. De Gennes, P.G. (1985). *Rev. Mod. Physics,* **57**, 827.
10. Blokhuis, E.M. and Hartland, S. (2004). Surface and Interfacial tension measurements, theory and applications*. Dekker, New York,* 149.
11. Lorber, B., Jenner, G. and Giege, R. (1996). *J. Cryst. Growth*, **158**, 103.
12. Tomos, A.D. and Leigh, R.A. (1999). *Annu. Rev. Plant. Physiol*. **50**, 447.
13. Bodenstaff, E.R., Hoedemaeker, F.J., Kuil, M.E., de Vrind, H.P.M and Abrahams, J.P. (2002). *Acta Cryst.* D**58**, 1901.
14. Chernov, A.A. (2003). *J. Struct. Biol.* **142**, 3.
15. Vekilov, P.G. and Galkin, O. (2003). *Colloids Surf.* A**215**, 125.
16. ten Wolde, P.R. and Frenkel, D. (1997). *Science*, **277**, 1975.
17. Galkin, O. and Vekilov, P.G. (1999). *J. Phys. Chem.* B**103**, 10965.
18. Pusey, M.L. and Nadarajah, A. (2002). *Cryst. Growth. Des*. **2**, 475.
19. Mc Pherson, A., Kuznetsov, Y.G., Malkin, A. and Plomp, M. (2003). *Acta Cryst.* D57, 1053.
20. Yau, S.T. and Vekilov, P.G. (2001). *J. Am. Chem. Soc.* **123**, 1080.

21. Chayen, N.E., Saridakis, E., El Bahar, R. and Nemirovski, Y. (2001). *J. Mol. Biol*. **313**, 591.
22. Pechkova, E. and Nikolini, C. (2001). *J. Cryst. Growth*, **231**, 599.
23. Knight, J. (2002). *Nature,* **418**, 474.
24. Nijman, E.J., Merkus, H.G., Marijnissen, J.C.M. and Scarlett, B. (2001). *Appl. Opt.* **40**, 4058.
25. Hansen, C.L., Skordalakes, E., Berger, J.M. and Quake, S.R. (2002). *Proc. Natl. Acad. Sci. USA,* **99**, 16531.
26. Zheng B., Tice J.D., Roach L.S. and Ismagilov, R.F. (2004). *Angew Chem Int. Ed.* **53**, 2508.

Chapter 2

Heterogeneous nucleation of 3D protein nano-crystals

Adapted from: Georgieva, D.G., Kuil, M.E., Oosterkamp, T.H., Zandbergen, H.W. and Abrahams, J.P. (2007). Heterogeneous crystallization of protein nanocrystals. *Acta Cryst.* D63*, 564.*

Abstract

Nucleation is the rate-limiting step in protein crystallization. Introducing heterogeneous substrates may in some cases lower the energy barrier of nucleation and thereby facilitate crystal growth. So far the mechanism of heterogeneous protein nucleation remains poorly understood. In this study, the nucleating properties of fragments of human hair have been investigated. The four proteins that were tested - lysozyme, glucose isomerase, a polysaccharide specific Fab fragment and potato serine protease inhibitor nucleated preferentially on the hair surface. Macro- as well as showers of tiny crystals of a few hundred nanometre thickness were obtained also under conditions that did not produce crystals in the absence of the nucleating agent. The mechanism of heterogeneous nucleation was studied by using confocal fluorescence microscopy, demonstrating that the protein is concentrated on the nucleating surface. Substantial accumulation of protein was observed on the sharp edges of the hair's cuticles, explaining the strong nucleating activity of the surface.

2.1 Introduction

An essential part of most protein crystallographic studies is finding suitable conditions for growing crystals, which is often the rate-limiting step. To predict protein nucleation and crystallization, techniques such as dynamic light scattering [1, 2] and fluorescence correlation spectroscopy [3] have been applied.

Most of the studies based on these techniques showed that the second viral coefficient B_{22} of dilute protein solution is closely related to protein crystallization [2, 4, 5]. Only a narrow range of slightly negative B_{22} values is favourable for crystallization. Working under conditions outside this range reduces the possibility for a successful outcome. However, using conditions which correspond to the so called "crystallization window" does not guarantee a successful crystallization trial. Thus, finding a crystallization condition remains a process of trial and error.

Progress in miniaturization and automation of crystallization experiments led to the development of protein nano-crystallization, which made it possible to set up thousands of crystallization trials in a single experiment. Despite the large number of proteins and screening conditions that have been tested, the success rate is lower than expected [6]. It appears that nano-crystallization is not a simple miniaturization of a protein crystallization experiment and one cannot reduce the crystallization volume without paying a penalty. In nano-volumes the surface tension forces become more important, affecting possible nucleation events. Furthermore, Bodenstaff *et al*. [7] showed that the mean number of nuclei formed per unit volume is linearly proportional to the total volume of the mother liquor present in the experiment. Moreover, when working in a nano-litre regime, the time before the first nuclei are formed increases dramatically. Usually, this is on top of already poor crystallization even in larger volumes.

It is known that the protein should be in the metastable phase for crystal growth, but higher levels of saturation are needed for nucleation. In many crystallization experiments the required saturation levels are not reached, so that nucleation does not occur. To create an environment favouring nucleation, so-called nucleant agents are introduced in the crystallization droplet which locally create a higher concentration of macromolecules, thus lowering the energy barrier for nucleation. A search for a “universal nucleant” has been ongoing for two decades. So far the following lines of research have been pursued.

(i) McPherson introduced the idea of controlling nucleation by using mineral substrates as epitaxial nucleants for protein crystallization [8]. His initiative has been pursued for more than 15 years, a variety of substrates have been employed, but so far none of them was generally adopted.

(ii) Later on the idea of using lipid layers and protein monolayers of 2D crystals was introduced, which also improved the crystallization of 3D protein crystals [9]. More

recently Fermani *et al*. [10] demonstrated that substrates containing ionisable surface groups can also enhance crystallization of certain proteins.

(iii) The idea of using natural seedlings (whiskers, seeds, fibres etc.) to generate nucleation in crystallization experiments is another approach which has been successfully applied in protein crystallography [11].

(iv) The development of lithography techniques allowed the fabrication of variety of Si substrates with different surface characterizations – terraces, steps, even pores of a few nm sizes which can match a crystal lattice. Sanjoh *et al*. [12] and Chayen *et al*. [13] explored this field intensively and showed that, in general, structured surfaces appear to be more efficient than non-structured.

We decided to search for materials which combine, as well as is possible, most of the properties mentioned above – surface ordered at the molecular level, ionisable groups, lipid layers, local concentration cavities, nano- and mesoscopic structure. We found that a prime candidate for such a material is abundantly available: the surface of human hair matches the above criteria quite well. We observed that not only standard proteins such as lysozyme and glucose isomerase, but also more difficult proteins like a polysaccharide specific Fab fragment and a potato protease inhibitor under study in our lab, crystallized preferentially on the strands of hair.

By using a combination of advanced visualization techniques such as fluorescence confocal microscopy and atomic force microscopy (AFM), it was possible to visualize the distribution of protein on the surface of the hair and demonstrate accumulation of protein on the sharp edges of the hair's cuticles. This experimental observation correlated with numerical simulations published by Cacciuto *et al*. [14], which showed nucleation and crystallization of a model colloid to occur preferentially on curved surfaces.

2.2 Experimental procedures

2.2.1 Materials

Commercial proteins used in this study were chicken egg-white lysozyme (Sigma, EC 3.2.1.17), glucose isomerase (Hampton Research cat. No. HR7-100). Antipolymeric Lewis X Fab fragment 54 was expressed by papain digestion of MAP 54-5C10-A

followed by extensive purification, using affinity and ion-exchange chromatography as described previously by van Roon *et al*. [15]. Potato serine protease inhibitor 6.1 (the number representing the isolectric point) was expressed and purified following the procedure given by Thomassen *et al*. [16]. All the chemicals used for the crystallization experiments were purchased from Merck and solutions were further filtered with Millipore filters (0.22 μm) directly before use. Lysozyme was labeled with fluorescein iso-thiocyanate (Isomer I, Molecular Probes, Eugene, Oregon USA) for confocal fluorescence studies. Water clear urethane rubber (Clear Flex 50, Smooth-On, Inc., Pennsylvania USA) and polydimethylsiloxane (Sylgard R 184 curing agent silicone elastomer, Dow Corning Corporation, Michigan USA) were used for the preparation of the polymer hair replica. Dark pigmented human hair fibres were used as heterogeneous nucleant surfaces.

2.2.2 Crystallization experiments

Crystallization trials were carried out at 20°C using the sitting drop vapour diffusion technique in Q-crystallization plates (Hampton Research). Portions of 1 μL protein were added to 1 μL of reservoir solution in a sitting drop. For the heterogeneous crystallization of lysozyme strands of hair were introduced in crystallization droplets containing 7.5 mg/ml lysozyme, 0.1 *M* acetate buffer (pH 4.5), 1.6 *M* NaCl and in conditions containing 7.5 mg/ml lysozyme, 0.1 *M* acetate buffer (pH 4.5), 30% glycerol and a salt concentration varying between 0.65 - 1.6 *M* NaCl.

Glucose isomerase was crystallized in the presence of 2 *M* ammonium sulphate and 0.1 *M* sodium citrate (pH 6.5). A final protein concentration of 15 mg/ml was used for the heterogeneous crystallization. Antipolymeric Lewis X Fab fragment 54 was crystallized in 100 m*M* citrate buffer (pH 5) and 11% PEG 3350. Potato serine protease inhibitor was crystallized in 0.1 *M* HEPES (pH 7.5), 5% PEG8000 and 4% ethylene glycol complemented with 0.1 *M* glycine (Hampton Additive screen 2). The protein concentration was 7 mg/ml.

2.2.3 Chemical modification of the hair surface

Removal of the lipids was done by soaking in petroleum ether. Single hairs were treated with petroleum ether for 30 minutes, 3 times for 10 minutes each time with fresh portions of the solvent. The remaining petroleum ether was allowed to evaporate

and the fibres were washed out with the buffer used in the crystallization experiment. A combination of delipifying and a slight denaturatum of the keratin surface was done with 50% ethanol solution. The same treatment was performed as described for the petroleum ether. For denaturation of the surface proteins, the hair was subjected to 3 *M* NaOH for 3 minutes and washed with buffer as in the previous two treatments.

2.2.4 Preparation of polymer hair replica's

A hair fibre cleaned with petroleum ether from dust and impurities as described in section 2.2.3 was placed on a microscope glass slide and glued to the slide at both ends. Polyurethane (Clear Flex 50) was poured on the fibre to form a thin layer on the microscope slide. The polyurethane rubber was left overnight at 70ºC for complete polymerization. Then the polymer layer was detached from the cover slide and the replica of the fibre formed in the polyurethane mold was fillled with a second polymer - polydimethylsiloxane and left for 24 h at 70ºC. After that, the polydimethylsiloxane fibre was removed from the polyurethane rubber mold and further crystallization experiments were performed.

2.2.5 Atomic Force Microscopy studies

In situ crystallization experiments of lysozyme were set up in a liquid AFM cell. The crystallization composition was the same as described in section 2.2.2. The hair fibre used in this experiment was treated with petroleum ether (section 2.2.3) in order to remove impurities from the fibre which might retract the AFM tip. Petri dishes with water were placed in the AFM to prevent the sample droplet from drying out during imaging. The experiment was performed with a Digital Instruments Nanoscope IIIa scanning-probe. Silicon nitride tips were used throughout in tapping and contact mode with scan frequency varying from 1 to 10 Hz.

2.2.6 Confocal fluorescence experiments

10 mg/ml lysozyme dissolved in 10 m*M* carbonate/bicarbonate buffer (pH 9.2) was mixed with 1 mg/ml fluorescein isothiocyanate in dimethylformamide in a ratio of (1:0.65). The mixture was wrapped in Al-foil and incubated on a rotor shaker at low speed at room temperature. To separate the free dye from the labeled protein, a desalting gel-filtration column PD-10 (Sephadex TM G-25M prepacked column) was

used. The column was first equilibrated with milliQ water, then loaded with 825 μL reaction mixture and eluted with milliQ water. Fractions of 1 ml were collected. The labeled protein eluted in fractions 4-7. After gel electrophoresis on a 15% PAGE-gel of the collected fractions, in fractions 4-7 fluorescent bands were clearly visible. After Coomassie staining, these bands corresponded to protein of MW 14.5 kD, as expected for labeled lysozyme.

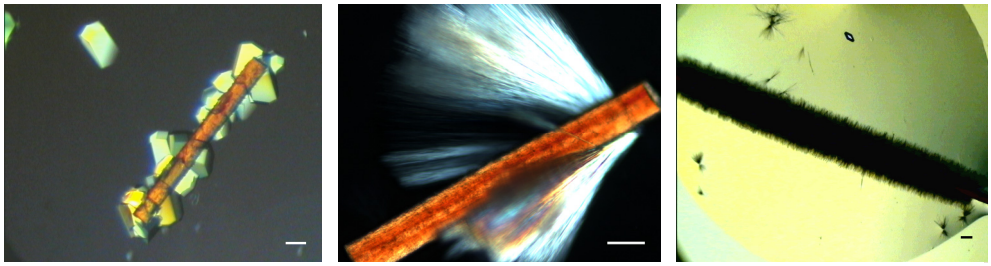
Strands of hair were introduced in droplets containing 0.1 *M* acetate buffer (pH 4.5) 1.6 *M* NaCl and protein concentrations of 2.5, 1, and 0.5 mg/ml labeled protein and incubated for 4 h at room temperature. The fibres were then removed from the original droplets and placed in new ones containing the same crystallization agents but now without protein. This was done in order to enhance visibility of bound fluorescent protein and to reduce the background fluorescence due to unbound fluorescent protein. Only in this way differences in distribution of protein on and near the hair strands could be studied using confocal microscopy. For imaging, we used an upright Zeiss Axioplan epifluorescence microscope and a confocal inverted Leica IRBE microscope coupled to a SP1 scanhead with a separate Argon and Krypton laser. The Argon laser was used for excitation at 488 nm the Krypton for excitation at 568 nm.

2.3 Results and discussion

2.3.1 Examples of heterogeneous crystallization on the surface of human hair

Chicken egg-white lysozyme was used as a model protein for our initial studies. In the presence of hair strands, introduced in droplets containing 7.5 mg/ml lysozyme, 0.1 *M* acetate buffer (pH 4.5) and 1.6 *M* NaCl, it was observed that lysozyme has a clear tendency to crystallize on the fibres. Figure 2.1 shows three morphological crystal forms of the protein to nucleate preferentially on the selected heterogeneous substrates. The induction time for nucleation was also much shorter - crystals were observed to appear within 3 to 4 h on the hair strands and more than 10 h were needed to form crystals in bulk solution. In some cases co-existence of tetragonal and “sea-urchin” crystals on the same nucleant surface was observed (see Figure 2.2). Such co-existence of the crystal forms has been observed earlier in bulk solution [17]. A tetragonal

lysozyme crystal grown on hair was mounted in a capillary and X-ray diffraction confirmed the space group P 4_3 2_1 2 (data not shown).



(a) (b) (c)

Figure 2.1 Lysozyme crystals grown heterogeneously on hair fibres in crystallization conditions containing 0.1 *M* sodium acetate (pH 4.5), 1.6 *M* sodium chloride and 7.5 mg/ml protein: (a) tetragonal, (b) needle like and (c) "sea-urchin" crystal forms. The bars correspond to 100 μm.

In our further experiments we consistently reproduced the crystallization of the macro and nano "sea-urchin" crystals (see Figure 2.3) on the selected nucleant surface. In the presence of hair fibres crystals were obtained also in conditions in which nucleation does not occur or occurs very rarely. For example, inclusion of 30% glycerol prevents nucleation of lysozyme in the bulk at protein concentration of 7.5 mg/ml. However, when hair strands were introduced into a solution like that, we observed lysozyme crystals nucleating and growing on the hair in a range of salt concentration between 0.65 - 1.6 *M* NaCl, confirming the strong nucleating properties of this surface.

Our studies on heterogeneous crystallization were extended further to other proteins that are more difficult to crystallize. Of special interest to us was the potato protease inhibitor (6.1) as this protein is not easy to crystallize and the structure is still not resolved by X-ray crystallography [16].

The crystals are difficult to manipulate and all attempts to soak them with heavy metals in order to get the phase information of some reflections have been unsuccessful so far. Moreover, crystals tend to intergrow, making the formation of large single crystals suitable for X-ray diffraction rather difficult. Potato protease inhibitor was observed to crystallize heterogeneously readily on the hair fibres, as shown in Figure 2.4.

Moreover, we could easily select individual crystals. The size of the crystals showed that they are suitable for electron microscopy studies. Also with glucose isomerase and Fab fragment 54 needle like crystals on the hair surface were obtained. In the case of glucose isomerase we observed also bulk crystals usually appearing at the extremities of the fibres as shown in Figure 2.4.

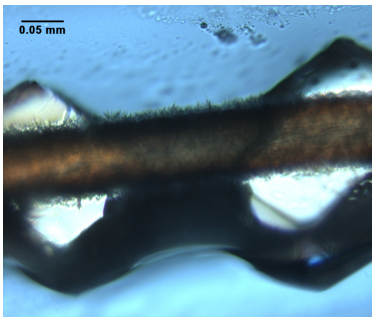


Figure 2.2 Co-existence of two different crystal forms - tetragonal (macro crystals) and sea-urchins (small crystals covering the surface) on the same heterogeneous substrate.

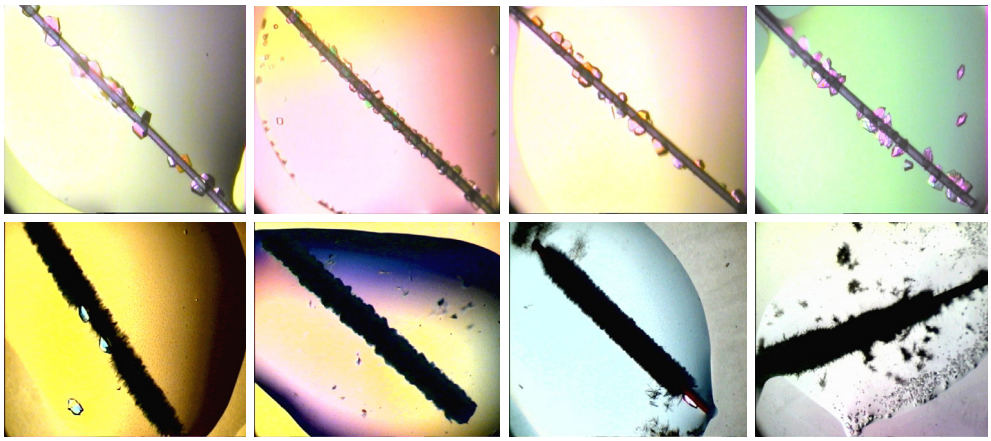


Figure 2.3 Gallery of lysozyme macro (first row of images) and nano-crystals (second row of images) grown heterogeneously on hair fibers.

In the presence of hair strands glucose isomerase was crystallized at two times lower protein concentration (15 mg/ml) than needed for homogeneous crystallization of the protein in bulk solution (30 mg/ml). When hair fibres were introduced in crystallization droplets of Fab fragment and potato protease inhibitor, both proteins were observed to nucleate preferentially on the surface of the hair. In the case of Fab fragment crystals were observed to appear in bulk solution as well, although less as a number. Potato protease inhibitor crystals were obtained also in droplets in which precipitation of the protein was added.

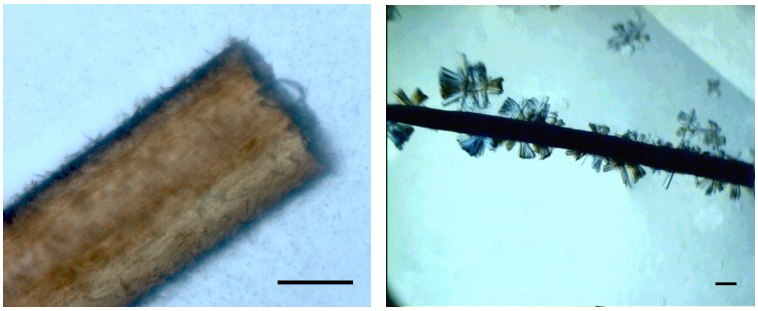


Figure 2.4 Examples of protein crystals grown heterogeneously on strands of hair: hair surface covered with showers of micron-sized crystals and one macro crystal of glucose isomerase - (left image). The protein was crystallized in presence of 2 *M* ammonium sulphate and 0.1 *M* sodium citrate (pH 6.5), bar 50 μm. Potato serine protease inhibitor crystals originating from heterogeneous nucleation on the hair surface - (right image), bar 100 μm.

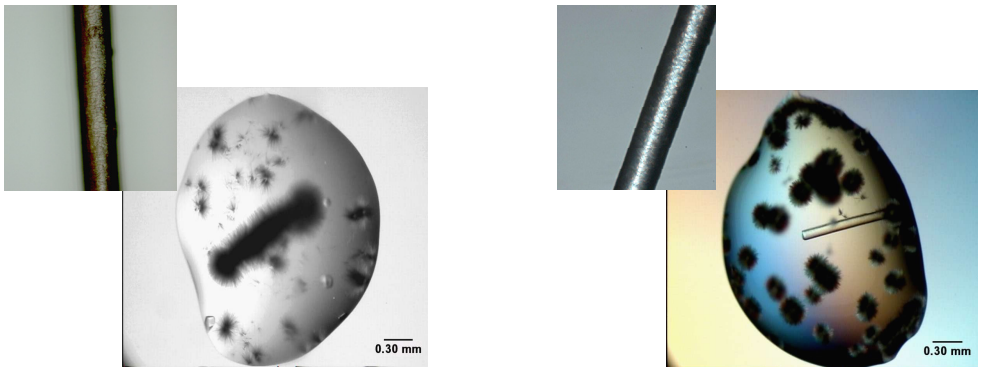
2.3.2 The mechanism of heterogeneous nucleation

Our experiments show that the surface of the hair is an effective nucleant for proteins. In order to understand this, we further investigated the effects of various chemical and morphological properties of the hair surface.

Keratins are the most predominant proteins of hair. They are ordered and may provide a semi-crystalline interface at the surface. Their pI varies from 4.7-8.5, which provides some buffering properties to the system. Moreover, the surface of the hair is structured: regularly repeating overlapping terraces of different size and depth can be recognized

in scanning electron micrographs as shown in Figure 2.6 (a). In order to understand the mechanism of the observed heterogeneous crystallization, all previously discussed properties needed to be excluded one by one. For this purpose, the fibres were treated with different chemicals – petroleum ether, ethanol, sodium hydroxide. After washing with mother liquor, they were introduced in crystallization droplets. Petroleum ether is well known as de-lipidifying agent and is used routinely in phytochemistry. Ethanol is not only de-lipidifying but also a moderately denaturing agent for proteins. Sodium hydroxide is a severe denaturing agent for proteins in the concentration used. After 24 h, no crystallization was observed on the fibres treated with sodium hydroxide. Only small "sea-urchin" crystallites were noticed on the hair treated with ethanol while both macro crystals and sea-urchin crystals were found on the fibres treated with petroleum ether only. All the experiments were performed three times and the same results were obtained. This suggested that removal of the lipids from the hair surface does not essentially influence the crystallization behavior. Full loss of nucleation, inferred from the absence of crystals, was observed when the surface keratins were denatured and most probably also the terraced structures were etched away by sodium hydroxide. However, based on this experiment we could not distinguish between the importance of the chemical role of the surface proteins and the structure that they form: the effect of the surface structure and the role of the overlapping terraces, if any, remained unclear. We made a polymer replica of a hair, including its terraced surface, and set up a series of crystallization trials. Protein crystals were observed to grow occasionally on the polymer substrates (see Figure 2.5), but there was no clear preference contrary to the crystallization droplets with natural hair fibres. We concluded that the presence of keratin seems to be essential for nucleation.

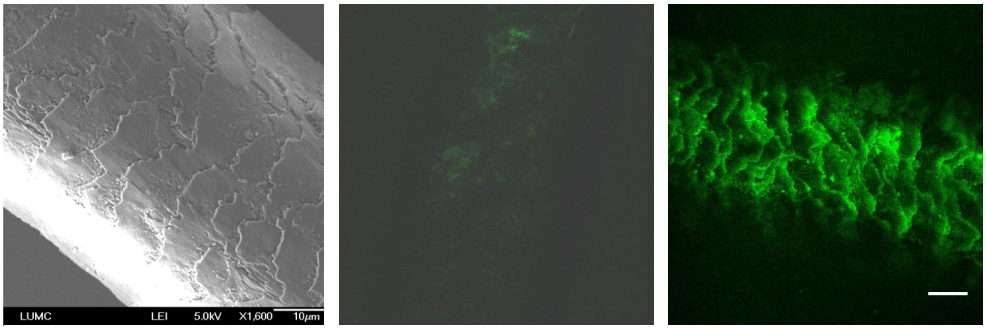
To visualize the distribution of lysozyme on the hair surface, fluorescence studies were performed with labeled protein. In order to distinguish between the protein of the hair fibres that fluoresce mostly in the red and the lysozyme (non fluorescent in the visible spectrum), the latter was labeled with green fluorescing dye. We verified that this modification did not influence crystallization (data not shown). At a protein concentration of 7.5 mg/ml labeled protein, heterogeneous crystallization could be followed, but the overall fluorescent signal was too strong to differentiate concentration differences next to, or on the fibres. However, at a concentration of 2.5 mg/ml we observed a non uniform distribution of lysozyme on the hair surface. At 1.0 and 0.5 mg/ml protein concentration we could clearly visualize that there is a much higher protein accumulation on the edges of the terraces (see Figure 2.6).



(a) (b)

Figure 2.5 Polymer and natural hair fibres as nucleant agents: sea-urchin lysozyme crystals show preference for nucleation on natural fibres - (a), but not on polymer fibres - (b).

When fibres were pre-incubated overnight with 1 mg/ml non fluorescent lysozyme in crystallization conditions (0.1 *M* acetate buffer (pH 4.5), 1.6 *M* NaCl) and after that 0.5 mg/ml labeled protein was added, green signal was detected on the edges of the cuticles, suggesting that the protein binding is reversible in view of the apparent exchange of fluorescent versus non - fluorescent protein.



(a) (b) (c)

Figure 2.6 Visualization of protein distribution by confocal fluorescent microscopy on the surface of human hair fibres: (a) scanning electron micrograph of a human hair fibre,

showing regularly repeated terraces on the surface, bar 10 μm, (b) distribution of protein on the surface of a hair fibre in crystallization droplets containing 1.6 *M* sodium chloride and 0.1 *M* sodium acetate (pH 4.5) at protein concentration of 0.05 mg/ml and (c) 0.5 mg/ml fluorescently labeled lysozyme. The green signal on the confocal photograph indicates presence of fluorescent protein. The stronger the signal is, the more fluorescent protein is accumulated. The dark parts on the photo indicate respectively absence of fluorescent protein. The length of the bar is 20 μm.

In dynamic AFM studies we followed the initial protein aggregation and crystal formation on the hair surface at a higher resolution. Imaging on the sharp edges was not possible as the height differences caused retraction of the AFM tip. In most of the cases, protein aggregates were observed to form on rougher parts or "irregularities" like clumps or scratches on the flat hair surface of the scanned areas (see Figure 2.7). When scanning was performed on relatively flat parts of the hair cuticles, no aggregation was observed for at least a few hours imaging. However, we also observed the scanning tip to interfere with crystallogenesis, possibly by disturbing pre-nuclei.

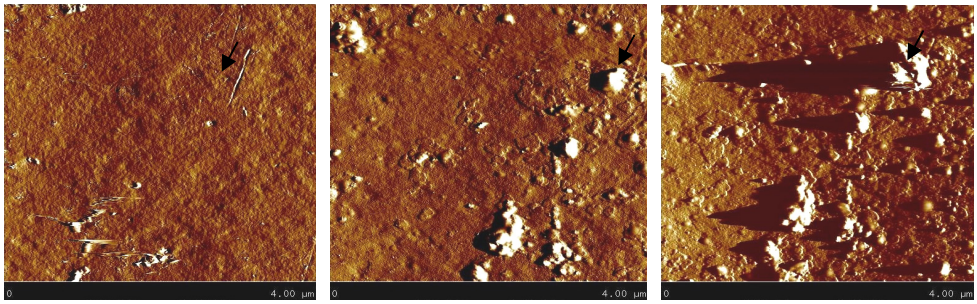


Figure 2.7 Following protein aggregation formation on the surface of a hair fibre. Aggregation of proteins was observed to occur mostly on surface areas with irregular landscape. With a black arrow it is shown a clump on which (in the vicinity of which) an aggregation was observed to occur. However, it has to be considered as well that in the AFM technique the scanning is done by using a tip which comes into contact with the surface (in this case the potential nucleus) and therefore the possibilities of destruction cannot be ignored.

2.4 Conclusions

We identified human hair to be a versatile nucleant surface and applied it successfully to the crystallization of various proteins. Moreover, despite the rather complex mechanism of heterogeneous crystallization, we discovered that the surface properties and the chemical composition together define the nucleation properties of the selected surface. No significant change in crystallization behavior was observed after modifying the natural hair fibres with petroleum ether while destroying the keratin with ethanol or sodium hydroxide caused partial and full loss of nucleation properties respectively. A polymer replica of a hair fibre introduced in crystallization experiment did not have the nucleant properties of the natural fibres. On the basis of these observations, we conclude that the native protein surface is vital for nucleation and that the lipid layer is not required. AFM imaging combined with confocal fluorescence studies of the surface of hair fibres in crystallization conditions indicated that the structured surface also plays an important role in the nucleation. Furthermore, we showed by confocal imaging that there is an uneven distribution of protein on the nucleant substrate in crystallization conditions which can explain the strong nucleating activity of hair fibres.

It has been observed, also by others, that protein crystals tend to appear on the edges of natural or engineered nucleant surfaces (see introduction (i) and (iii)). So far this phenomenon remained somewhat enigmatic. Here, we provide visual evidence for protein accumulation on the edges of such a nucleant substrate. According to the classical nucleation theory, fluctuations in protein concentration are the driving force for crystallization.

By using heterogeneous nucleants the high kinetic barrier of spontaneous nucleation can be bypassed. Still, most of the initial trials produce only nano or micro crystals that require further improvement in order to be used for X-ray studies. Optimization is difficult to automate since it must be adapted for each case individually. An alternative is to use electron sources in order to study sub-micron crystals.

Solving structures using electron diffraction data of 3D protein crystals is currently not yet feasible, but if certain technical obstacles can be overcome, it may provide an excellent alternative to X-ray diffraction for proteins that give only very small crystals.

References

1. Berne, J.B. and Pecora, R. (1976). *Dynamic light scattering.* New York: Willey and Sons.
2. George, A. and Wilson, W. (1994). *Acta Cryst*. D**50**, 361.
3. Schmauder, R., Schmidt, T., Abrahams, J.P. and Kuil, M.E. (2002). *Acta Cryst*. D**58**, 1536.
4. Velev, O.D., Kaler, E.W. and Lenhoff, A.M. (1998). *J. Biophys*. **75**, 2682.
5. Narayanan, J. and Liu, X.Y. (2003). *J. Biophys*. **84**, 523.
6. Dale, G.E., Oefner, C. and D'Arcy, A. (2003). *J. Struct. Biol*. **142**, 88.
7. Bodenstaff, E.R., Hoedemaeker, F.J., Kuil, M.E., de Vrind, H.P.M. and Abrahams, J.P. (2002). *Acta Cryst*. D**58**, 1901.
8. McPherson, A. and Shlichta, P.J. (1987). *J. Cryst. Growth*, **85**, 206.
9. Hemming, S.A., Bochkarev, A., Darst, S.A., Kornberg, R.D., Ala, P., Yang, D.S.C. and Edwards, A.M. (1995). *J. Mol. Biol*. **246**, 308.
10. Fermani, S., Falini, G., Minnucci, M. and Ripamonti, A. (2001). *J. Cryst. Growth*, **224**, 327.
11. D 'Arcy, A., Mac Sweeny, A. and Haber, A. (2003). *Acta Cryst*. D**59**, 1343.
12. Sanjoh, A., Tsukihara, T. and Gorti, S. (2001). *J. Cryst. Growth*. **232**, 618.
13. Chayen, N.E., Saridakis, E., El Bahar, R. and Nemirvsky, Y. (2001). *J. Mol. Biol*. **312**, 591.
14. Cacciuto, A., Auer, S. and Frenkel, D. (2004). *Nature*, **428**, 404.
15. van Roon, A.M., Pannu, N.S., Hokke, C.M., Deelder, A.M. and Abrahams, J.P. (2003). *Acta Cryst*. D**59**, 1306.
16. Thomassen, E.A., Pouvreau, L., Gruppen, H. and Abrahams, J.P. (2004). *Acta Cryst*. D**60**, 1464.
17. Lorber, B., Jenner, G. and Giege, R. (1996). *J. Cryst. Growth*, **158**, 103.

Chapter 3

Preparing 3D protein nano-crystals for electron diffraction studies

Abstract

Single crystal X-ray diffraction is still the leading technique for structure determination in the solid state. However, one of the major bottlenecks in general and particularly in protein crystallography is the limited availability of well diffracting macro-crystals. Often only micro- or nano-sized protein crystals grow and there are no standard protocols for optimizing such crystals (if they can be optimized at all). Since electrons interact much more strongly with matter than X-rays we investigated whether electron diffraction can be applied as an alternative technique for studying proteins. 3D nano-crystals of potato serine protease inhibitor and lysozyme, grown as described in Chapter 2, were used for testing. The diffraction signal obtained from nano-crystals of both proteins indicated that the crystals are well ordered and can be used for diffraction studies.

In this chapter we describe how to optimally prepare 3D protein nano-crystals for electron diffraction studies.

3.1 Introduction

Despite substantial progress in crystallization techniques, growing protein crystals remains a process of trial and error. Development of protein nano-crystallization methods allows thousands of crystallization trials to be set up in a single experiment. However, even if crystals are obtained in nano-volumes there is no guarantee that the crystallization experiment can be reproduced in larger volumes and that macro-crystals can be further grown [1].

Heterogeneous crystallization techniques relax to a certain extent the requirement for high protein concentrations needed to induce nucleation, facilitating crystallization in this respect. However, even using heterogeneous nucleants, crystal growth is not easily controlled. Often showers of crystals are obtained and those crystals need to be further optimized for X-ray studies. If optimization is not possible, X-ray diffraction techniques cannot be used for structure determination (nano-crystals are far too small to be studied even with synchrotron X-ray sources).

The development of synchrotron X-ray powder techniques relaxes the requirement for macroscopic single crystals in special cases [2]. Powder diffraction usually yields accurate unit cell parameters and in some cases it may allow a full structure determination. Unfortunately, the collapse of three-dimensional reciprocal space into a one-dimensional powder diffraction pattern leads to a significant loss of information. There is not only the usual phase problem, but due to significant peak overlap it is frequently not even possible to determine the intensity of the individual Bragg reflections, but only their sum. This has limited *ab initio* structure determination from powder diffraction to smaller molecules, though in favorable conditions protein structures have been determined with this technique. Another practical requirement which is also a severe limitation in the case of proteins is that milligram quantities of crystalline powder need to be available for successful structural studies.

Protein crystals absorb energy during data collection, which reduces their crystallinity and the resolution in a predictable fashion. The X-ray dose (energy per unit mass) a cryo-cooled crystal can absorb before the diffraction pattern decays to half of its original intensity is defined as 2×10^{7} J/kg (the "Henderson limit") [3]. Since modern synchrotron X-ray sources have sufficient brilliance to very quickly reach the Henderson limit, beyond which further data collection is no longer useful, radiation damage has become the second main limiting factor in X-ray crystallography of proteins, especially when only micro-crystals are available.

However, the statistics are much more favorable in the case of electrons. The main advantage of electrons is that for each elastic scattering event, accompanying inelastic event they deposit at least three orders of magnitude less energy in the sample, compared to X-rays [4]. Hence even sub-micron crystals can still yield useful diffraction data when electrons are used rather than X-rays. This explains the success of electron diffraction of 2D crystals of (membrane) proteins, which are far too thin for diffraction analysis by X-rays.
In view of the success of electrons in 2D protein crystallography, electron diffraction

should be an attractive alternative for 3D protein nano-crystals as well, since the high scattering cross-section of electrons allows very small crystalline volumes (more than 10^6 times smaller than those required for X-rays) to be studied. A comparison of the atomic cross-section for electrons, X-rays and neutrons as a function of wavelength is given in Figure 3.1. The ratio of inelastic to elastic scattering events is also more favorable in the case of electrons (see Table 3.1)

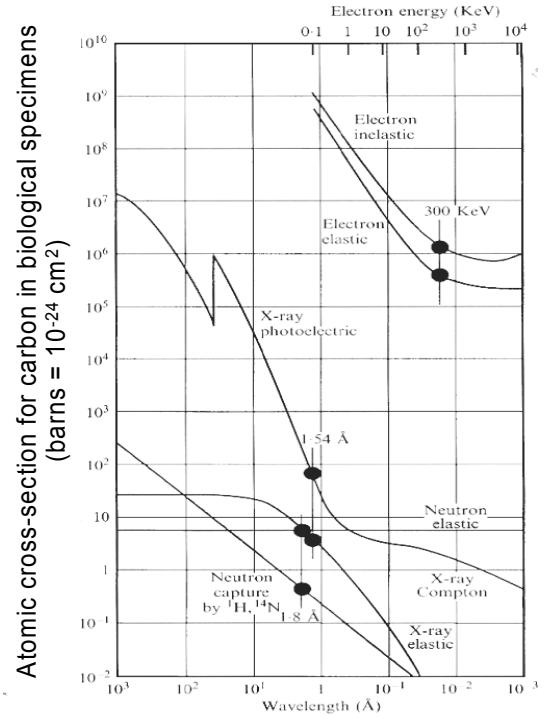


Figure 3.1 Comparison of the atomic cross-section for electrons, X-rays and neutrons as a function of wavelength. (Henderson, 1995)

Table 3.1 shows that relative to the number of elastic scattering events, electrons cause less radiation damage compared to X-rays [4]. Because of the lack of sufficiently bright neutron sources the use of neutrons for studying biological samples is not feasible yet. For X-rays, the amount of damage per useful elastic scattering event is several hundred times greater than for electrons at all wavelengths and energies. Consequently, the

requirements on specimen volume are correspondingly higher. Therefore, electrons provide at present the most information for a given amount of radiation damage.

Table 3.1 Comparison between the energy deposited per elastic and inelastic scattering by electrons, X-rays and neutrons. (From Henderson,1995)

	Electrons 80-500 keV	X-rays 1.5Å	Neutrons 1.8Å
Ratio (inelastic/elastic) scattering events	3	10	0.080
Energy deposited per inelastic event	20 eV	8 keV	2 keV
Energy deposited per elastic event	60 eV	80 keV	160eV
Energy deposited relative to electrons			
(inelastic)	1	400	100
(elastic)	1	1000	2.5

The development of cryo-electron microscopy was an important step, as it allowed at present thin biological specimens to be preserved in a frozen-hydrated state. While maintaining specimens at liquid nitrogen temperature or colder, the samples can be introduced into the high vacuum of the electron microscope column and observed at cryogenic conditions. The low temperature of cryo-electron microscopy provides also an additional protective factor against radiation damage [5].

Since protein crystals contain 50-70% water and need to remain intact, most of the cooling techniques, including high pressure or spray freezing, cannot be used. Because of the very small size of the crystals, the “fishing” techniques used in X-ray crystallography of selecting and mounting individual crystals for freezing, are anything but trivial. Immersion or plunge freezing therefore seems to be by far the best choice

for preservation of 3D nano-sized protein crystals. The higher thermal capacity of liquid ethane compared to liquid nitrogen, as well as the development of robotic systems for automated high-speed plunging of the specimen, can reproducibly lead to successful vitrification of three-dimensional biological samples.

A protocol for vitrification of thin 3D protein nano-crystals (in particular lysozyme) is presented in this chapter. Ice formation or secondary contamination of ice during the sample transfer into the microscope present additional problems and may impede successful cryo-electron experiments. An optimization of the cryo-transfer system is introduced and the advantages of the system over existing cryo-transfer systems are discussed. Initial diffraction experiments using selected area diffraction mode are also reported.

3.2 Experimental procedures

3.2.1 Materials

Commercial proteins used in this study were chicken egg-white lysozyme (Sigma, EC 3.2.1.17). Potato serine protease inhibitor 6.1 (the number representing the isolectric point) was expressed and purified following the procedure given by Thomassen *et al*., [6]. All the chemicals used for the crystallization experiments were purchased from Merck and solutions were further filtered with Millipore filters (0.22 μm) directly before use.

3.2.2 Crystallization experiments

Crystallization trials were carried out at 20°C using the sitting drop vapour diffusion technique in Q-crystallization plates (Hampton Research). Portions of 1 μL protein were added to 1 μL of reservoir solution in a sitting drop containing 7.5 mg/ml lysozyme, 0.1 *M* acetate buffer (pH 4.5), 1.6 *M* NaCl. Potato serine protease inhibitor was crystallized in 0.1 *M* HEPES (pH 7.5), 5% PEG8000 and 4% ethylene glycol complemented with 0.1 *M* glycine (Hampton Additive screen 2). The protein concentration was 7 mg/ml. Strands of hair were introduced in the crystallization droplets of both proteins, inducing nucleation and the growth of nano-crystals on its surface.

3.2.3 Sample preparation and preliminary electron diffraction of the nano-sized protein crystals

Chemical fixation

A hair fibre covered with tiny crystals of lysozyme or potato protease inhibitor was placed on a negatively charged 300 mesh continuous carbon grid. Next, a few microliter of the crystallization solution with the same composition as used for the crystal growth was pipetted onto the grid, resulting in detachment of most of the crystals. After that, the hair fibre was removed. The excess of buffer was blotted away with a filter paper and the crystals were negatively stained with 2% uranyl acetate.

Plunge freezing

Thin lysozyme crystals were grown as described in Chapter 2. For the freezing of the crystals three different cryo-solutions were prepared - one containing 1.6 *M* NaCl and 0.1 *M* Acetic buffer (pH 4.5) as in the mother liquor and the other two with the same composition, but additionally including 10% and 15% glycerol as cryo-protectant. The crystals were transferred on an electron microscopy grid as described in [7] and a few µL of a cryo-solution were pipetted onto the grid. Using the Leiden Vitrification System the grid was blotted between filter papers to remove the excess of the cryo-solution and plunge-frozen in liquid ethane. Two different kinds of 300-mesh electron-microscopy grids were used, with continuous and holey carbon film respectively.

Preliminary electron diffraction studies

Electron diffraction measurements were performed on a Philips CM30T LaB6 and CM200 FEG working at 300kV and 200kV respectively. A Gatan liquid nitrogen specimen holder was used for the data collection. The data were collected at - 160º C, using selected area diffraction in low dose mode and recorded using a CCD camera.

3.3 Results and discussion

Small protein crystals are generally considered to be unsuitable for crystallographic studies. Whereas this assumption is valid for X-ray single crystal diffraction, it is not true for electron diffraction. In fact, for electron diffraction experiments the crystals have to be very thin because electrons undergo also coulombic interactions with matter which are 10^8 stronger than electromagnetic interaction.

For electron diffraction experiments we used nano-sized lysozyme and potato protease inhibitor crystals grown on strands of hair. The advantage of using heterogeneous crystallization was that all the small crystals could be transferred easily onto an electron microscopy grid by transferring the fibre. No additional "fishing" techniques were needed.

Lysozyme was chosen as a test model since the protein has been studied extensively by X-ray crystallography techniques. Therefore it was possible to compare structural information obtained from X-ray and electron diffraction studies. Potato protease inhibitor (6.1) was selected as an example of a protein from which mostly small crystals are obtained and the optimization of the crystals for X-ray diffraction studies is rather difficult. Moreover, the crystals tend to intergrow which makes fishing out a single crystal for X-ray analysis is more difficult. Potato protease inhibitor was observed to crystallize readily on the hair fibres, as shown in Figure 2.4 in Chapter 2.

Transferring of the crystals on an electron microscopy grid did not present additional technical difficulties. It was also possible to image individual crystals, showing that they are suitable for electron crystallography studies (see Figure 3.2).

Optimizing the specimen preparation is an important step towards the acquisition of high quality diffraction data. Substantial progress in rapid freezing techniques allows biological structures to be studied in their hydrated state. However, formation of crystalline ice often occurs when the excess of water can not be removed completely. Ice crystals yield strong diffraction signals which disturb the diffraction produced by the protein crystals. Most importantly the formation of ice causes structural damage and may destroy the crystallinity of proteins.

There are no generally applicable rules for specimen preparation. Optimizing the thickness of the ice film is largely a process of trial and error However, from crystallization theory it is known that on increasing the viscosity of liquid water all rearrangements including the formation of nuclei occur more slowly.

Therefore compounds with high viscosity such as PEG, glycerol etc. are generally used as cryo-protectants for sample preparation in X-ray crystallography studies.

For the vitrification of the lysozyme nano-crystals we combined aspects of freezing techniques from X-ray crystallography and cryo-electron microscopy. The nano-crystals were plunge-frozen in liquid ethane. Three different cryo-solutions were used – containing 0%, 10% and 15% glycerol, respectively. Compared to X-ray crystallography, where the use of cryo-protectants is essential for the preservation of macro-crystals and their selection is often the most difficult step in the freezing procedure, the protein nano-crystals could be vitrified successfully in liquid ethane without any additives or cryo-protectants. This relaxes sample preparation requirements compared to X-ray crystallography where cryo-protectants are essential for the preservation of macro-crystals and the optimal cryo-protecting conditions need to be identified for each case individually. However, the formation of hexagonal ice was substantially minimized when the nano-crystals were frozen in cryo-solutions containing 10% glycerol. Vitrification of the crystals with 15% or even higher percentages of glycerol appeared to be less successful as this increased the background in the diffraction patterns. A schematic representation of plunge freezing and an electron micrograph of vitrified lysozyme nano-crystals are given in Figure 3.3.

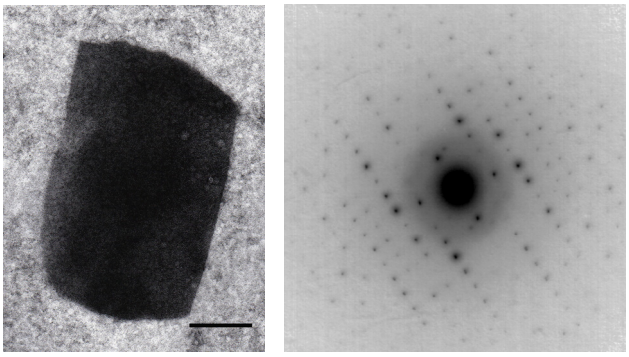


Figure 3.2 Electron micrograph of a negatively stained crystal of potato protease inhibitor (left image); Inset of diffraction pattern acquired from a negatively stained crystal of potato protease inhibitor (right image).

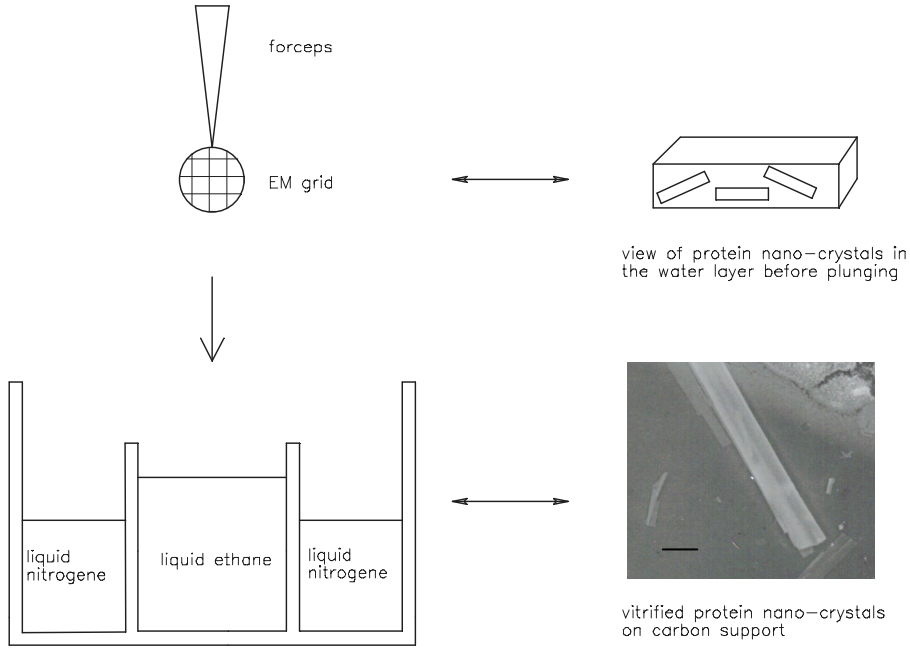


Figure 3.3 Schematic representation of the vitrification of lysozyme nano-crystals by using immersion (plunge) freezing techniques. The bar is 2 μm.

Optimizing the cryo-transfer system is a continuing problem in electron microscopy. From the user's point of view the specimen-exchange procedure is one of the most important and technically most difficult parts of cryo-microscopy. Therefore the transfer system needs to provide a fast, smooth and relatively easy mechanism of exchange. Most of the commercially available Gatan cryo-holders require the use of threaded rings which need to be screwed on top of the grid. The system is difficult to manipulate in a fast and accurate way and especially the screwing on the clamping ring liquid nitrogen causes problems. This often limits the speed and efficiency of specimen exchange. The design of the home modified transfer device we used for our experiments is given in Figure 3.4.

This system uses a sliding mechanism that is easier to manipulate. The sliding devices are also relatively big and therefore easier to handle under liquid nitrogen. This facilitates specimen exchange shortening the time for loading and reducing the chance for specimen warming and the formation of cubic ice.

The diffraction signal itself is determined by the ordering of the atoms in the crystal. It cannot be controlled but the signal-to-noise ratio which co-determines the quality of the diffraction information may be influenced by external factors. Some of these, such as the increasing background noise caused by the radiolysis products formed during exposure, are not easily controlled. Other factors determined by the diffraction conditions, filtering out inelastically scattered electrons, the state of preservation of the crystals and the choice of support or grids as well as the choice of detector can be controlled to some extent. We have investigated whether the background diffraction of the carbon support or the vitrified water can be reduced by freezing the crystals in the thinnest possible water layer on a holey carbon support. Vitrification of the crystals in the holes was possible, but it was also observed that a lot of crystals shrunk or became slightly bent. Although the signal-to-noise ratio of crystals frozen in holey carbon should theoretically be better, bending of the crystalline specimen was of particular concern as it may affect the quality of the intensity data. The crystals vitrified on the carbon layer of the holey carbon film were better preserved and in the majority of the cases showed better diffraction than crystals frozen within the holes.

By using selected area diffraction, data of resolution close to 3.3Å was obtained from vitrified protein nano-crystals (see Figure 3.5). Insertion of a diffraction (selected area) aperture allows quasi-parallel illumination and therefore formation of diffraction spots in stead of diffraction discs. Spots are easier to analyze, especially when the diffraction signal is relatively weak as is the case for protein crystals.

Traditionally in electron crystallography (when inorganic crystals are studied), three dimensional diffraction data are collected through the combination of rotation and additional tilt (beta tilt) which allows orienting the crystal and recording a set of diffraction patterns at various chosen crystallographic zones. Protein crystals absorb energy during data collection, which reduces their crystallinity and consequently the resolution obtained. Therefore orienting the crystals in the beam and acquiring only well oriented diffraction patterns is largely limited by the high beam sensitivity of the proteins and often even not possible. A different approach is used for collecting 3D diffraction data in protein X-ray crystallography where the beam sensitivity of the crystals is also a pertinent problem. In this case the protein crystal is rotated during exposure and (usually) for every degree a diffraction pattern is taken. Applying this method allows maximizing the diffraction information that can be obtained before the crystal is exhausted as no diffraction is wasted on orienting the crystal. The diffraction patterns are subsequently integrated, scaled and merged.

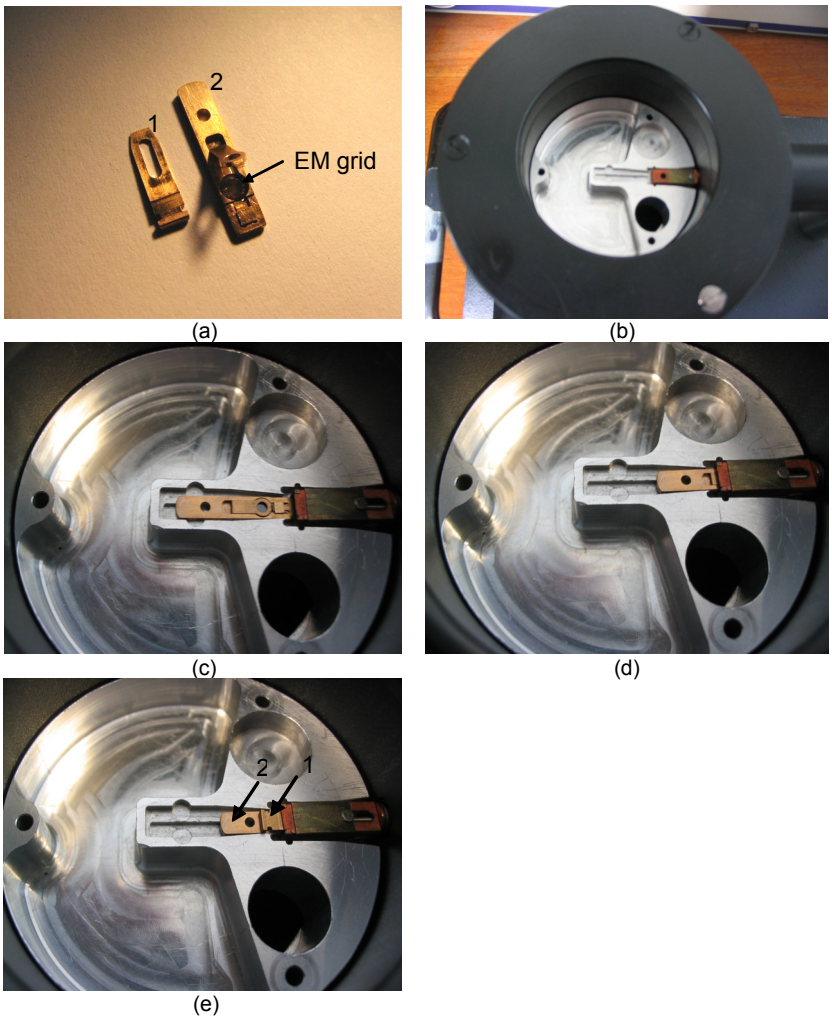


Figure 3.4 Illustration of the home modified cryo-transfer system. In figure 3.4 (a) is shown the tip in which the EM grid is placed (indicated with a black arrow). There are two removable parts as given in figure 3.4 (a) and marked with 1 and 2. After the grid is loaded, the part indicated with 1 is positioned on the top of part 2 as shown in figure 3.4 (e) and used to prevent possible instability or moving of part 1 in the holder. The cryo-holder itself was also modified such that it can accommodate the removable tips as shown in figure 3.4 (b). After the grid is placed in the tip, the latter is sliced in the cryo-holder as illustrated on figure 3.4 (c) and 3.4 (d). The final set-up before inserting the holder in the microscope, is given on figure 3.4 (e).

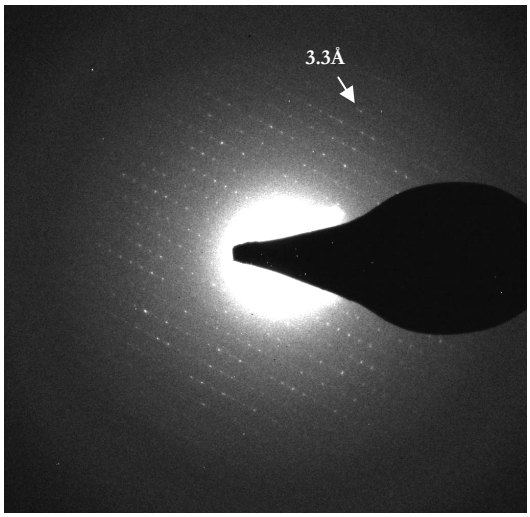


Figure 3.5 Inset of diffraction pattern of a vitrified lysozyme nano-crystal recorded with a CCD camera. No ice rings or diffraction from crystalline ice is present on the diffraction pattern.

In the case of protein nano-crystals it is possible to obtain multiple diffraction patterns from a single crystal using low dose diffraction mode (see Figure 3.6). However, since the crystals used for electron-microscopy are 10^6 smaller than those used for X-ray diffraction studies (even though electrons are three orders of magnitude less damaging than X-rays) the number of diffraction patterns that can be acquired from a protein nano-crystal is much less compared to the number of patterns that can be collected from a macro-crystal. The diffractograms in Figure 3.6 show that after each exposure the protein diffraction signal deteriorates significantly. If a full 3D dataset of sufficient quality can be obtained from a protein macro-crystal, diffraction data-sets from different nano-crystals need to be collected and merged.

3.4 Conclusions

Sample preparation requirements for electron microscopy studies are significantly relaxed compared to X-ray single crystal and powder diffraction. In the case of protein crystals the development of plunge freezing techniques allows vitrification of nano-crystals without additional cryo-protectants

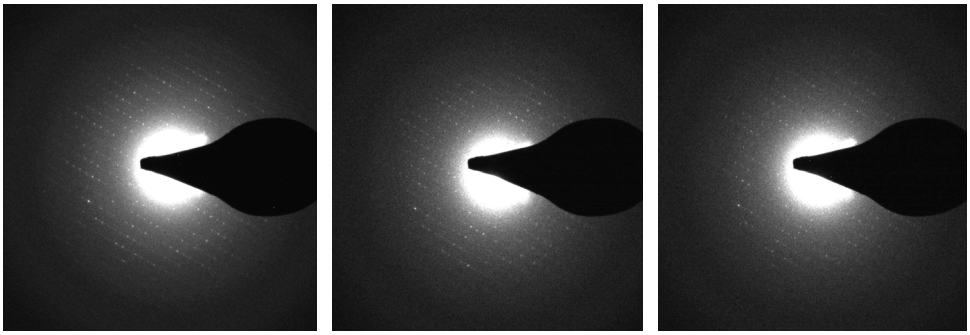


Figure 3.6 Diffraction patterns acquired from the same area of a vitrified lysozyme nano-crystal. The rotation angle is 1º, exposure time 1s. After every exposure and tilt the diffraction signal deteriorates significantly. It is hardly possible to distinguish the weak diffraction spots on the last diffractogram.

For preservation of macro-crystals in X-ray crystallography, the use of cryo-protectants is essential and the optimal cryo-protecting conditions need to be identified for each case individually.

The strong scattering cross-section of electrons allows diffraction information to be acquired from thin protein nano-crystals. The small electron wavelength permits collection of a complete diffraction zonal pattern at a single crystal setting.

Due to the high beam sensitivity and the small size of the protein nano-crystals it is not possible to apply existing strategies for data collection from X-ray protein crystallography or electron crystallography of small molecules. Since diffraction patterns from different crystals need to be collected in order to reconstruct the 3D reciprocal space, exhausting a single crystal in order to collect as many as possible diffraction patterns at the expense of poor quality and low resolution is highly unlikely to provide a successful structure solution. An alternative is to collect patterns with high quality diffraction information (high signal-to-noise ratio, high resolution, less dynamical perturbations) preferably from undamaged areas in a wide range of crystals settings. In this case the angular relationship between the diffraction patterns is lost. This makes data analysis definitely more difficult. However, the more diffraction information is present on a single pattern and the more kinematical the data are, the easier will be to identify the crystal phase and index the reflections.

References

1. Georgieva, D.G., Kuil, M.E. and Abrahams, J.P. (2006). Protein nanocrystallization. *Springer seriesin Biophysics, Advanced Techniques in Byophysics, Springer-Verlag Berlin Heidelberg*.
2. Margiolaki, I., Wright, J.P., Fitch, A.N. and Gavin, C. (2005). *Acta Cryst*. D**61**, 423.
3. Henderson, R. (1990). *Proc. R. Soc. London*, B**241**, 6.
4. Henderson, R. (1995). *Q. Rev. Biophys*. **28**, 171.
5. Dubochet, J., Adrian, M., Chang, J., De Homo, J.C., McDowall, W. and Shultz, P. (1988). *Q. Rev. Biophys*. **21**, 129.
6. Thomassen, E.A., Pouvreau, L., Gruppen, H. and Abrahams, J.P. (2004). *Acta Cryst*. D**60**, 1464.
7. Georgieva, D.G., Kuil, M.E., Oosterkamp, T.H., Zandbergen, H.W. and Abrahams, J.P. (2007). *Acta Cryst.* D **63**, 564.

Chapter 4

High resolution electron diffraction of 3D protein nano-crystals: optimizing 3D data collection and data analysis

to be published as: Georgieva, D.G., Zandbergen, H.W., Nicolopoulos, S., Sarakinou, E. and Abrahams, J.P. Optimal data collection of beam sensitive 3D crystals of proteins and pharmaceuticals by electron diffraction.

Abstract

For 2D protein crystals and small molecule nano-crystals, electron diffraction is the only option, but 3D protein crystals present additional problems. Limitations in collecting diffraction data of sufficient quality, the complexity of the scattering and the lack of procedures for processing non-oriented diffraction patterns from different crystals have so far frustrated the promise electrons hold in this respect. Here we report that these problems can, to a certain extent, be overcome by vitrifying the sample, applying low dose diffraction techniques (such as microdiffraction) and precession of the electron beam. Our procedures, specifically aimed at gathering high-resolution, 3D reciprocal space data, allowed electron diffraction data to be collected up to 2.1 Å resolution of 3D nano-crystals of lysozyme. By precessing the beam not all reflections are simultaneously excited which renders the diffraction pattern less dynamical, reducing intensity variations caused by multiple scattering. This facilitated diffraction pattern recognition and identification of the crystal phase. Unit cell determination and indexing of the resulting diffraction patterns were done with our newly developed software: the algorithm is also described in this chapter. The parameters inferred were confirmed by existing indexing programs in the case of well oriented diffraction patterns.

4.1 Introduction

Structure determination by electron diffraction of 3D protein nano-crystals remains to be validated. There are several theoretical and practical reasons why it is difficult to collect 3D electron diffraction data of a quality sufficient for structure determination:

Dynamical diffraction of electrons cannot be completely ignored when protein crystals reach a thickness of more than about 100 nm[*]. The dynamical effect can be treated as an auto-convolution of the diffracted structure factor data and it is determined by the thickness of the crystal and its orientation in the electron beam. Without knowledge of the phases, the thickness of the crystal and its precise orientation, the dynamical effect cannot be modelled accurately.

It is not possible, using current electron microscopes, to accurately and uniformly rotate the crystal during exposure, analogous to the rotation method in protein X-ray crystallography. Hence, none of the recorded reflections will have fully moved through the Ewald sphere, currently limiting the method to the collection of "stills".

Due to the high beam sensitivity of proteins it is not possible to accurately orient the crystal in the beam and certainly not to collect a full 3D data set from a single nano-crystal.

[*] Assuming the mean elastic scattering cross section of the average atom within a protein to be in the order of $5x10^{-5}$ nm^2 (the elastic scattering cross section of carbon at 300 kV;) and the average density of a protein to be in the order of 50 scattering atoms per nm^3 (there are about 30 H_2O molecules per nm^3 of liquid water), the chance of an electron being scattered elastically within a protein volume of 1 nm^3 is about $(50*5x10^{-5}) = 0.0025$. These numbers imply that about 90% of all electrons that are scattered elastically by a 100 nm thick protein crystal (which in the case of lysozyme would correspond to 15 to 30 unit cells) would scatter only once, whilst 10% would be scattered more than once. For protein data, where the discrepancy between model structure factor intensities and the measured data is usually in the order of 20% in the outermost resolution shells, even for well-determined X-ray structures, this is just about tolerable.

For crystals that are substantially thicker than 100 nm, when dynamical scattering is contributing significantly to the diffracted intensity, the electron beam precession technique proposed by Vincent & Midgley [1] offers a partial solution by decreasing the dynamical behaviour. This technique is equivalent to the Buerger precession technique used in X-ray diffraction, where the crystal is precessed with respect to the incident X-ray beam. In the electron precession technique the electron beam is tilted and precessed along a conical surface, having a common axis with the optical axis of the TEM (see Figure 4.1). As a result not all of the reflections are excited simultaneously, which renders the resulting diffraction pattern less dynamical.

Much of the systematic dynamic scattering which extends more strongly in on-axis conditions for incoherent multiple scattering, leading to the appearance of kinematically forbidden reflections, is also reduced as the precessing beam is entering the sample from off-axis directions [2].

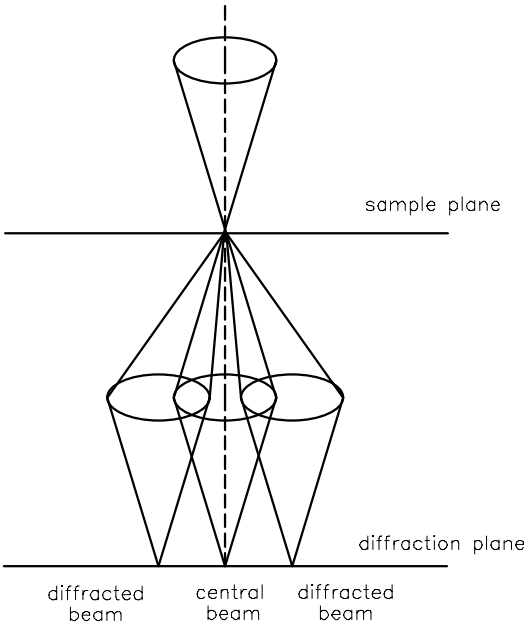


Figure 4.1 Schematic representation of precession electron diffraction. (Adapted from Own C.S. 2005)

Due to multiple scattering events, the electron diffraction intensities oscillate with specimen thickness. This makes the analysis of data acquired from different crystals of non uniform thickness particularly difficult. It was shown for a number of inorganic crystals that by averaging over incident beam directions and different thicknesses while precessing the beam, the intensity variations caused by non-systematic dynamic effects are reduced.

In X-ray crystallography the crystal is rotated during exposure which allows more reflections (including high resolution reflections) to pass through the Ewald sphere, resulting in the collection of integrated intensities. Rotation or tilting of the crystal in the electron microscope during exposure is not possible. However, precessing the beam during exposure is optically equivalent to precessing the crystal and allows the collection of integrated intensities and high resolution reflections.

When the precession angle is larger than the rocking curve, at least some reflections will have fully passed the Ewald sphere during exposure to a precessing electron beam (see Figure 4.2). This will facilitate scaling and merging the integrated intensities recorded from different crystals with different orientations: not only is dynamical scattering reduced, which renders the diffracted intensities of symmetry related reflections more equal and renders the measured intensities more proportional to the structure factors, but there are also data from fully recorded reflections.

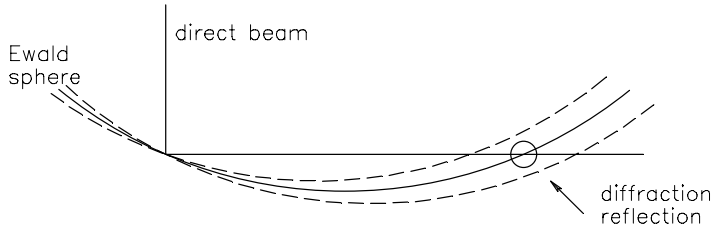


Figure 4.2 Intersection of the Ewald sphere with the reciprocal space in conventional electron diffraction (whole line) and in precession mode (dashed line). (Adapted from Own C.S. 2005)

In view of these potential advantages of precession electron diffraction, we set out to optimize 3D data collection with this technique.

Traditionally, in electron crystallography the unit cell of unknown crystalline phases is determined by the method of reciprocal lattice reconstruction from an electron diffraction tilt series. For the purpose three-dimensional (3D) diffraction data is collected by tilting a crystal around a selected crystallographic axis and recording a set of oriented diffraction patterns (a tilt series) at various crystallographic zones. However, this method is not always applicable in the case of 3D protein and other organic nano-crystals. The high beam sensitivity of the materials often does not allow collection of tilt series from a single nano-crystal. This has limited so far the application of electron diffraction for studying beam-sensitive molecules.

We measured diffraction from different crystals, assuming that it would be possible to subsequently extract the unit cell, the orientations of the single shot diffraction patterns and the intensities of their reflections. A new algorithm for unit cell determination from randomly oriented electron diffraction patterns is also discussed. This chapter covers the first part of the data-processing procedure and describes the steps leading to determining the unit cell parameters.

4.2 Experimental procedures

4.2.1 Electron diffraction data collection

Electron diffraction measurements were performed on a Philips CM30 LaB6 and CM200 FEG working at 300kV and 200kV respectively. For diffraction experiments a 30μm C2 aperture was inserted. For the acquisition of the diffraction patterns the beam was defocused on the desired area in image mode typically (1μm). No beam blanker was used. Once the correct adjustments to the microscope were done and the first diffraction patterns were obtained, the grid was further scanned only in diffraction mode. Switching from diffraction to imaging mode and *vice versa* was done only in cases when the diffraction signal deteriorated significantly. A single tilt, home modified Gatan liquid nitrogen specimen holder was used for data collection. The data were collected at - 160º C. A tilt of +/- 45º was applied to collect diffraction patterns from different crystallographic zones. The tilt angle was mainly limited by the dimensions of the holder and the goniometer.

Diffraction data with precession were collected using the Spinning Star P020 (NanoMegas Company). The diffraction patterns were recorded on a CCD camera or DITABIS image plates. Single diffraction patterns were exposed for 2 s.

4.3 Results and discussion

4.3.1 Conventional and precession electron diffraction of 3D protein nano-crystals

The typical approach to collecting electron diffraction data is to use a standard selected area aperture (SAD) or select via the illumination using convergent beam electron diffraction (CBED). When SAD is used the beam incident on the specimen is parallel and relatively large (1-10μm). In CBED, the beam is converged on the area of interest at the specimen plane, resulting in convergent beam electron diffraction. Although diffraction patterns obtained via CBED techniques can provide a wealth of extra information not available in SAD, the use of CBED is limited by the stability of the material in the beam. It is not the method of choice when highly beam-sensitive materials like proteins or pharmaceuticals are studied. SAD is generally regarded as inducing less beam damage. However, it has to be considered that the selected area aperture is inserted in the image plane below the specimen and the whole area is still being exposed during diffraction, whilst only diffraction from the selected part is recorded. This, as well as the limited spatial resolution of SAD can reduce the resolution and the quality of diffraction, especially for highly beam sensitive materials.

A different approach is to use a small second condenser (C2) aperture (30μm) and a spot size of around 8-10 nm. Selected area aperture is not used. This set up allows for a quasi-parallel illuminated area [3]. At these illumination conditions only the area from which the diffraction is obtained is exposed to the beam. We used a similar approach but we defocused the beam in image mode (see Figure 4.3). In this case the diffraction pattern is not focused on the back focal plane but instead slightly above or below. One of the major advantages of this method is that more atom planes contribute to the formation of the diffraction signal and therefore the obtained diffraction spots are stronger which allows also very weak high resolution spots to be detected. Compared

to selected area diffraction only a small part of the crystal is illuminated, so the damage remains local. When the conventional microdiffraction mode is applied, diffraction discs are obtained. The change in diffraction focus (when the beam is defocused in image mode) leads to the formation of spots in stead of discs in the diffraction pattern. Therefore the signal- to-noise ratio is improved as well.

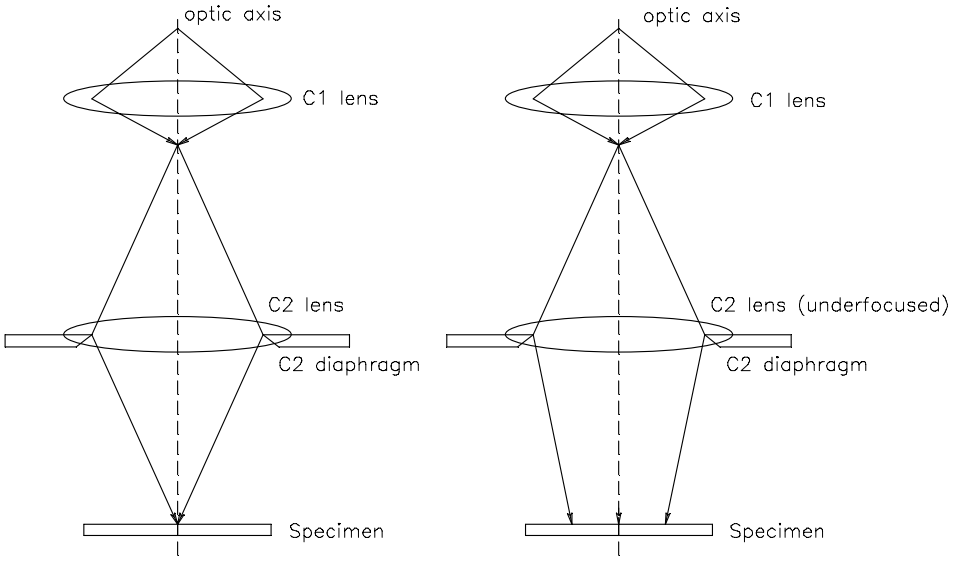


Figure 4.3 A focused C2 lens illuminates a small area of the specimen (left image); by underfocusing the C2 lens the beam is defocused on the area of interest and more quasi - parallel illumination is obtained.

The choice of the detector was shown to be important for optimizing diffraction data collection. CCD cameras have certainly a lot of advantages and are often the preferred recording media in electron microscopy. However, due to their high dynamic linear range and low intrinsic noise, image plates allow not only high frequency diffraction reflections to be recorded (in the case of lysozyme up to 2.1 Å resolution) but also small differences in intensities to be detected (see Figure 4.4). This is of particular concern for protein crystals since generally they yield very weak diffraction signals. Another major advantage of image plates is that they relax the need of a backstop. We

concluded that image plates are currently the preferred detector for diffraction data collection, even though they have to be scanned off-line.

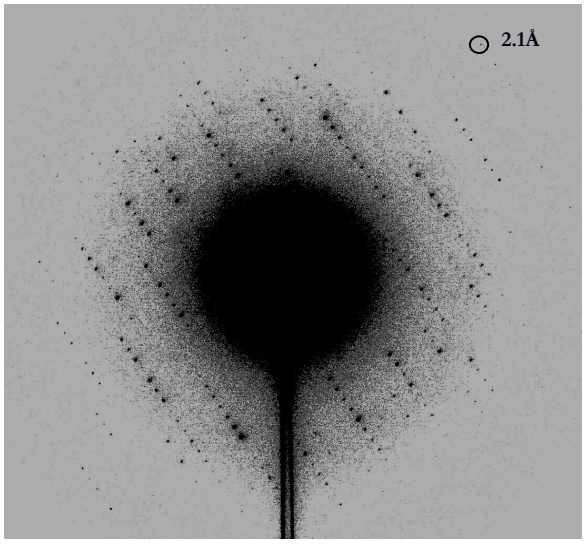


Figure 4.4 Diffraction pattern from a vitrified lysozyme nano-crystal recorded on an image plate. The high dynamic range of image plates allow very weak high resolution diffraction reflections to be recorded from highly beam sensitive crystals as well as small differences in intensities to be detected.

In the case of lysozyme, precessing the beam was used in order to reduce the effect of dynamical scattering and to allow full motion through the Ewald sphere of the precessing diffraction spots - reducing the number of partially recorded reflections. Therefore, the technique proposed for collecting diffraction data by defocusing of the beam was also tested in precession mode.

The crystals had a preferred orientation on the carbon support, with the long unit cell axis normal to the plane. By applying alpha tilt diffraction, data in a wide range of crystal settings were collected (see Figure 4.5). Acquiring data of sufficient quality from high-index zones is more difficult at the current state since in this case high alpha tilts are often required. Due to limitations caused by the geometry of the holder, high tilts could not be applied if the crystals were close to the edge of the grid. It has to be

considered as well that if a plate-like crystal is tilted (as in the case of lysozyme crystals) the thickness from which the diffraction is obtained increases with increasing of tilt angles. In this case the dynamical effects also become stronger. This is of special concern if the crystals are thicker than 80 - 100nm (for crystals composed mostly of light atoms) and it needs to be taken into account when crystals to be subjected to tilt are selected.

In total more than 300 diffraction patterns from different nano-crystals were collected. (see Figure 4.5 for examples)

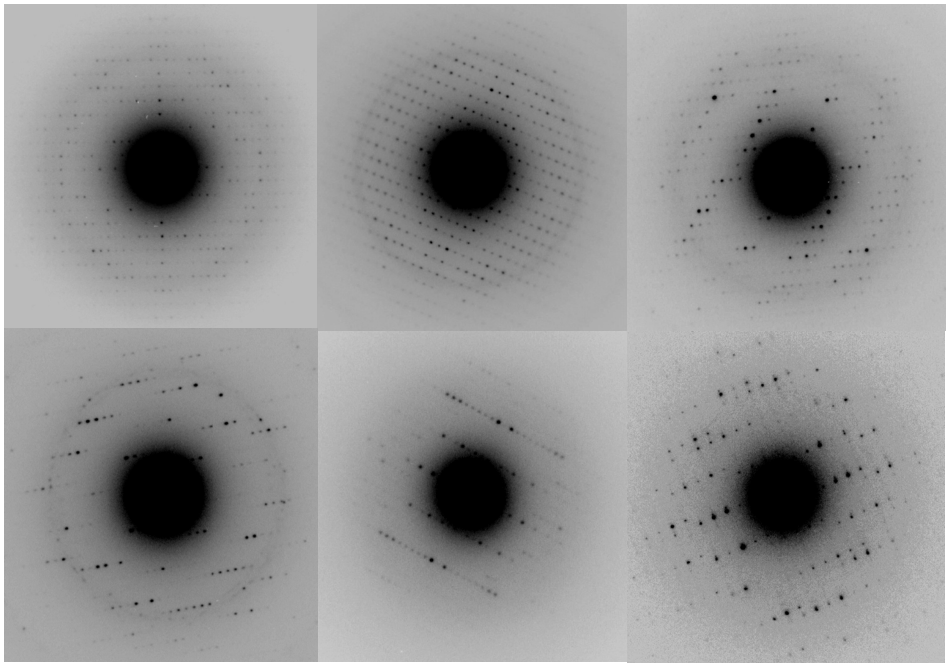


Figure 4.5 Diffraction patterns from vitrified lysozyme nano-crystals collected in a wide range of crystal settings.

4.3.2 A new algorithm for unit cell determination from electron diffraction patterns acquired from randomly oriented crystals

A new algorithm for unit cell determination from randomly oriented electron diffraction patterns is presented.

It first requires pre-processing of the data to generate auto-correlation maps of the diffraction patterns. Then the peak positions of the auto-correlation patterns are analyzed to generate the most likely unit cell. These two steps are discussed separately.

Data pre-processing

Prior to reconstruction, the original data should go through a number of corrections. At this stage, the centre of the diffraction patterns is defined, the central beam removed and the background noise subtracted.

The centre in a diffraction pattern is found based on a search for the connected spots of the highest intensity. For the purpose the original diffraction pattern is scanned from the top to the bottom and the coordinates of the most high intensity consecutive pixels as well as the centre of this consecutive sentence are stored. The same search is repeated also from the left to the right. The coordinates found from the two scans (the scan from the top to the bottom and from the left to the right) are compared and averaged. If there is a substantial deviation between the coordinates, the centre is defined manually. In the later procedure, the position of the centre is refined.

In the next step the central beam is removed and the background noise subtracted (see Figure 4.6). It is assumed that the preliminary centre in the diffraction pattern is already known. From this centre the radius is increased by 1 pixel and for each ring (with a thickness of 1 pixel) the average intensity is calculated by dividing the summed intensities of all the pixels from the selected ring by the number of the pixels. The calculated average intensity is then subtracted from the image.

Next, autocorrelation maps are generated from the diffraction patterns (see Figure 4.7).

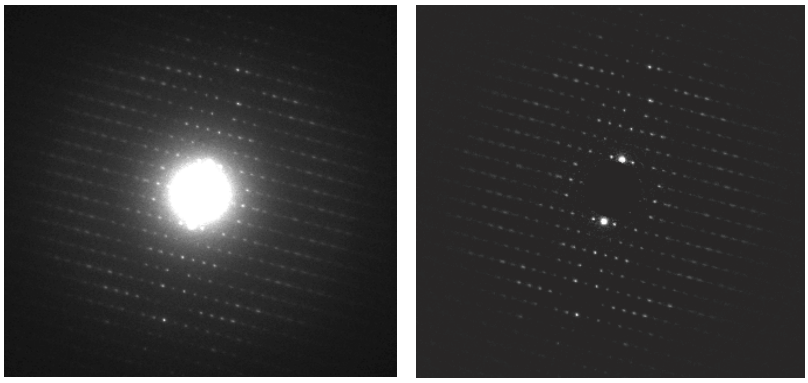


Figure 4.6 Removal of the central beam and background subtraction. The left image is an electron diffraction pattern of a lysozyme nano-crystal. The right image is the same diffraction pattern after removal of the central beam and background subtraction.

By definition the autocorrelation function is a cross-correlation of the signal with itself. It is used in the data pre-processing step for finding repeating patterns e.g. the presence of a diffraction periodic signal buried under noise or identifying missing frequencies in the signal such as systematic absences in the diffraction pattern. By constructing the autocorrelation image, possible artifacts and the background noise are further reduced. The diffraction reflections are also enhanced. This facilitates finding the positions of the diffraction spots as well as calculating the repetitive distances between them. In the cases when the signal- to- noise ratio of the original diffraction pattern is high, the autocorrelation map will also display diffraction peaks with intensities substantially higher than the background. Those peaks can easily be selected by setting a fixed threshold value. Only the pixels above this set value will be further treated in the analysis.

However, in practice the problem is more complicated since some of the diffraction reflections e.g. high frequency reflections may have roughly the same intensities as the background noise in the low-resolution area (close to the central beam). Therefore, the intensity landscape needs to be examined locally at any part of the image. For the purpose, the background noise from the autocorrelation images was subtracted using the same algorithm applied for the background subtraction from the original diffraction

pattern. The intensity landscape is calculated by adding up the intensities of all pixels in a ring divided by the total number of pixels in this ring.

Furthermore, a Gaussian filter is applied to the autocorrelation images. Gaussian filtering is used to smooth the images i.e. reduce the amount of intensity variation between two neighboring pixels and therefore remove noise and artifacts. If a big Gaussian filter is applied, the image gets blurred. Each intensity peak is broadened until it becomes almost flat. The obtained "blurred" image is used further as a background and subtracted from the original image. As a result the "new" autocorrelation image displays sharper peaks with lower intensities.

In the next "peak searching" step each peak consisting of multiple pixels is reduced to a single (x, y) coordinate. For the purpose, each pixel is checked against its neighboring pixels. Only those pixels with the highest value are treated as single peaks and stored in a *.plt file. The resulting peaks are shown in Figure 4.8 where each spot consists of a single pixel and corresponds to a peak in the autocorrelation map.

In the final step the centre of the diffraction pattern is refined. An autocorrelation image is by definition cantered. This property is used in the refinement step.

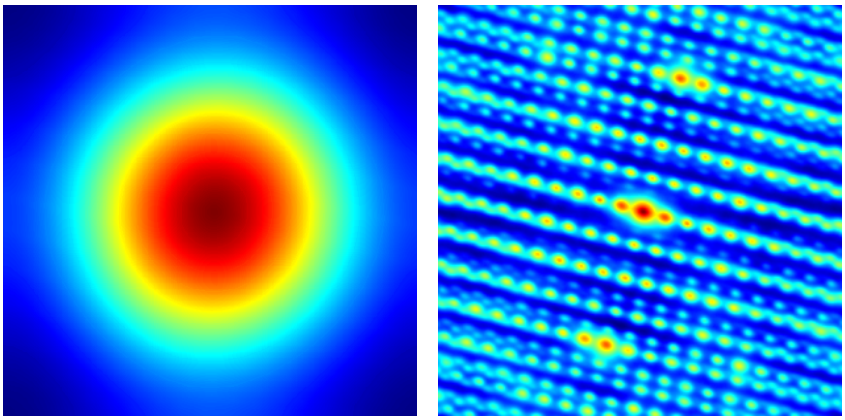


Figure 4.7 Autocorrelation images generated from an electron diffraction pattern of a lysozyme crystal. The left image is the autocorrelation map generated from the diffraction pattern before and the right image after beam centre removal and background subtraction.

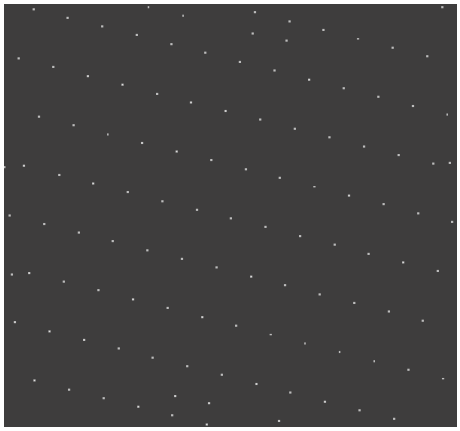


Figure 4.8 Inset of a *.plt file. In the "peak searching step" each pixel is checked against its neighboring. Only the pixels with the highest value are treated single peaks and stored as a *.plt file. The resulting peaks are shown. Each spot consists of a single pixel and corresponds to a peak in the autocorrelation map.

After the original image and the autocorrelation map have been brought to the same size, the distances between the autocorrelation maps and the intensity peaks of the original image are minimized. The *.plt images are used as "templates" for the minimization and refinement step since they have little or almost no background noise. A flowchart of the steps involved in the data pre-processing are given in Figure 4.9.

Description of the algorithm for unit cell determination from randomly oriented diffraction patterns

The method discussed in the following paragraphs currently deals successfully only with primitive cells. Non-primitive, centered cases will be investigated later. The method uses a least mean square approach to search for the optimal unit-cell model which fits the experimental data best. First, a unit cell with 6 parameters is generated and the low resolution vectors of the corresponding reciprocal lattice are computed. Second, the best matched orientation or simulated intersection face is found for every autocorrelation function of each diffraction pattern. All the difference values between the observed and simulated images are summed in a penalty function.

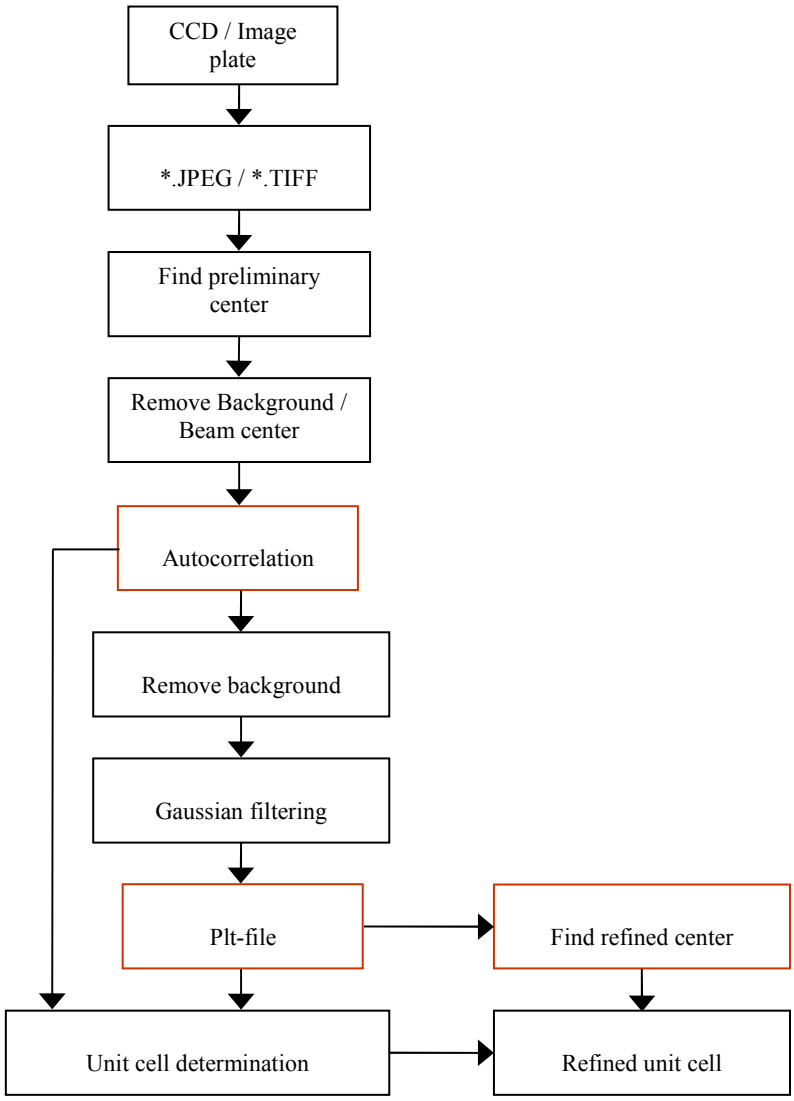


Figure 4.9 A flowchart of the steps required in the data pre-processing. With red boxes are indicated the output files which are subsequently fed into the program for unit cell determination.

Last, an optimization of the model is performed in order to get the smallest penalty by exhaustive grid search of different parameter combinations or linear approximations starting from a local optimal parameter set.

Cell model and 3D reflection lattice

Given 6 parameters, a, b, c and α, β, γ, we can define a primitive cell, where a, b and c are the three principal axes in a crystal, α is the angle between b and c-axis, β is the angle between a and c-axis, and γ is the angle between a and b. Using these 6 parameters all possible unit cells can be simulated by varying the dimensions of the axes and the angles. If chemical or structural information is available beforehand about the structure, it is used to define the search range (the range in which the unit cell parameters are expected to be) as well as the step size with which the parameters will be varied. For instance, if the structure is inorganic, it is known that the searched lattice parameters will be relatively small. In contrast, proteins or biological macromolecules have large unit cell dimensions. The simulated unit cells are used as starting models for the lattice cell determination of the structured studied.

Using a set of cell parameters, a cell matrix can be constructed. The position of any reflection point 'P' of the 3D reflections lattice in Fourier space for a defined cell can be calculated using the equation:

$$p = hkl * M \qquad (1)$$

Where *hkl* (*h, k, l*) are the indices of P, M is the cell matrix, and p is the (x, y, z) coordinate of P.

The equation above is unlimited for possible *hkl* values, therefore it can not be used straightforwardly to simulate a model unless it is not known which hkl satisfy 'P' in a given resolution range.

The hkl indices that satisfy 'P' for a chosen resolution range can be found by applying the following constraints.

$d_{min} \leq |p| \leq d_{max}$ (2)

where d_{min} is the lower boundary of resolution range and d_{max} the upper boundary resolution. The idea presented is also illustrated schematically in Figure 4.10.

The algorithm was implemented in the computer program EDiff and all the possible positions of reflection spots in 3D Fourier space under a given resolution can be generated quickly.

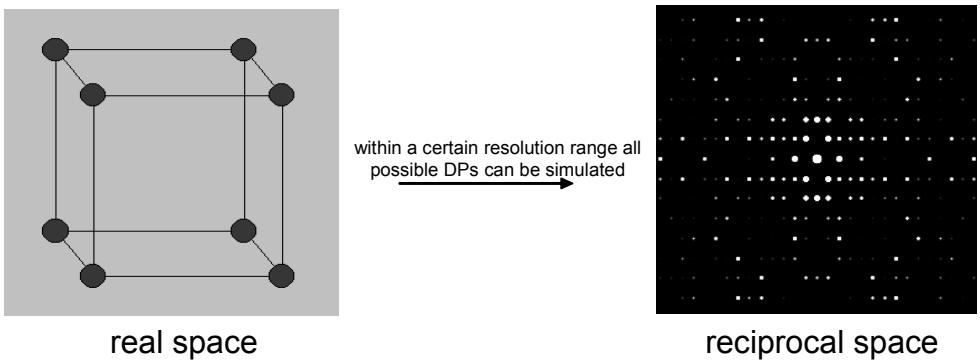


Figure 4.10 From a cell model to a reciprocal cell lattice within a certain resolution range.

Experimental diffraction patterns and simulated intersection faces

Theoretically, the diffraction patterns are 2D intersections of the 3D reciprocal space. The Ewald sphere cutting planes of the 3D reciprocal lattice generates the diffraction patterns. Given an orientation, a 2D cutting plane or a diffraction pattern can be generated from a 3D reflections model. The 2D plane of a diffraction pattern always passes through the origin of the 3D lattice. And all the 2D diffraction patterns share the same centre with the 3D reciprocal lattice model.

The Ewald sphere has a radius of $1/\lambda$ where λ in the case of electron diffraction is very small, e.g. 0.01938 Å for electrons with energy 300KeV and therefore the Ewald sphere is very "flat", especially in the low resolution area. In order to find the dimensions of

the unit cell, the low resolution information is enough. In a later step, the high resolution data will be used to refine the cell parameters.

In a first approximation we can simply use a plane to simulate the Ewald sphere in order to find the unit cell parameters. The electron diffraction image in Figure 4.11 displays a regular lattice and a non-scattered central beam. The recorded diffraction spots show how the Ewald sphere cuts the 3D reciprocal reflection lattice. Only the reciprocal lattice nodes that pass through the Ewald sphere are excited and observed as Bragg reflections on the diffractograms. A given reflection on the diffraction pattern can be characterized by the vector from the centre of the pattern to the spot. The short vectors contain the basic low resolution information. The two shortest vectors and the angle between them determine the periodic units in the reflection lattice. In the original diffraction patterns it is usually hard to find the first order diffraction reflections since they may be covered by the central beam or be absent (i.e. systematic absences) (see Figure 4.11). Therefore, the shortest distances are determined from the autocorrelation maps.

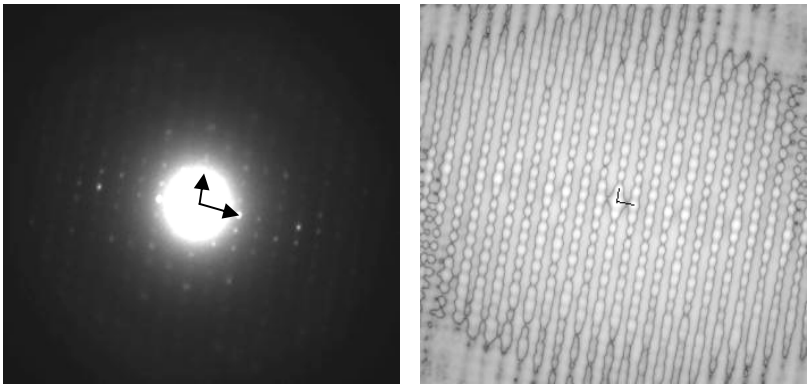


Figure 4.11 It is often difficult to select the two shortest vectors from the original electron diffraction pattern (left image). Often the first order reflections are hided by the central beam or they are missing due to systematic absences. However, it is fairly easy to select the shortest vectors from the auotocorrelation map (right image).

For each diffraction pattern the shortest vector pair (facet) is found (see Figure 4.12). A vector pair consists of the vectors between two spots and the origin as well as the angle

between those two vectors. From these three parameters one can form triangles with different shapes (facets). Each facet corner should overlap with three (out of the eight) corner points of the integer multiple of a unit cell. The model facets are built up from imaginary unit cells of different sizes. For each size, one can construct all unique facets that fit this imaginary unit cell. In the previous section, it was discussed how the cell model is built. Given an orientation, a 2D diffraction pattern or the possible intersection planes from the 3D reciprocal lattice can be simulated. There must be at least one simulated 2D pattern which will match the observed diffraction pattern the best.

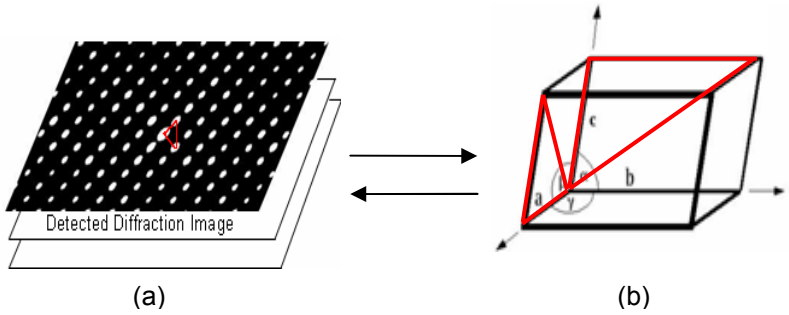


Figure 4.12 For each diffraction pattern the shortest vector pair (facet) is found - (a). A vector pair consists of the vectors between two spots and the origin as well as the angle between those two vectors. From these three parameters (the vectors and the angle between them) one can form triangles (facets). Each facet corner should overlap with one corner of the integer multiple of the unit cell - (b).

The task of finding the unit cell parameters is equivalent to finding a set of unit cell parameters from which we can simulate 2D diffraction patterns from certain orientations (or unique facets) and fit all the experimental diffraction images with the smallest error.

Diffraction patterns can be generated only when the Ewald sphere passes through or touches some of the reflections. The same principle applies for the simulated diffraction patterns which are 2D intersections of an imaginary 3D cell lattice. Therefore, when we are sampling and trying to find the orientation of a experimental diffraction pattern, it is not necessary to perform an exhaustive search for all possible orientation angles. We need to consider only the orientations or intersection planes

passing through the origin and two reflections which do not lie on the same line in the 3D reciprocal lattice. Three points which are not on one line define a unique plane and also they can define the Ewald sphere if its radius is known. Beside the origin (the central spot) if we select another two diffraction reflections which do not lie on the same line, we select a unique plane which we further call "main facet".

Since the 2D diffraction patterns are intersections of the 3D reflection lattice, the facets in the 2D diffraction patterns should be a subset of the facet in the 3D reflection lattice. And any facet in the 2D diffraction pattern should match one of the facets in the 3D reflection lattice. Finding the orientation angles of each diffraction image is the same as finding a facet in the 3D reciprocal cell lattice which also shows up in the diffraction pattern.

Function approximation

In an ideal case, all facets from the experimental data should exactly fit the facets of one specific model unit cell. However, in practice it is impossible to measure the exact vector length and the angles in the diffraction pattern. The diffraction spots have a certain size, and one has to calculate the weighed peak point. Furthermore, images consist of a certain number of pixels, meaning that one needs to work with integer peak coordinates. Measurement errors can therefore not be ignored hence, function approximation needs to be performed. In general, function approximation is a mathematical approach used to select a function among a well-defined class that closely matches or "approximates" a target function in task-specific way.

In the case of the unit cell determination task, if we know the correct cell parameters and search for all the possible reflection pairs (every reflection pair is corresponding to a facet) in the 3D reciprocal cell lattice, we can certainly find two reflections which will match the corresponding reflections from the diffraction pattern. However, if we don't know the correct unit cell parameters but use random simulated 3D reciprocal cell lattices, searching for all possible vector pairs, we can still find the best fitting pairs but within a certain error (there will be a certain difference between calculated distances from the diffraction patterns and the simulated distances from the imaginary cells). The squared sum of those differences from all the diffraction patterns can be used as an evaluation of the fitting between the simulated cell parameters and the real ones.

The idea can be illustrated and implemented also in terms of facets. For given unit cell parameters, a 3D reflection lattice can be simulated. For each facet found from the experimental diffraction pattern, we can find the corresponding facet in the 3D reflection lattice which fits the best. The squared difference between the experimentally found facets from all the collected diffraction patterns and the simulated ones are accumulated in a penalty function. The function used to calculate the accumulated differences between the experimental values and the values obtained from the simulations are used as a target function. Optimization of the cell parameters is done by using a grid search method in a later step.

In practice, a lot of computational time is needed in order to match each facet from a diffraction pattern with all possible facets in an imaginary cell. If we can select two main vectors or any other unique vector pairs in the diffraction pattern, we could use only the main facet or that unique facet in the target function rather than search for all possible facets.

The "square difference function" is used to calculate the least fitting square error of two facets. If we assume that A_0 and A_1 are the vector pair of the observed facet and B_0 and B_1 are the vector pair of the simulated facet, the square error is defined as:

$$SQER = |A_0\text{-}B_0|^2 + |A_1\text{-}B_1|^2$$

The square error we call here also facet residue and in fact we try to find the smallest facet residues A_0 B_0 and A_1 B_1 (see Figure 4.13)

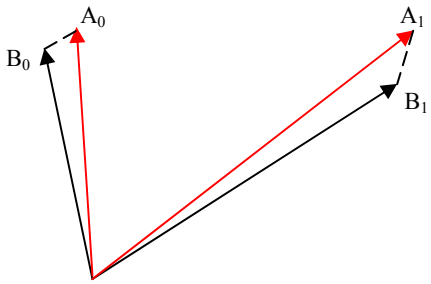


Figure 4.13 Illustration of the idea of the facet residues used to calculate the least fitting square error of two facets.

Grid search and optimization of the model

In the previous section the application of the two target functions where all the possible facets are considered, and when only the "main" facets are used to calculate the least square deviation between experimental and observed facets, was discussed. These two target functions can be used in the optimization step of the search for the correct primitive cell parameters.

The primitive cell is defined by six parameters - the cell axes (a, b, c) and the angles between them. In the first test calculation, a simple grid search method is performed. This is a six dimensional exhaustive search. It is also possible to use a different method in the optimization step such as the quasi-Newton method (which is another well-known algorithm used for finding local minima and maxima of functions). In a later step, a more detailed search is performed to refine the unit cell parameters. This is based not only on the dimensions of the cell axes and the angles between them but also includes the orientation angles and the refined centre of the patterns.

Generally, a six parameters primitive cell model can be used in the case of all different seven crystal systems (cubic ... etc see Table 4.1). If the crystal system is known beforehand, constraints can be applied to the cell parameters in the exhaustive search. For example if the crystal system is tetragonal, it is known that two of the cell axes are equal and all three angles are 90º etc. Using this information in the input will speed up the search and improve the precision of the obtained results.

However, in the search for the six parameters of the primitive cell using the functions expressed in formula 1 and formula 2, a problem arises when the reciprocal cell lattice is very dense (the reciprocal reflections are very close to each other). In these cases we can always find some simulated reflection pair (facet) which will match the observed reflection pairs. The RMSE value which is used to evaluate the fit in these cases will be very low. In a simulation test it was found that the square error (facet residue SQER) statistically decreases to $(1/N)^2$ when the dimensions of the simulated axes increase N times. If the Root Mean Square Error is used in the target function, it decreases with (1/N). This suggests that if we consider a large imaginary cell it is always possible to find facets which will match the facets calculated from the collected diffraction patterns and the denser the reciprocal lattice (the bigger the considered cell) is, the smaller the squared differences between experimental and simulated facets would be found. Obviously, this problem needs to be addressed.

Table 4.1 A six parameters primitive cell model can be used in the case of all different crystal systems. Depending on the crystals system, constraints can be put on the cell dimensions and the angles when searching for the unit cell parameters.

Crystal system	Constraints on cell dimensions & angles
cubic	a=b=c, α=β=γ=90˚
tetragonal	a=b, α=β=γ=90˚
hexagonal	a=b, α=β=90˚, γ=120˚
orthorhombic	α=β=γ=90˚
monoclinic	α =γ=90˚
triclinic	none

We know the index of the reflections in the simulated 3D reciprocal cell lattice. When we fit an observed facet with a simulated facet, we assign indices to the observed reflections of the facet at the same time. The index contains also information about the lattice, i.e. how "dense" the lattice is. If we get two possible indices for a certain reflection e.g. a high index and a low index when different imaginary cells are considered (one with large unit cell dimensions and another one with smaller dimensions), we can multiply the square error with the square of the maximum value of the *hkl* index weighing the result.

In a test with mayenite (a mineral with a cubic crystal system and a=11.99Å) if the problem with the dense 3D lattice is not taken into account, large unit cell parameters were found, such as 24Å. After the weighted error was used as a target function evaluation, the correct answer was obtained.

The optimization of the target function can be used not only to find the unit cell parameters but also to define how the Ewald sphere has cut the reciprocal lattice in every diffraction pattern. The orientation of the Ewald sphere that we find for each image is actually the closest main zone. In reality, the Ewald sphere does not go through the middle of all the reflections (the reflections are not fully excited). For the reflections which are partially excited, we may get larger errors in the target function. It

needs to be considered as well that some the reflections lying on the Ewald sphere and recorded as diffraction peaks might be from different crystallographic zones or different Laue zones, for instance in the case of heavily mistilted diffraction patterns. To get more precise unit cell parameters and the exact orientation of the Ewald sphere for each diffraction pattern requires indexing all the reflection on the patterns (not only the low resolution ones). In the refinement step many more parameters need to be optimized at the same time including the cell axes and the angles between them, the orientation angles, and the centre position.

4.3.3 Comparison of the existing electron diffraction programs to EDiff

Automated methods for unit cell determination and indexing of electron diffraction patterns are becoming more common. In the cases when tilt series can be acquired from a single nano-crystal, there are well established algorithms used to determine the cell dimensions and to index the acquired oriented diffraction patterns. Such algorithms are e.g. implemented respectively in the computer program TRICE [4] and PhIDO [5]. In the general case, it is necessary to solve patterns from all crystal systems and at any sample orientation. This is a more difficult task than for instance indexing on-axes patterns.

If the unit cell dimensions and the type of crystal system are known (e.g. from X-ray powder diffraction or single-X-ray diffraction techniques) or structures similar to the one studied are already solved, existing algorithms can be applied to index single electron diffraction patterns. In these cases data extracted from existing databases is used as a basis for indexing of the patterns. The face that proves to have the best crystallographic match is deemed to be the correct face. Such principle is implemented in the programs PhIDO [5] of the package CRISP used for crystal phase (form) identification.

In fact, the indexing done with PhIDO is based on a comparison or matching of the original diffraction patterns with simulated patterns from a database of known structures (the database includes lattice parameters, crystal system and lattice type). The indexing done with EDiff is based on a comparison or matching of the diffraction

patterns with simulated patterns from a database of imaginary cells via an exhaustive search.

4.3.4 Crystal phase identification of lysozyme nano-crystals from electron diffraction data

Electron diffraction patterns collected from 3D lysozyme nano-crystals were analyzed with the software EDiff and CRISP. Based on the data analysis with EDiff, the crystal system was identified as orthorhombic with unit cell parameters approximately 31x52.5x89Å. Analysis of the data was performed in three steps: in the first step the diffraction pattern was centered and the central beam was removed, in the second step the autocorrelation patterns were calculated and in the last step the autocorrelation patterns were compared with simulated patterns (see Figure 4.14). The unit cell and crystal settings which proved to give the best match with the experimental diffraction pattern were further used for the indexing of the pattern.

Analysis of the diffraction data with CRISP confirmed the results inferred by EDiff in the case of well oriented patterns.

An example of the lattice refinement of a diffraction pattern from lysozyme performed with ELD (a sub-program of CRISP), is given in Figure 4.15. For each lattice, the length of the shortest two lattice vectors U and V (in $Å^{-1}$) and the angle between them was calculated. By comparison of the calculated d-spacings (U and V) and angles between them with a database of known substances (the database includes cell parameters of crystal forms of lysozyme reported by X-ray and found by EDiff), the crystal phase of lysozyme was identified as orthorhombic with a primitive Bravais lattice and cell parameters 31x52.5x89Å. Heavily mistilted diffraction patterns and patterns collected at high tilts could not be successfully processed with CRISP.

Examples of the [011] and [001] zone indexed are given in Figure 4.16. The indexing was done with EDiff and confirmed with CRISP.

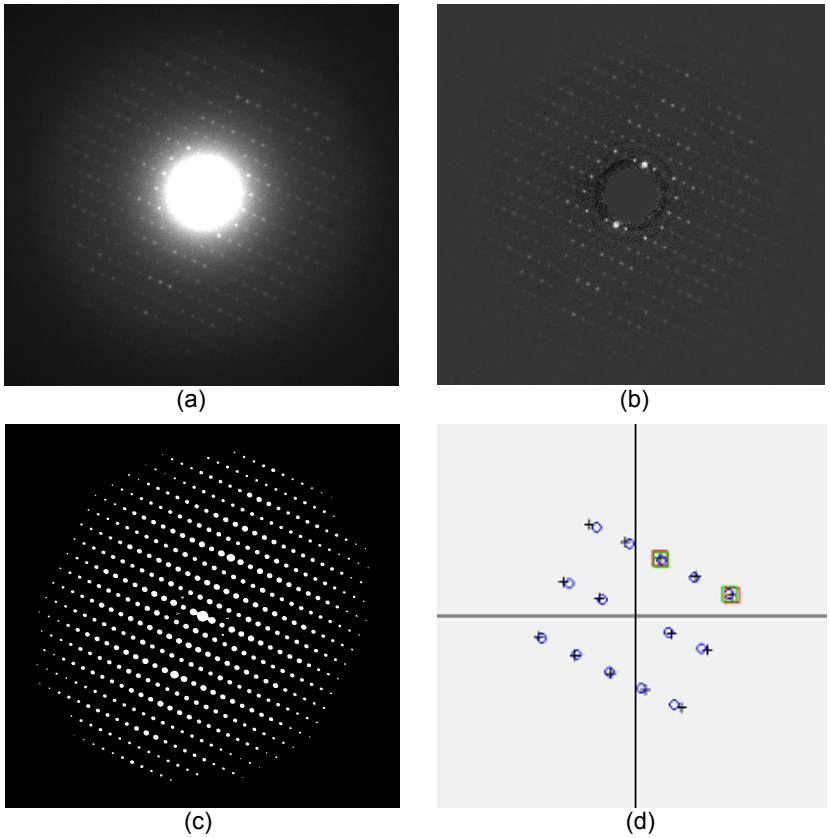


Figure 4.14 Electron diffraction pattern of a vitrified 3D lysozyme nano-crystal - (a); centering of the pattern and removal of the central beam - (b); calculated autocorrelation pattern - (c); matching of a simulated and an experimental pattern - (d).

4.4 Conclusions

The combination of plunge freezing vitrification techniques and cryo-electron microscopy provides a basis for collecting high resolution diffraction data from protein nano-crystals.

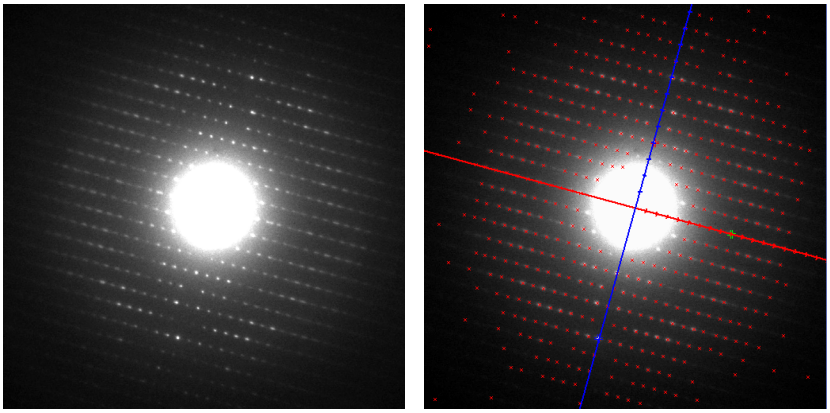


Figure 4.15 Diffraction pattern from a lysozyme nano-crystal - left image; lattice refinement performed with ELD - right image. The length of the two shortest vectors (given in blue and red) and the angle between them were calculated from the diffraction pattern. These three parameters were further used for crystal phase identification by comparison with a database of known substances.

Although X-ray single crystal diffraction is definitely the technique of first choice for structural studies of beam sensitive macro-crystals, electron diffraction was shown to have potential. Protein crystals of nanometer size which are too small for single crystal X-ray diffraction experiments, can be studied with electron diffraction techniques and yield high diffraction resolution.

With the development of precession electron diffraction dynamical diffraction is reduced, resulting in more quasi-kinematical data. Especially when protein crystals are thicker than 100 nm, this should facilitate the data analysis. Precessing the beam allows also integrated intensities to be collected in a single exposure.

The high beam sensitivity of proteins does not allow full 3D data collection from a single nano-crystal. Sets of diffraction patterns sets from many different crystals are needed in order to build up a complete dataset and solve the structure. Unit cell determination and indexing of protein electron diffraction patterns from different randomly oriented crystals is possible. Integration of the data is the next step, which is not trivial and requires the development of novel routines, or the modification of existing programs

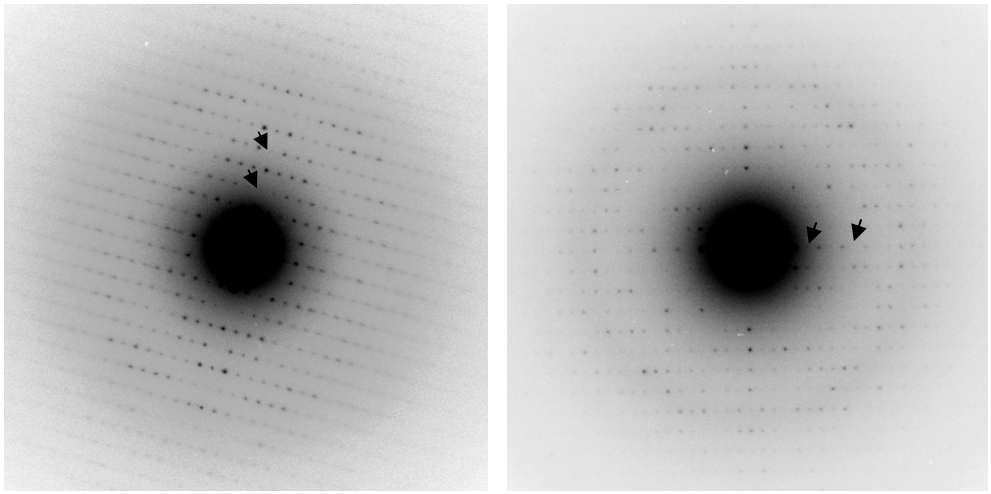


Figure 4.16 Electron diffraction patterns of vitrified 3D lysozyme nano-crystals: inset of diffraction pattern from (011) zone - left image and (001) zone - right image. The systematic absences shown with arrows indicate that the patterns do not suffer from strong systematic dynamic effects.

Acknowledgements The work was supported by a grant of FOM (Stichting voor Fundamenteel Onderzoek der Materie), The Netherlands. I would like to thank Linhua Jiang and Kim Ijsper for their work on the programs for data pre-processing and unit cell determination.

References

1. Vincent, R. and Midgley, P.A. (1994). *Ultramicroscopy*, **53**. 271.
2. Own, C.S. (2005). System design and verification of the precession electron diffraction technique, *PhD thesis*.
3. Jansen, J. and Zandbergen, H.W. (2002). *Ultramicroscopy,* **90** 291.
4. Zou, X., Hovmöller, A. and Hovmöller, S. (2004). *Ultramicoscopy,* **98**. 187.
5. PhIDO - Phase identification and indexing from ED patterns, *Calidris, Solentuna, Sweden*, 2001 www.calidris.em.com.

Chapter 5

High resolution electron diffraction of penicillin-type nano-crystals: 3D data collection and data analysis

(Testing the algorithm for unit cell determination from randomly oriented electron diffraction patterns)

Adapted from: Georgieva, D.G., Jiang, L., Zandbergen, H.W. and Abrahams, J.P. (2008). Unit cell determination from non-oriented electron diffraction patterns. *Acta Cryst. D64*. (in press)

Abstract

The structural studies of large organic pharmaceutical molecules are often frustrated by the same limitations as biological macro-molecules (proteins): difficulties in the growing of well diffracting macro-crystals limit the application of single-crystal X-ray diffraction techniques. Despite substantial progress, powder diffraction can not always provide complete crystallographic information and often fails for nano-systems (nano-particles, nano-crystals). Nevertheless, the problem of structure identification of pharmaceutical compounds is a pertinent one. Pharmaceuticals can exist in more than one polymorphic form. Differences in crystal form often result in different physical and chemical behavior and altered pharmaceutical activity. The methodologies that yielded promising results for obtaining electron diffraction data from protein nano-crystals were employed in the studies of single nano-crystals from microcrystalline powders of penicillin analogues. Diffraction resolution beyond 1Å could be obtained from single penicillin crystals using electrons. Based on the electron diffraction patterns, it was possible to identify the crystal form of the given pharmaceutical compound and to differentiate between pharmaceutical crystals and inorganic "impurity" crystals present

in the bulk. 3D electron diffraction data were acquired of penicillin G potassium and oxacillin sodium crystals and analyzed with the programs ELD and PhIDO (sub-programs of the software package CRISP). The results confirmed that the nano-crystals studied have the same unit cell parameters as those reported by X-ray crystallography of macro-crystals of the same compound. The electron diffraction data were further subjected to pre-processing which involved removing of the central beam and generating of autocorrelation maps. The latter were fed into the program EDiff (presented in Chapter 4) and used to determine the unit cell parameters from randomly oriented electron diffraction patterns of the two types of penicillin antibiotics. The cell dimensions reported by X-ray crystallography were confirmed by the newly developed algorithm for unit cell determination.

5.1 Introduction

To obtain structural information needed to understand the physical properties of large molecule pharmaceutical compounds can be a challenging task, especially when only sub-micron crystallites are available and considering that they are administered to patients in formulations which contain other bulk carrier compounds. Added to this is the high beam sensitivity of the pharmaceuticals which severely limits the time available for diffraction data collection.

Developments in instrumentation, computer technology and powder diffraction methodology contributed to the increased success rate in structures solved or polymorphic phases identified from powder samples. However, limitations inherent to the data quality often frustrate structural studies of organic molecules. It is relatively rare in powder diffraction, particularly when dealing with organic crystal structures, for good quality diffraction data to be obtained to atomic resolution. The presence of impurities and differences of size and shape of the crystallites in the bulk sample present additional problems and often create ambiguities in the interpretation of the data [1].

The essential difference between single-crystal and powder diffraction and also one of the intrinsic limitations of the powder diffraction technique is the loss of information resulting from the rotational projection of the three dimensional reciprocal lattice points on to the single (angular) dimension of a powder diffraction pattern. This effect is often aggravated in the case of organic materials by line broadening arising from

structural imperfections and poor crystallinity. The degree of reflection overlap becomes increasingly severe with increasing angle (the number of diffraction points scales with d^{*3} ($d^*=2\sin\theta/\lambda$)). Every discrete intensity y measured in a powder diffraction pattern at point $2\theta_i$ is actually a summation of the contribution of all Bragg reflections.

$$y(2\theta_i) = \sum_{hkl} H_{hkl}(2\theta_i)$$

Recovering the individual components $H_{hkl}(\theta_i)$ requires the deconvolution of a powder pattern into its individual reflection profiles. This explains why peak overlap may hinder data analysis, easily leading to misinterpretations.

The diffraction resolution especially at low angles is a factor which contributes significantly to the reliability of the indexing solution of a powder diffraction pattern. This can be understood from the definition of the figure of Merit M_{20} used to assess the physical plausibility of powder pattern indexing.

$$M_{20} = \frac{Q_{20}}{2 < Q > N_{20}}$$

Here $Q_{hkl} = \frac{10^4}{d_{hkl}^2}$; d_{hkl} is the interplanar spacing related to the diffraction angle by Bragg's law ($n\lambda=2\,d\sin\theta$); N_{20} is the number of different calculated Q values up to Q_{20}, which is the Q value for the 20th observed and indexed line; < Q > is the average discrepancy in Q for these 20 lines [1].

With higher resolution, more peaks may be detected, particularly for low crystal symmetry. Since most of the organic pharmaceutical compounds crystallize in low symmetry space groups, high requirements are put to the instrumental resolution and often access to short-wave sources (synchrotron radiation) is needed.

The design of a powder diffraction experiment regarding data collection is not trivial either and requires careful tuning of the experimental condition for each individual case [1]. Most of the factors involved in the experimental design (such as the choice of X-ray source and wavelength, optimum step width and intensity with which to sample the

powder pattern, the number of reflections to be collected etc.) need to be selected by the experimentalist and have the potential to affect the outcome of the data analysis. This makes the use of the technique rather difficult and requires the expertise of crystallographers who are highly specialized in powder diffraction.

Even when synchrotron X-ray sources are employed and the experimental design is set up for optimal diffraction data collection (providing high resolution patterns with well resolved peaks), X-ray powder diffraction cannot always yield the correct unit cell parameters when larger structures are studied (large unit cells create problems at the indexing stage) [1]. Moreover, the technique fails for materials if only nano-crystals are available. In this respect electron diffraction has certain advantages. By using electron sources 2D diffraction information can be obtained from single nano-crystallites. There are no intrinsic limitations of the technique which prevent studying structures with large unit cells or low symmetry. Since electron diffraction can be obtained from single nano-crystals the requirements to the amount and purity of the sample are also reduced.

In view of these advantages of electron diffraction and the fact that the size of the crystallites composing pharmaceutical powder samples can be down to $1\mu m^3$ or even smaller, we investigated whether electron diffraction might be used as an alternative technique for structural studies of organic pharmaceutical compounds. Three main issues are addressed in this chapter: The first one is related to the diffraction resolution which can be obtained from a single crystallite and how the quality of the diffraction data is dependent on the instrumental design. Another important point is whether the electron diffraction data can be correctly interpreted and the crystal phase identified. For the purpose, datasets from different penicillin-type crystals (penicillin G and oxacillin) were analyzed with the program PhIDO [2] and with the algorithm presented in the previous chapter (without using prior information about the unit cell). Last, since more than one crystal form can be present in a bulk powder sample (including different polymorphs or "impurity" crystals from the drug carriers), it is important whether the method can be used to differentiate between different crystal phases present in the same sample

5.2 Results and discussion

Five different types of penicillin were studied - amoxycillin, oxacillin, flucloxacillin, cloxacillin and penicillin G. The diffraction resolution obtained from single crystallites

from powder samples using electron diffraction is given in Table 5.1. The grade of crystallinity and the beam sensitivity of organic samples are definitely the factors which govern the quality of the diffraction information. However, provided the pharmaceutical nano-crystals are well ordered, diffraction resolution close to atomic can be obtained using electron sources. (see Figure 5.1).

Table 5.1 Electron diffraction resolution obtained from single penicillin-type nano-crystals at cryo-conditions.

Penicillin-type antibiotic	Electron diffraction resolution
Penicillin G potassium	0.9Å
Amoxicillin sodica sterile	1.1Å
Flucloxacillin sodium	1.1 Å
Cloxacillin sodium	1.1 Å
Oxacillin sodium	0.8 Å

The instrumental design of an electron diffraction experiment involves the choice of the electron source, diffraction mode, and detector. Experiments were performed with a thermionic (LaB6) and a field emission gun (FEG) microscope. FEG sources have definitely certain advantages over thermionic in terms of spatial coherence and current stability (see Table 5.2). However, diffraction data of sufficient resolution and good signal-to-noise ratio can be obtained also using LaB6 microscopes on organic compounds. This relaxes to a certain extent the requirements for the type of instrument. It has to be considered though that electron diffraction is a multiple scattering event and the intensities of the electron diffraction reflections oscillate with crystal thickness. Therefore, it is preferable if the diffraction pattern is taken from an area with uniform thickness. This would facilitate monitoring the multiple scattering in the later phase if atomic structure determination is the ultimate goal. By using a field emission gun a spot size down to 0.01μm can be obtained which allows the nano-diffraction mode (see Chapter 1) to be used for data collection.

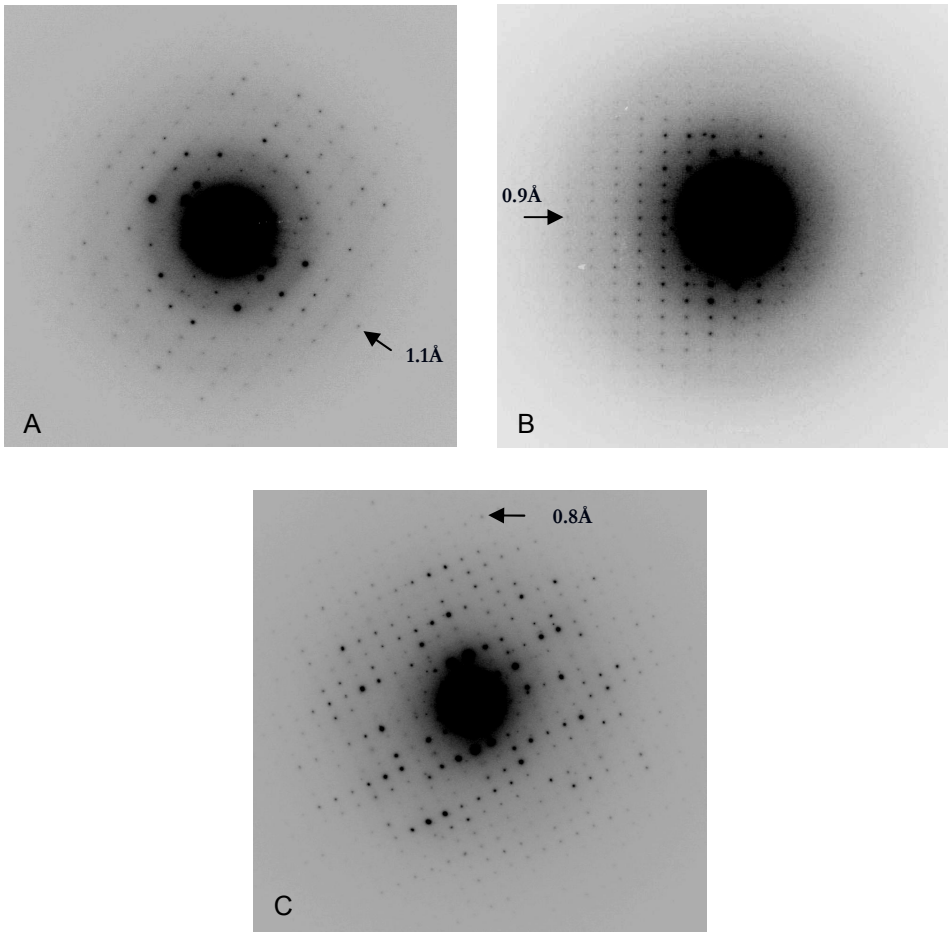


Figure 5.1 Electron diffraction patterns of amoxicillin sodica sterile - A, penicillin G potassium - B and oxacillin sodium - C. By using electron sources (LaB_6 source) and with the development of cryo-methods, diffraction information of high resolution and good signal-to-noise ratio can be obtained from pharmaceutical nano-crystals. The high beam sensitivity of the crystals does not always allow manual tilt in order to acquire an oriented diffraction pattern, precessing the beam (optically equivalent of precessing the crystal) was applied as a successful alternative in order to obtain relatively well oriented patterns (fig 5.2).

Table 5.2 Chararacteristics of the two principle electron sources operating at 100 kV. (From Williams & Carter, 1996).

	LaB_6	Field Emission
Crossover size (μm)	10	<0.01
Energy spread (eV)	1.5	0.3
Emission current stability (%/hr)	<1	5

It was confirmed also in the case of penicillin type crystals that by applying microdiffraction mode but with defocusing the beam on the sample (typically around 1μm) (the technique was used for protein diffraction data collection), high special resolution can be obtained from beam sensitive materials (see Figure 5.1). The experiments were performed at cryogenic conditions to increase the stability of the sample in the beam and to slow down phase transitions. Image plates were used as recording media in view of their high dynamical range.

Traditionally, in electron crystallography three-dimensional (3D) diffraction data are collected by manually tilting and orienting a crystal in order to obtain diffraction data from on-zone conditions [3, 4]. Due to the high beam sensitivity of the pharmaceutical crystals it is not always possible to orient the sample and if it is possible it occurs at the price of a decrease in resolution. By precessing the beam (optically equivalent to precessing the crystal) relatively well oriented high resolution diffraction patterns were collected (see Figure 5.2). In this case the intensities have been gathered from off-zone conditions (the precessing beam is entering the sample from an off-axis direction) [5]. However, the patterns may be indexed as conventionally obtained diffraction patterns. Another advantage of using precession electron diffraction is that the gathered intensities suffer less from dynamical perturbations [6]. This eased diffraction pattern recognition of data from thicker crystals. With increasing crystal thickness the dynamical effects become stronger, leading to the appearance of symmetry forbidden

reflections. In cases when the indexing of the zone based on the d-spacings is ambiguous, the interpretation of the symmetry of the pattern (e.g. presence of a screw axis etc.) is essential for determining the lattice type of the crystal, indexing of the patterns and identification of the crystal form.

One of the commercially available electron diffraction programs for indexing and crystal phase identification is PhIDO. Based on the d-spacings calculated from the electron diffraction patterns, the program assigns possible indices and crystal forms within a certain tolerance to a given pattern (zone) by comparing the calculated d-spacing with simulated ones from a database with known crystal forms. In cases when the electron diffraction patterns are heavily mistilted or contain reflections from different zones, it is impossible to give a certain index to the pattern and the crystal phase cannot be identified. By increasing the tolerance of indexing (respectively crystal phase identification) meaning that the calculated d-spacings are treated within a deviation which can be up to 9%, it is possible in principle to find a close zone. However, in practice a list of many matches is generated, including possibilities for different compounds and crystal forms which make it difficult to find the correct answer. This explains why one of the primary conditions for successful data analysis and crystal phase identification is to feed the program (if possible) with relatively well oriented electron diffraction patterns and therefore special care must be taken to acquire such patterns.

Electron diffraction data of penicillin G potassium sub-micron crystals collected in precession mode were subject first to analysis with the program ELD. In this step the crystal lattice of the patterns were refined (examples are given in Figure 5.2). Based on the crystal lattice refinement, the d-spacings were calculated from the electron diffraction patterns.

The electron diffraction patterns were further indexed with the program PhIDO (part of the software package CRISP). The main crystal form in the powder sample was identified as orthorhombic with a primitive Bravais lattice and unit cell parameters a= 6.3 Å, b=9.3 Å, c=30 Å (a tolerance of 2% was used for crystal phase identification). The same crystal form was reported from single-crystal X-ray diffraction of potassium penicillin G macro-crystals [7]. The absences of kinematically forbidden reflections (dynamically allowed) on the electron diffraction patterns were also taken into account for the final identification of the crystal form. They were found to be consistent with the obtained indexing solutions from the program PhIDO for the given crystal form.

This led to the conclusion that the penicillin G nano-crystals from the powder sample studied are of the same crystal form as reported by X-ray crystallography.

The electron diffraction data of penicillin G potassium were analyzed also with the program EDiff, which allows unit cell parameters to be determined "*ab initio*" from electron diffraction patterns collected from randomly oriented crystals. In electron crystallography, unit cell parameters are determined by tilting a crystal along a certain crystallographic zone (preferably a main zone) and collecting a tilt series of diffraction patterns. However, the high beam sensitivity of organic nano-crystals does not always allow multiple patterns to be collected from a single crystal which has limited the application of electron diffraction for studying beam sensitive materials. The major advantage and the novel part of the algorithm implemented in the program EDiff (the algorithm is described in Chapter 4) is that the knowledge of the angular relationship between the different electron diffraction patterns is not required and therefore diffraction data from different crystals can be used.

Before analyzing the electron diffraction data of penicillin G potassium with the program EDiff, the patterns were subject to a number of corrections. At this stage, the central beam was removed, the preliminary centre of the patterns was found and autocorrelation maps were generated. The autocorrelation maps were further used for unit cell determination in the program EDiff.

Three different search modes are implemented in EDiff. For employing the Main Vector Search mode the two shortest vectors need to be selected manually from the autocorrelation maps. The selected main vectors are than fitted to all the possible unit cell parameters that the program tests. Therefore, for finding the correct indexing solution (respectively unit cell dimensions), using this search mode it is vital that the main vectors are selected correctly. It is also possible to mark the autocorrelation maps as "important", "good", "normal" and "bad" giving more preference or weight to some patterns over others. This is important when the quality of the autocorrelation maps does not allow finding or selecting the shortest vectors. On the other hand this pre-selection of the data may introduce also certain bias to the analysis influencing the outcome. Therefore the estimation of the data needs to be done with special caution.

In the other two search modes - Full Vector Matching and the Unique Facet Matching - the vectors are selected by the program. No manual input in this respect is required.

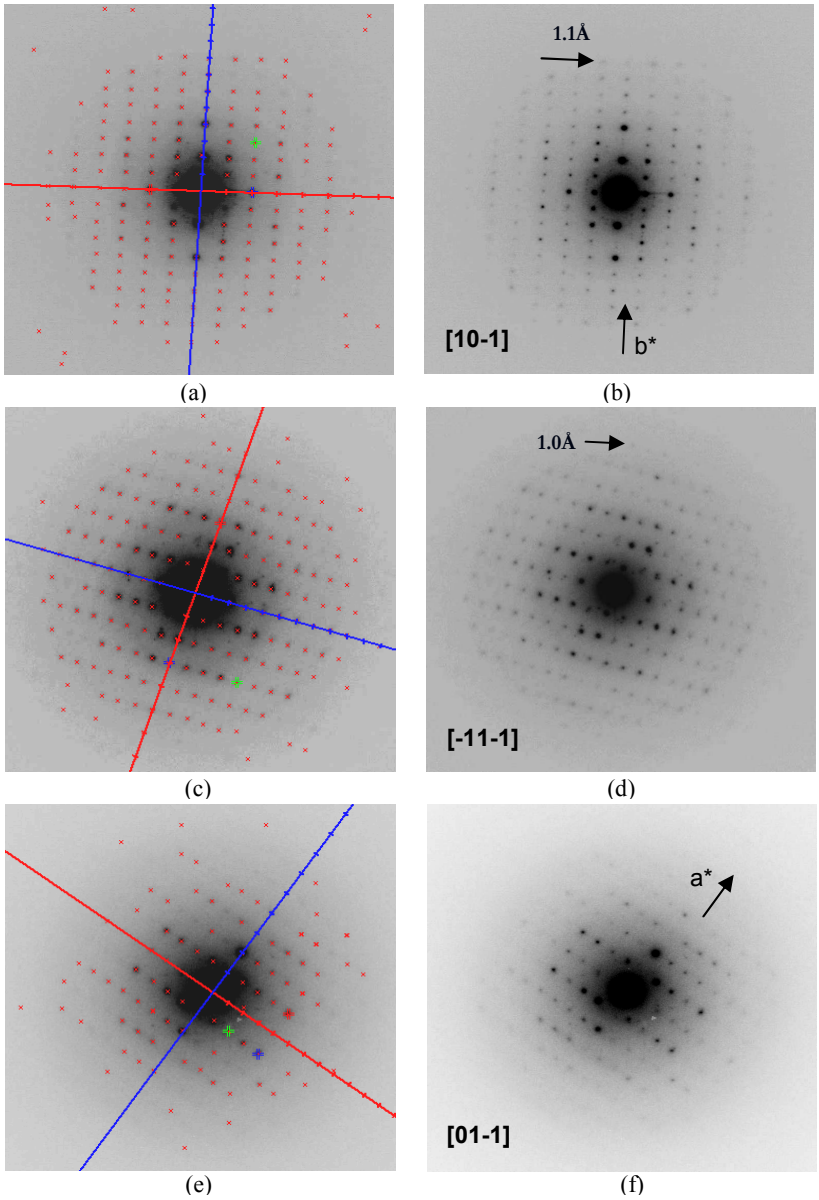


Figure 5.2 Electron diffraction patterns of single nano-crystals of penicillin G potassium acquired in precession mode. Based on the electron diffraction patterns the main crystal form in the powder sample was identified as orthorhombic with a primitive Bravais lattice

and unit cell parameters 6.3x9.3x30Å. The same crystal phase was reported from single-crystal X-ray diffraction of potassium penicillin G macro-crystals. The electron diffraction data were analyzed with the software CRISP. The lattice refinement and the estimation of the intensities were performed with ELD (a sub-program of CRISP). The diffraction patterns were indexed with the program PhIDO (a sub-program from the CRISP package). For diffraction pattern 5.2(b) the following indices were assigned: [101], [-10-1], [10-1], [-101]; for diffraction pattern 5.2(d): [111], [-11-1], [1-1-1], [-1-11] and for diffraction pattern 5.2(f): [011], [0-1-1], [0-11], [01-1]. A tolerance of 2% was used for crystal phase identification. The indices given on the diffraction patterns were confirmed with the program EDiff.

Another difference compared to the Main Vector search mode is that not always the shortest vectors from the autocorrelation maps are selected. In the Full Vector search mode, the program is considering all the possible vectors calculated from the maps as well as the possible combinations between them. In the Unique Facet Matching, a search is performed for finding the "unique facet" which consists of two vectors (not necessarily the shortest) and the angle between them.

In total thirteen diffraction patterns of penicillin G potassium acquired at different crystal settings were used for the unit cell determination. The data were analyzed with the three search modes. The autocorrelation maps generated from the electron diffraction patterns were of good quality, allowing fairly easy determination of the main vectors. Therefore, the data were subjected first to analysis with the Main Vector Matching mode (the search with this mode is much quicker compared to the other two). The initial search range was set as given in Table 5.3. The results obtained are presented in Table 5.4. In the next step, the Full Vector Matching and Unique Facet Matching search modes were employed. In this case a narrower search range was used (see Table 5.5) and the search step size was also reduced. The results are given in Table 5.6 and Table 5.7.

As a result of the analysis performed with the three search modes, possible unit cell parameters are generated and also indices are assigned to the diffraction patterns. Examples of the indexing solution obtained with the Full Vector Matching mode of some of the patterns of penicillin G potassium are presented in Figure 5.3. The root mean square error (RMSE) for the different zones (diffraction patterns) is given in Table 5.8. The indexes for the well oriented zones were verified also with the program PhIDO.

In the case of penicillin G potassium, unit cell parameters consistent with those reported by X-ray crystallography were obtained using all the three different search modes implemented in EDiff.

Table 5.3 Initial search range for finding the unit cell parameters of penicillin G potassium from randomly oriented electron diffraction patterns.

Search mode	a (Å)	b (Å)	c (Å)	angles	step size
Main Vector Matching	3-30	3-30	3-30	3x90°	0.5

Table 5.4 Output of the initial search performed with the Main Vector Matching.

Search mode	a (Å)	b (Å)	c (Å)	angles
Main Vector Matching	6.5	9	29.5	3x90°
	6.5	9	30	3x90°
	6.5	9	29	3x90°
	6.5	9	28.5	3x90°
	6.5	9	28	3x90°

Table 5.5 A narrow search range used for finding the cell parameters of penicillin G potassium using the Full Vector Matching and the Unique Facet Matching. In this case the search step size was also smaller.

Search mode	a (Å)	b (Å)	c (Å)	angles	step size
Full Vector Matching &Unique FacetMatching	5-8	8-10	28-31	3x90°	0.1

Table 5.6 Output of the search range performed with the Full Vector Matching, sorted on quality of RMSE 's.

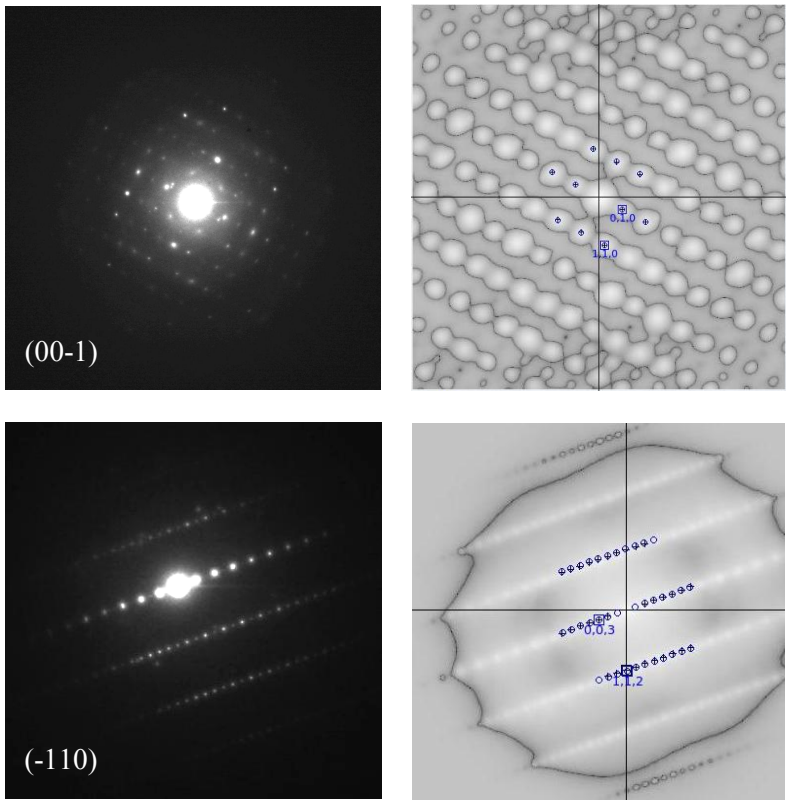
Search mode	a (Å)	b (Å)	c (Å)	angles
Full Vector Matching	6.3	9.3	29.8	3x90°
	6.3	9.3	29.7	3x90°
	6.3	9.3	29.6	3x90°
	6.3	9.3	29.9	3x90°
	6.4	9.3	30.8	3x90°

Table 5.7 Output of the search range performed with the Unique Facet Matching**.**

Search mode	a (Å)	b (Å)	c (Å)	angles
Unique Facet Matching	6.4	9.4	30	3x90°
	6.4	9.4	29.9	3x90°
	6.4	9.4	30.1	3x90°
	6.4	9.4	30.8	3x90°
	6.4	9.3	30.8	3x90°

For determining the unit cell parameters of penicillin G potassium different low-index main zones such as [001], [011], [101] and [110] were used, meaning that information about the three cell parameters was well expressed in the data. However, in practice it is not always possible to acquire all low index zones by random data collection. Moreover, crystals often show texture or preferred crystal orientation, which may frustrate electron diffraction data collection. This is one of the factors which impede electron diffraction data collection of full 3D data. In this case information about the third unit cell dimension is usually more difficult to obtain. The presented algorithm for unit cell determination was also tested with data from oxacillin sodium. In this case

low index main zones such as [011] were also present in the data. However, most of the diffraction patterns were acquired from zones such as [021], [212] (see Figure 5.4). The data collected from oxacillin sodium sub-micron crystals were subject to the same analysis procedure as penicillin G potassium. In total eleven electron diffraction patterns at different crystal settings were used for the analysis. The electron diffraction patterns were first indexed with the program CRISP to confirm that the nano-crystals studied are of the same crystal form as reported in the literature. Further, the electron diffraction patterns went through the pre-processing steps discussed previously. Autocorrelation maps were generated and used for the determining the unit cell dimensions



(see next page)

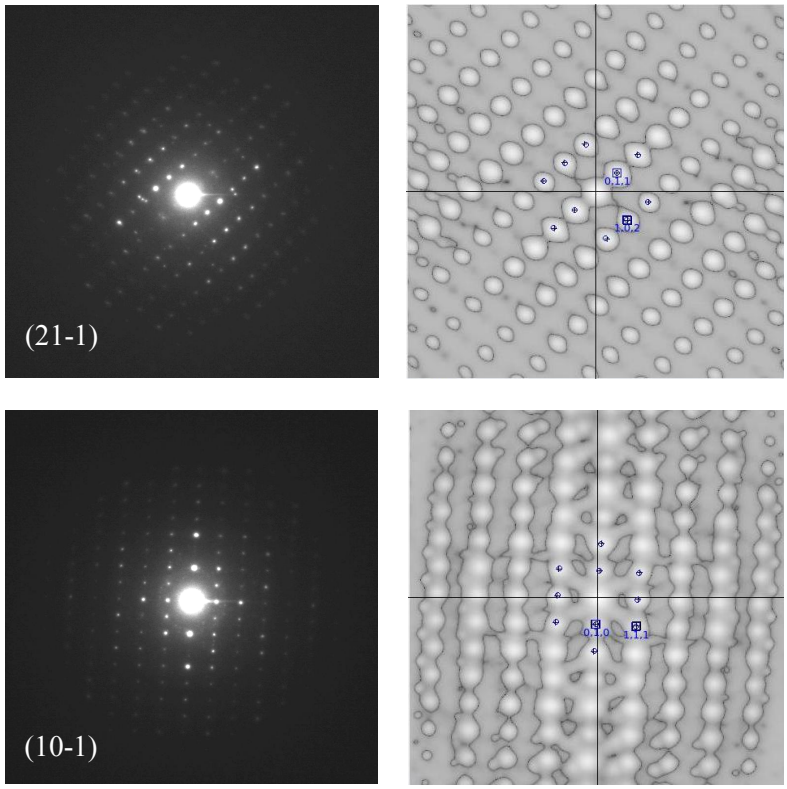


Figure 5.3 Electron diffraction patterns acquired from different crystallographic zones of penicillin G potassium nano-crystals - (left column of images). Next to the experimental diffraction patterns the calculated autocorrelation images are given. With crosses are indicated the peak position on the autocorrelation image and with circles the peak position in the simulated diffraction pattern. The root mean square error (RMSE) of the experimental and simulated patterns for the different zones (diffraction patterns) are given in table 5.8.

The data of oxacillin sodium were analyzed with the three search modes and the unit cell parameters calculated are presented in Table 5.9. The root mean square error (RMSE) for the different zones (diffraction patterns) is given in Table 5.10. In the case of oxacillin sodium the indexing of the diffraction patterns (see Figure 5.4) was also verified with the program PhIDO.

Table 5.8 The root mean square error (RMSE) for the different zones (diffraction patterns) given in fig. 5.3.

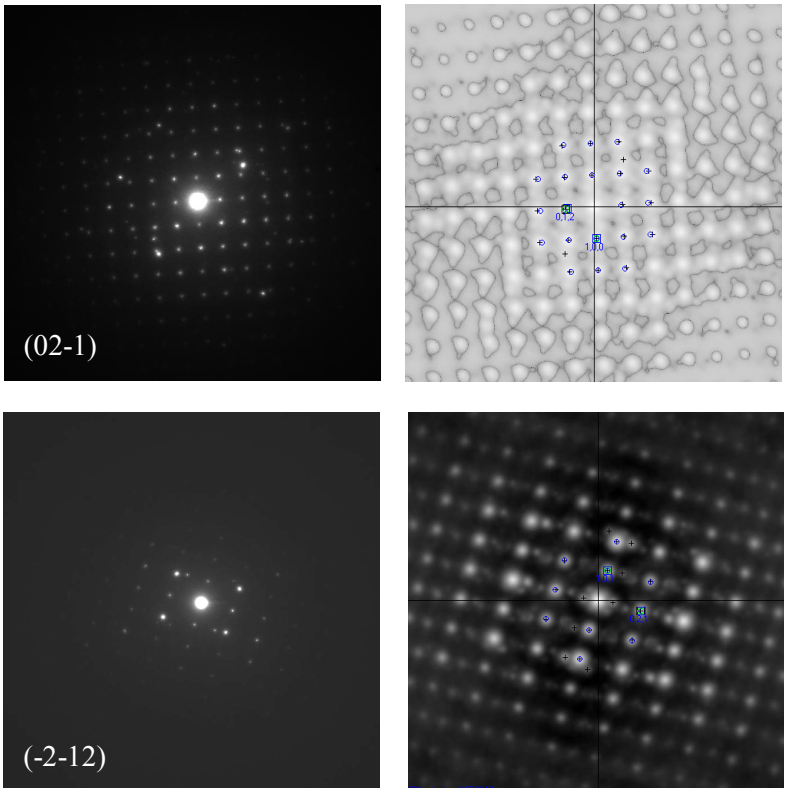
zone	RMSE (%)	angular error
(00-1)	0.61	0.4
(-110)	0.70	0.01
(21-1)	1.40	0.17
(10-1)	1.63	1.45

Table 5.9 Unit cell parameters of oxacillin determined by single-crystal X-ray diffraction and electron diffraction of single nano-crystals from a powder sample using the EDiff program.

	a (Å)	b (Å)	c (Å)	angles
X-ray diffraction	7.3	10.3	26.7	3x90°
Unique Facet Matching	7.3	10.1	27	3x90°
Full Vector Matching	7.3	10.5	27	3x90°
Main Vector Matching	7.4	10.2	27	3x90°

In both cases of penicillin G potassium and oxacillin sodium oriented and misoriented electron diffraction patterns were used in the search for the unit cell parameters. It is definitely helpful if most of the low-index zones are present in the dataset (as in the case of penicillin G potassium) for determining the correct cell dimensions from different randomly oriented nano-crystals with the program EDiff. However, good results could also be obtained in the case of oxacillin sodium where mostly only one

main zone was present in the dataset. All three search modes implemented in the program EDiff and used for determining the cell dimensions in the cases studied were found to be very sensitive to the quality of the electron diffraction patterns respectively the autocorrelation maps. However, the Main Vector Matching algorithm enables the user to select the two main vectors (the spots closest to the centre). Therefore, this search mode should theoretically give the most accurate and quick determination.



(see next page)

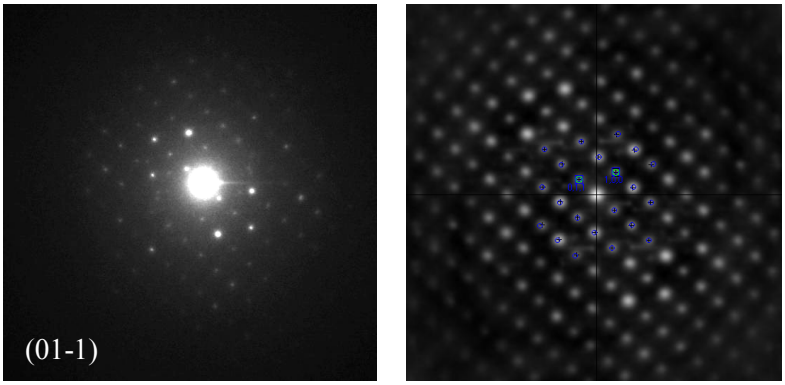


Figure 5.4 Electron diffraction patterns acquired from different crystallogarphic zones of oxacillin nano-crystals - (left column of images). Next to the experimental diffraction patterns the calculated autocorrelation images are given. With crosses are indicated the peak position on the autocorrelation image and with circles the peak position in the simulated diffraction pattern. The root mean square error (RMSE) of the experimental and simulated patterns for the different zones (diffraction patterns) are given in table 5.10.

Table 5.10 The root mean square error (RMSE) for the different zones (diffraction patterns) given in fig. 5.4.

zone	RMSE (%)	angular error
(02-1)	3.23	0.20
(-2-12)	0.91	0.90
(01-1)	1.71	0.52

The Unique Facet and Full Vector matching algorithms search for different facets automatically. Errors made in the autocorrelation mapping may cause these two algorithms not to select the facets correctly and therefore not to yield the most accurate solution, but at least they give a good estimate. A criterium which can be employed to evaluate the performance of the program and the correctness of the solution found is

the consistency of the unit cell parameters determined by all three search modes. The program needs to be further tested for systems of lower symmetry than the orthorhombic such as monoclinic and triclinic.

On the basis of the electron diffraction patterns (d-spacing, symmetry and signal-to-noise ratio), it was also possible to differentiate between penicillin G crystals and "foreign" crystals (crystals not sharing the same crystallographic features as the penicillin crystals) that were present in the sample (see Figure 5.5). This showed that electron diffraction can be used as a diagnostic method to identify impurities in pharmaceutical samples (even if they are present in very small quantities since individual crystallites can be studied) including polymorphic transitions, should these occur.

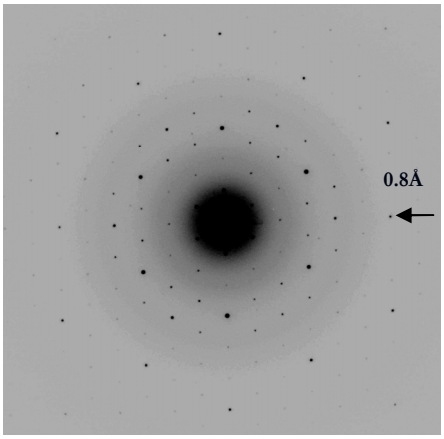


Figure 5.5 Based on the different (higher) crystallographic symmetry and the smaller lattice spacing compared to the lattice parameters of penicillin G potassium (6.3x9.3x30Å, orthorhombic crystal system), a second crystal phase was also identified in the sample. The high beam resistance of the crystals from the second crystal phase indicated that they are of inorganic origin and do not belong to the penicillin type. This shows that electron diffraction can be used to differentiate between different crystal forms in pharmaceutical samples and to identify "impurity" crystals even if they are in small quantities.

5.3 Conclusions

Using electron sources and with the developments in cryo-methods, high spatial diffraction resolution can be obtained from single beam-sensitive pharmaceutical nano-crystallites. Since individual crystals can be studied, the requirements to the amount of the sample are much reduced (compared to powder diffraction techniques). The experimental design for electron diffraction data collection is straightforward and does not need to be adjusted for each case individually. Crystal thickness and the related to this problem of multiple scattering of electrons often frustrate data analysis. The precession electron diffraction technique offers a partial solution to the problem of dynamical scattering. Precessing the beam allows also relatively well oriented diffraction patterns to be acquired (if not too far from a main crystallographic zone). This facilitates the automated indexing of the data and crystal phase identification by existing program (such as PhIDO). New EM software allows unit cell dimensions to be determined within a certain tolerance from randomly oriented electron diffraction patterns acquired from different crystals, without initial knowledge about the angular relationship between the patterns. Although X-ray single crystal and powder diffraction are definitely techniques of first choice for structure determination in the solid state, electron diffraction was shown to have potential for studying beam sensitive pharmaceutical compounds of which only sub-micron crystals are available.

References

1. David, W.I.F., Shankland, K., McCusker, L.B. and Baerlocher, Ch. (2002). Structure determination from powder diffraction data*, IUCR monographs on crystallography 13, Oxford University Press Inc., New York.*
2. PhIDO - Phase identificarion and indexing from ED patterns*, Calidris, Solentuna, Sweden*, 2001 www.calidris.em.com.
3. Zou, X., Hovmöller, A. and Hovmöller, S. (2004). *Ultramicoscopy*, **98**. 187.
4. Li, X.Z. (2005). *Ultramicroscopy,* **103**. 269.
5. Own, C.S. (2005). System design and verification of the precession electron diffraction technique, *PhD thesis*.
6. Own, C.S., Marks, L.D. and Sinkler, W. (2006). *Acta Cryst*. D**62**, 434.
7. Dexter, D.D. and van der Veen, J.M. (1978). *J. Chem. Soc. Perkin Trans.* **1**. 185.

Chapter 6

Summary, conclusions and future perspectives

Without understanding the mesoscopic phase behavior of proteins, the development of new products in biotechnology is largely a matter of hit-and-miss. The aberrant aggregation behavior of proteins does not only limit production of commercially important proteins, but has also major consequences in entirely different fields, ranging from human pathology to biochemical research. For instance, existing methods of recovering native protein from aggregates are often inefficient and time-consuming. Aggregation has also been identified as the cause of numerous human diseases, like cystic fibrosis, Alzheimer's disease, and prion spongiform encephalopathies such as Creutzfeldt-Jacob disease.

Visualization of bio-molecules to establish their 3-D conformation and their atomic composition is indispensable in research that aims at the prediction of the phase behavior of such molecules. In cases when macro-crystals can be obtained, X-ray crystallography is undoubtedly the technique of first choice. However, most of the proteins are difficult to crystallize and there are no definite rules on determining the crystallization conditions.

This thesis aims at developing the technique of electron diffraction for structural studies of protein and other beam-sensitive molecules from which only small 3D crystalline volumes can be obtained.

In the first part of the study protein nano-crystals of well known proteins (lysozyme and glucose isomerase) as well as less studied proteins such as potato protease inhibitor were obtained using heterogeneous substrates. There are several practical and theoretical considerations for choosing the heterogeneous over the homogeneous crystallization approach. By definition heterogeneous crystallization is a process of preferential nucleation, for instance crystallization of proteins on an introduced heterogeneous substrate. This gives the possibility not only to grow protein crystals at a lower protein concentration but also to study and control better the process of

crystallization (since it can be predicted where it will occur) which is otherwise more difficult in the bulk crystallization solution. Additionally, growing of protein nano-crystals on a surface offers certain practical advantages related to the visualization, recognition and handling of the crystals.

Electron diffraction data collection was performed at cryo-conditions. Image plates were used as recording media because of their high dynamic range (compared to CCD camera). This allowed strong first order reflections as well as very weak high resolution reflections to be recorded. It was also experimentally found that by using microdiffraction but with a defocused beam in image mode, high spatial diffraction resolution could be obtained. When the beam is defocused on the specimen in image mode, in diffraction mode the pattern is formed not at the back focal plane but slightly above or below. In this case the beam is also more parallel which results in obtaining of spots in stead of discs. Spots are generally easier for detection and analysis especially when the diffraction signal itself is weak (the light atoms composing the protein crystals don't scatter electrons strongly). Moreover, when the beam is defocused on the specimen, more atom planes contribute to the formation of the diffraction pattern which can also explain why the diffraction signal is stronger and high frequency diffraction reflections can be detected. The presented mode of data collection is not conventional and it needs still to be explored further. However, the initial results are quite promising since this mode provides a possibility to acquire high resolution diffraction information from highly beam sensitive materials (no additional low dose software is required). The mode can be used also in combination with the precession diffraction technique where the beam is conically tilted during data recording and therefore integrated intensities are gathered.

Since it was not possible with the current instrumentation to collect a 3D dataset from a single protein nano-crystal, diffraction data from different crystals were needed. An algorithm for unit cell determination from randomly oriented electron diffraction patterns is presented also.

The algorithm was tested further with electron diffraction data collected from two different types of penicillin crystals. Organic pharmaceutical crystals (in this case penicillin-type) suffer from the same limitations regarding data collection as protein nano-crystals. The high beam sensitivity of the penicillin crystals doesn't allow collection of a 3D dataset from a single nano-crystal. However, in the case of penicillin, nano-crystals from a powder sample could be studied with electron diffraction. No further re-crystallization was required (re-crystallization always risks getting a different

polymorph form). The unit cell parameters obtained from the electron diffraction data and analyzed with the algorithms presented in both cases of penicillin G and oxacillin were consistent with the parameters reported earlier, obtained by X-ray crystallography.

Conclusions and future perspectives: The studies carried out and described in this thesis show that electron diffraction has the potential to enable structural studies on protein and organic pharmaceutical nano-crystals.

Provided that the crystals are well ordered, electron diffraction information of sufficient quality (resolution and signal-to-noise ratio) can be obtained with the current methods of preparation and instrumentation. The fact that the quality of the diffraction data, was largely improved by optimizing the mode of data collection and data recording, shows that with the development of the automated modes of data collection in electron microscopy such as tomography as well as the new generation detectors with zero background noise such as the medipix detector, the limits can be stretched further. Moreover, the development of the instrumentation will also allow a quicker and more elegant way of data acquisition which will speed the process of structure determination using electron diffraction.

The beam sensitivity of the materials studied is currently one of the rate-limiting step for their structure determination. The unit cell determination is the first step towards crystal phase identification and structure determination of an unknown crystal form. The algorithm for unit cell determination presented in this thesis can deal with diffraction patterns acquired from different nano-crystals. The method was tested successfully with datasets from structures with high symmetry (from inorganic materials, not discussed in this thesis) and low symmetry structures (from pharmaceutical crystals).

Integration, scaling and merging of the intensities are subsequent steps towards structure determination at atomic level. The complexity of the scattering and in particular the dynamical scattering is of special concern when electron diffraction data are to be analyzed. Precessing of the beam during data acquisition, is one of the approaches used to reduce systematic and non-systematic multiple scattering. Analysis of the protein electron diffraction intensities obtained with and without precession needs to be performed further, this is one of the future directions of development.

The research performed and described in this thesis has a mainly fundamental character and is aimed at getting a better insight into the mechanism of heterogeneous nucleation,

reproducible growth of sub-microcrystal and their diffraction studies with electron sources. Some of the methodologies developed in the course of this study found also directly practical applications for instance, in the case of pharmaceutical nano-crystals. The problem of polymorphism and purity of pharmaceutical compounds is and will remain of special concern in the pharmaceutical industry. Electron diffraction has some advantages over techniques such as X-ray powder diffraction and microcalorimetry regarding the amount of sample required for the study. The technique allows also in a relatively fast way impurities (for instance of inorganic origin) to be identified based on their different crystallographic features. With the development of precession electron diffraction, relatively well oriented high resolution diffraction patterns from beam-sensitive pharmaceutical nano-crystals can be acquired (without additional tilting of the crystals). This facilitates indexing of the patterns and crystal phase identification with existing program such as CRISP (in which prior information about the unit cell is required) or without using prior information as described in Chapters 4 and 5 of this thesis.

Nederlandse samenvatting

Zonder begrip van het mesoscopische gedrag van eiwitten is de ontwikkeling van nieuwe producten in de biotechnologie grotendeels een kwestie van trial en error. Het afwijkende aggregatiegedrag van eiwitten stelt niet alleen een limiet aan de productie van commercieel belangrijke eiwitten, maar heeft grote gevolgen in totaal verschillende velden, van menselijke pathologie tot biochemisch onderzoek. De bestaande methoden om natieve eiwitten terug te winnen uit aggregaten zijn vaak inefficiënt en tijdrovend. Aggregatie is bekend als de oorzaak van talrijke menselijke ziektes zoals cystic fibrosis, ziekte van Alzheimer en prion spongiform encephalopathie zoals de ziekte van Creutzfeldt-Jacob.

De visualisatie van biomoleculen om hun 3-D structuur en hun atomaire samenstelling te bepalen is onontbeerlijk in onderzoek dat tot doel heeft het gedrag van dergelijke moleculen tevoorspellen. In gevallen waarin macrokristallen verkregen kunnen worden is Röntgen kristallografie zonder twijfel de eerste keuze. De meeste eiwitten zijn echter moeilijk te kristalliseren en er zijn geen vastomlijnde regels om de juiste voorwaarden voor kristallisatie te bepalen.

Dit proefschrift behandelt het ontwikkelen van de techniek van elektronen diffractie voor structurele studies van eiwitten en andere stralingsgevoelige moleculen waarvan slechts kleine 3D kristallen verkregen kunnen worden.

Het eerste deel van het onderzoek betreft de proteïne nanokristallen van algemeen bekende eiwitten (lysozyme en glucose isomerase), alsmede van minder bestudeerde eiwitten als aardappel protease inhibitor verkregen met behulp van heterogene kristallisatie. Er zijn verscheidene praktische en theoretische afwegingen om de heterogene aanpak boven de homogene benadering te verkiezen voor het kristalliseren. Heterogene kristallisatie, bijvoorbeeld kristallisatie van eiwitten op een ingebracht heterogeen substraat, is per definitie een proces van voorkeurs nucleatie. Dit geeft de mogelijkheid om niet alleen eiwitkristallen bij een lagere concentratie van eiwitten te verkrijgen, maar ook om het kristallisatieproces beter dan in een bulk kristallisatieoplossing te kunnen regelen en bestuderen. Bovendien geeft het aangroeien van eiwit nanokristallen op een oppervlak zekere praktische voordelen ten aanzien van visualisatie, herkenning en behandeling van de kristallen.

De verzameling van elektronendiffractie is uitgevoerd onder cryo-condities. Image plates werden gebruikt om de patronen te registreren (vanwege het hoge dynamische bereik van image plates in vergelijking met een CCD camera). Daardoor was het mogelijk zowel sterke, lage resolutie reflecties als zwakke, hoge resolutie reflecties te registreren. Eerder is experimenteel vastgesteld dat door middel van microdiffractie, maar dan met een gedefocuseerde elektronen bundel in "image mode", hoge diffractie resolutie verkrijgbaar is. Als de elektronen bundel is gedefocuseerd op het preparaat in "image mode", wordt het diffractie patroon niet exact geproduceerd in het zogenaamde "back focal plane" maar enigszins er boven of beneden. In dit geval is de bundel ook weinig divergent en als resultaat worden diffractie punten verkregen in plaats van schijfvormige reflecties. Diffractie punten zijn, in het algemeen, makkelijker te registreren en analyseren, vooral als het diffractie signaal zwak is (zoals bij de bestudering van eiwit kristallen). Bovendien, als de bundel gedefocuseerd is op het preparaat, nemen meer kristal lagen deel aan het ontstaan van een diffractie patroon. Dit feit kan ook verklaren waarom nu het diffractie signaal sterker is en hoge resolutie reflecties gemeten kunnen worden. De besproken manier van data collectie is geen routine en zij moet nog nader worden onderzocht. Echter, de eerste resultaten zijn veelbelovend, gezien de mogelijkheid om hoge resolutie informatie te verkrijgen van stralingsgevoelige nanokristallen. De besproken manier van data collectie kan ook gebruikt worden in combinatie met precessie diffractie technieken waarbij de bundel op het kristal precedeert tijdens de data collectie. Daardoor worden geïntegreerde intensiteiten geregistreerd.

Aangezien het met de huidige instrumentatie niet mogelijk is om een 3D dataset van één enkel eiwit kristal te registreren, werden diffractie patronen afkomstig van verschillende kristallen geregistreerd. Een algoritme om *ab initio* de cel dimensies te bepalen uit diffractie patronen afkomstig van willekeurige georiënteerde kristallen is voorgesteld en besproken.

Het algoritme is getest met elektronen diffractie data van twee verschillende types penicilline. De cel dimensies verkregen uit the elektronendiffractie data en geanalyseerd met het besproken algoritme in zijn beide gevallen overeenkomstig met de cel dimensies verkregen uit Röntgen kristalgrafisch onderzoek.

Curriculum Vitae

Dilyana Georgieva was born on March 5, 1980 in Sofia, Bulgaria. In 1998, she completed a secondary Spanish Language School in Varna, Bulgaria. Between 1998 and 2004, she studied Pharmacy at the Medical University of Sofia. After a specialization in drug delivery systems at the University of Antwerpen, a thesis project on ciprofloxacin nanoparticles for ophtalmic use, she received a Master Degree in 2004 from the Medical University of Sofia.

Shortly after, she started a PhD project on protein nano-crystallography at the Department of Biophysical Structural Chemistry, University of Leiden, the results of which are summarized in this thesis.

Currently, she is working as a post-doctor at the University of Leiden, carrying out research on electron crystallography of proteins and pharmaceuticals.

Acknowledgements

I would like to end this thesis by expressing my gratitude to all the people who contributed to the success of this work and more importantly supported me and helped me during the last four years.

I am very much indebted to Prof. Dr. Henny Zandbergen for his support, encouragement and guidance during the last few years. Furthermore, I am very grateful to him for his advices, for teaching me electron microscopy and for the patience with which he was listening and helping me to solve scientific and technical problems. He gave me also the opportunity to explore and apply new techniques such as the precession electron diffraction technique in the field of protein nanocrystallography

I would like to thank Dr. Stavros Nikolopoulus for the opportunity to learn, work and understand better precession electron diffraction. The discussions with him inspired me to pursue the development of electron diffraction and to apply the method not only for protein crystals but also for pharmaceuticals. I would like to thank Stavros for all his help throughout my PhD and for his kind friendship. I really enjoyed my time with him.

Sincerely thanks to Dr. Rag de Graaf for his guidance, fruitful discussions and for reading my manuscript.

I am particularly grateful to Dr. Tjerk Oosterkamp and his students for teaching me and helping me with the Atomic Force Microscopy experiments.

I am very thankful to all my colleagues from Leiden for their support: Raj, Irakli, Pavol, Daniel, Ellen and Jan. I would like especially to thank Linhua who worked on the development of the software for the unit cell determination. I appreciate enormously his help. I was really fortunate to work with him and deeply impressed by his knowledge.

I would like to thank my colleagues from Delft: Vikas, Alex, Mauro, and Qiang for being not only excellent colleagues but also for being my friends, for their help and support and for the nice moments we had together. I really enjoyed my time in Delft.

Special thanks to my friends: Lidiya, Nadja, Svetla, and Eleni, who supported me and believed in me.

Last but not least I would like to thank the people I love the most: my parents, Alexander and my cousins Boyan and Tzvetelina. Without their love and support I couldn't be what I am now.

To those who didn't feel appreciated or whose names I forgot to mention I apologize.

Dilyana Georgieva

Leiden, 2008

List of publications

Georgieva, D.G., Kuil, M.E. and Abrahams, J.P. (2006). Protein nanocrystallization. *Springer series in Biophysics, Advanced Techniques in Byophysics, Springer-Verlag Berlin Heidelberg*.

Georgieva, D.G., Kuil, M.E., Oosterkamp, T.H., Zandbergen, H.W. and Abrahams, J.P. (2007). Heterogeneous crystallization of protein nano-crystals. *Acta Cryst*. D63. 564.

Georgieva, D.G., Jiang, L., Zandbergen, H.W. and Abrahams, J.P. (2008). Unit cell determination from non-oriented electron diffraction patterns. *Acta Cryst*. D64, (in press)

papers in preparation

Georgieva, D.G., Zandbergen, H.W., Nicolopoulos, S., Sarakinou, E. and Abrahams, J.P. "Optimal data collection of beam sensitive 3D crystals of proteins and pharmaceuticals by electron diffraction"

Georgieva, D.G., Jiang, L., Hoedemaker, F., Zandbergen, H.W. and Abrahams, J.P. "Identification and analysis of a new polymorph of mianserin using electron diffraction"

Presentations

Lunteren crystallographic research community 2005 (Lunteren, The Netherlands):

D. Georgieva: "Growing protein crystals for electron diffraction studies"

Lunteren protein research community 2005 (Lunteren, The Netherlands): D Georgieva: "Protein nanocrystallography" (AIO research prize)

Workshop on precession electron diffraction 2007 (Athens, Greece): D. Georgieva (invited) and J.P. Abrahams: "New frontiers for “spinning star”: pharmaceutical and protein 3D structures"

First precession electron diffraction user meeting 2007 (Martina Franca, Italy): D. Georgieva (invited): "Electron diffraction of protein and pharmaceutical nano-crystals"

Lorentz workshop 2007 (Leiden, The Netherlands): D. Georgieva (invited), L. Jiang and J.P. Abrahams: "New algorithms required for solving structures using electron diffraction of 3D beam-sensitive nano-crystals"