COHERENT X-RAY DIFFRACTIVE IMAGING AND NUCLEATION OF PROTEIN CRYSTALS

BY

SÉBASTIEN BOUTET

B.S., McGill University, 1999

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2005

Urbana, Illinois

Table of Contents

List of Figures . . . . . **v**

Chapter 1 Introduction **1**
1.1 Protein Crystallization 2
1.1.1 Proteins in Solution 3
1.1.2 Nucleation 4
1.1.3 Crystal Growth 6

Chapter 2 X-Ray Scattering **7**
2.1 Elastic Scattering 7
2.1.1 One Charge 8
2.1.2 Continuous Charge Distributions 10
2.1.3 One Molecule 11
2.1.4 One Supermolecule 12
2.1.5 A Collection of Supermolecules 13
2.1.6 Coherence 13
2.1.7 Orientational Average 15
2.1.8 Thermal Fluctuations and Disorder 16
2.2 Small Angle X-Ray Scattering 18
2.2.1 Solution Scattering 18
2.2.2 Low Q Approximation 19
2.2.3 High Q Approximation 20
2.3 Coherent X-Ray Diffraction 20
2.3.1 Crystal Diffraction 21
2.3.2 Miller Indices 21
2.3.3 Finite Size Crystal 22
2.3.4 Strain 23
2.3.5 2D Cuts Through Diffraction Pattern 24
2.3.6 Phase Retrieval 25

Chapter 3 Coherent X-Ray Diffractive Imaging from Protein Crystals . . **28**
3.1 Facility and Equipment 29
3.1.1 Source 29
3.1.2 Monochromator 30
3.1.3 Experimental Hutch 30
3.1.4 Detector 35
3.2 Sample and Preparation 36
3.2.1 Ferritin 36
3.2.2 High Quality Single Crystal 38
3.2.3 Protein Solutions 42
3.2.4 Crystal Shower 43
3.3 Experiment 45

3.3.1 Sample Mount . . . 46
3.3.2 Ewald Construction . . . 47
3.4 Results . . . 52
3.5 Thermal Fluctuations . . . 57
3.5.1 Freezing . . . 60
3.5.2 Gels . . . 62
3.6 Radiation Damage . . . 63
3.7 KB Mirror Focusing . . . 71
3.8 CXD Data . . . 72
3.9 Fitting Procedure . . . 75
3.10 Interpretation of the Fits . . . 78
3.11 Simulations . . . 79

Chapter 4 Nucleation at Low Temperature . . . 83
4.1 Facility and Equipment . . . 83
4.1.1 Experimental Setup . . . 83
4.2 Experiment . . . 84
4.2.1 Sample Preparation . . . 84
4.2.2 Data Collection . . . 84
4.3 Data and Data Treatment . . . 86
4.3.1 Calibration and Integration . . . 86
4.3.2 Normalization . . . 87
4.3.3 Form Factors and Concentration Dependence . . . 87
4.3.4 Temperature Dependence . . . 89
4.4 Structure Factors . . . 92
4.4.1 Holoferritin . . . 93
4.4.2 Protein Concentration Dependence . . . 94
4.4.3 Interpretation . . . 98
4.4.4 Ionic Strength Dependence . . . 100
4.4.5 Apoferritin . . . 103
4.5 Simulations . . . 107
4.5.1 Isotropic Clusters . . . 108
4.5.2 Planar clusters . . . 110
4.6 Fitting . . . 113
4.6.1 Holoferritin . . . 115
4.6.2 Apoferritin . . . 119
4.7 Interpretation of Results . . . 122

Chapter 5 Conclusion and Outlook . . . 126
5.1 Summary . . . 126
5.2 Future Directions . . . 127

Appendix A CXD Times Series . . . 129

References . . . 138

List of Figures

2.1 Scattering geometry of a single charge. 8
2.2 Transverse coherence length 14
2.3 Schematic representation of Gerchberg-Saxton method 27

3.1 View of vacuum chamber 31
3.2 Schematic of beamline setup 32
3.3 Bottom view of KB mirrors 33
3.4 Picture of the diffractometer 34
3.5 Atomic structure of ferritin molecule 37
3.6 Symmetry axes of ferritin molecule 39
3.7 Cadmium salt bridges 40
3.8 Reciprocal space of finite sized FCC crystal 42
3.9 Sample angles 44
3.10 Sample mount and temperature control 47
3.11 Picture of the sample mount 48
3.12 Ewald construction 49
3.13 Cuts through CXD pattern with different angles 51
3.14 SAXS pattern of solutions of apoferritin and holoferritin 52
3.15 Powder diffraction of apoferritin and holoferritin 53
3.16 Individual Bragg peaks from holoferritin 53
3.17 Integrated intensity for different $CdCl_2$ concentration 55
3.18 Number of crystals versus $CdCl_2$ concentration 56
3.19 Crystal size versus $CdCl_2$ concentration 56
3.20 Rocking curves of small holoferritin crystals 58
3.21 Crystal freely rotating in and out of the Bragg condition 58
3.22 Angular trajectories of small crystals at room temperature 59
3.23 Distribution of angular velocities 61
3.24 Scattered intensity from frozen small crystals 62
3.25 Radiation damage time series 64
3.26 Time series near a (111) Bragg peak 65
3.27 Time series near a (200) Bragg peak 65
3.28 Time variation of position and width of (111) Bragg peak 67
3.29 Time variation of position and width of (200) Bragg peak 68
3.30 Hendricks-Teller fit parameters versus time. 70
3.31 Time series of multiple Bragg peaks 72
3.32 Slices through a Bragg peak 73
3.33 Symmetrized slices through a Bragg peak 74
3.34 Fit of symmetrized slice 1 76
3.35 Fit of symmetrized slice 2 77
3.36 Multiple Bragg peaks with streaks 79
3.37 Octahedral FCC unit cell 80

3.38 CXD pattern of octahedral crystal . . . 81
3.39 Reciprocal space of strained FCC crystal . . . 82

4.1 Peltier current and voltage versus temerature . . . 85
4.2 Integrated powder diffraction of apoferritin and holoferritin . . . 86
4.3 Steps of the data treatment . . . 88
4.4 Intensity at different concentrations . . . 89
4.5 SAXS pattern from frozen holoferritin . . . 90
4.6 Integrated SAXS on holoferritin for multiple temperatures . . . 90
4.7 Thermal expansion of holoferritin small clusters . . . 91
4.8 Structure factor for apoferritin and holoferritin powder crystals . . . 93
4.9 Structure factors for holoferritin at 100mg/ml . . . 94
4.10 Structure factors for holoferritin at 200mg/ml . . . 95
4.11 Structure factors for holoferritin at 100mg/ml . . . 95
4.12 Radiation damage to small clusters of holoferritin . . . 96
4.13 Structure factors for holoferritin at 50mg/ml . . . 97
4.14 Structure factors for holoferritin at 35mg/ml . . . 98
4.15 Structure factors for holoferritin at 20mg/ml . . . 99
4.16 Structure factors for holoferritin at 10mg/ml . . . 100
4.17 Structure factors for holoferritin at 5mg/ml . . . 101
4.18 Effect of NaCl concentration on structure factors of holoferritin . . . 102
4.19 SAXS intensity for apoferritin at 50mg/ml and 150mg/ml . . . 103
4.20 Structure factors for apoferritin at 150mg/ml . . . 104
4.21 Structure factors for apoferritin at 50mg/ml . . . 105
4.22 Structure factors for apoferritin at 50mg/ml . . . 106
4.23 Thermal expansion of apoferritin small clusters . . . 106
4.24 Structure factors for apoferritin at 25mg/ml . . . 107
4.25 Structure factors for apoferritin at 10mg/ml . . . 108
4.26 FCC spherical clusters . . . 109
4.27 Simulations of structure factor from fcc clusters . . . 110
4.28 FCC unit cell and crystal contacts . . . 111
4.29 Simulations of structure factor from ⟨110⟩ rod clusters . . . 112
4.30 Results of fits for 100mg/ml holoferritin . . . 116
4.31 Results of fits for 200mg/ml holoferritin . . . 118
4.32 Results of fits for 100mg/ml holoferritin . . . 120
4.33 Results of fits for 50mg/ml apoferritin . . . 121
4.34 Results of fits for 50mg/ml holoferritin . . . 123

A.1 . . . 130
A.2 . . . 131
A.3 . . . 132
A.4 . . . 133
A.5 . . . 134
A.6 . . . 135
A.7 . . . 136
A.8 . . . 137

Chapter 1

Introduction

Crystallography, the study of the internal arrangements of crystalline matter is a fascinating branch of science and has been for a long time. Before the discovery of x-ray radiation allowing for diffraction measurements from crystals, the internal symmetries were deduced from the relative angles of each facets of a large crystal. X-ray diffraction allowed for direct measurement of the lattice structure of materials. The study of how atoms organize in periodic arrays in inorganic crystals is fascinating in its on regard, but this lead to greater advances in the understanding of materials such as the development of the theory of band structures and conductivity in crystals [3]. Soon after the discovery of x-rays, scientists began to determine the crystal structure of all the inorganic materials they could think of. Many decades later, structure determination using x-rays was even extended to the surface structure of crystals as well as thin films in what is called surface diffraction [67].

Somewhere along the way, it was realized that one might be able to determine the atomic structure of large macromolecules if those were arranged in a crystalline array. The fields of macromolecular and protein crystallography were slow to develop compared with inorganic crystallography due to the technical challenges presented. First of all, preparing crystals of proteins is rarely a straightforward proposition [45]. Secondly, high quality diffraction data is required to solve a protein structure and the quality of sources of x-rays used to limit the feasibility of this kind of experiment. And finally, the level of computation required to solve a protein structure can be enormous and only the advent of computers has made it possible for the technique to develop to what it is today.

New protein structures are solved everyday and the pace is only expected to increase. With the completion of the Human Genome Project, it is expected that a host of new proteins will be discovered and isolated and the focus now shifts to structure determination of all of these. The field of proteomics, in analogy to genomics should be a very busy field for many years to come.

There are however some major hurdles in the way of structure determination of all the proteins in the human proteome. Without even considering the isolation of the $\sim$400,000 proteins which will require some considerable effort, the production of high quality crystals of each of these proteins could prove to be an almost insurmountable challenge. In some cases, scientists spend years trying to crystallize a single protein. There exists some theories allowing to predict whether a certain protein can be crystallized such as the study of George *et al.* [22]. In this study, the authors concluded that the second virial coefficient obtained during static light scattering experiments on protein solutions could be used to determine

whether a certain solution could eventually produce crystals. While this is not the only attempt to develop a predictor of protein crystallization, no one has yet come up with satisfactory method of determining which conditions will lead to crystal formation for a specific protein. The trial-and-error method, with some way of narrowing the possibilities, is still the preferred method used to produce crystals of a new protein.

The determination of protein structures is in itself a valid scientific enterprise, but the possible applications of the knowledge of the structure is where the real advances are made. Once the structure is known, cures can be engineered in the case where the protein is associated with a disease. Also, the fairly new field of computational biology where interactions between large molecular complexes can be modeled using supercomputers would not exist without knowledge of the structure of the molecules as a starting point. Many theoretical studies such as light harvesting by bacteria have been performed [66] to name just one and one can expect this field of computational biophysics to grow with increasing computing power and more protein structure determination.

The bottleneck in structure determination is generally the growth of high quality crystals. Considerable effort has recently been put in developing experimental methods allowing for the growth of crystals of higher quality. Techniques such as the gel acupuncture technique and the growth of crystals in the microgravity of space have been developed [76, 20]. The success and usefulness of such methods have been somewhat unclear, with improved crystal quality in some, but not all cases. Furthermore, many scientists are now performing basics studies on the crystallization process itself to try to understand it better. In these experiments, the goal is not to have a large high quality crystal in the end, but rather to understand the steps involved in getting such a crystal. This is a fundamentally different approach to protein crystallization than that of protein crystallographers where the means to the end are irrelevant.

In this thesis, the techniques of coherent x-ray diffraction and small angle scattering were used to gain some insight on the structure of small crystals and nanoclusters of a model protein. After a quick introduction on the theory and previous experimental results on protein crystal nucleation and growth, Chapter 2 is dedicated to the formalism of x-ray scattering. This is followed in Chapter 3 by a discussion of the experimental setup and results of the coherent x-ray diffraction experiments performed. The small angle scattering results on small nano-sized clusters of proteins from solutions at low temperature is presented in Chapter 4. Finally, a summary of the results as well as possible future extensions are discussed in Chapter 5.

1.1 Protein Crystallization

When discussing protein crystallization, the proteins involved are generally globular proteins. This is because other types of proteins such as membrane proteins more often than not prove impossible to crystallize due to the low solubility of such proteins [42]. In the discussion below, we are concerned only with water-soluble globular proteins. The proteins must be kept in solution. The interactions in the solution can lead to crystallization which proceeds in two steps. The first one is the nucleation process of formation of stable clusters.

The second step is the actual growth of these clusters into large crystals.

1.1.1 Proteins in Solution

Proteins are chains made up of any combination of the 20 amino acids. Each amino acid consists of an amino group, an alpha carbon and a carboxyl group, which is common to all. They are distinguished only by a side chain attached to the C_α. The amino acids are linked together by the formation of a peptide bond between the carboxyl group and the amino group of the next amino acid, or residue in the chain. The peptide bond is planar which restricts somewhat the possible folding of the chain. The chain of peptide bonds forms what is called the backbone of the molecule.

With a given complete chain, each residue interacts in a certain way with the solvent as well as with the other residues in the molecule. This leads to the folding of the molecule into what is called, under physiological conditions, the native state of the protein. The folding and unfolding of proteins is in itself a very active field of research [75]. The side chains of the amino acids can be acidic, basic, neutral polar and neutral nonpolar. In general, the nonpolar residues which are hydrophobic tend to be buried in the interior of the protein with the hydrophilic side chains exposed on its surface.

In a given solvent at a certain pH, the protein will likely possess a net surface charge. This is due to the fact that certain amino acids will ionize at a given pH. Changing the pH allows one to control the total charge on the protein. The pI value corresponds to the pH value at which there is no charge on the protein. At this value, there is no Coulomb repulsion between the proteins which generally causes them to attract each other through primarily the Van der Waals interaction. Therefore, generally the solubility of proteins is at a minimum at the isoelectric point. This is the reason why no ordered crystallization can be expected at the isoelectric point. The attractive interactions are too strong leading to disordered aggregation or precipitation. However, since some attractive interactions are required for crystallization, it generally occurs at pH values not too far from pI.

The important factor of proteins in solution as far as crystal growth is concerned is the solubility. Many factors influence the solubility of proteins including temperature, pH and electrolyte concentration. Chernov *et al.* reviewed the dependence on each of these parameters [11]. The addition of salt often proves to be the simplest way to change the solubility of a protein. Depending on the protein and the conditions, the addition of salt may increase the solubility, what is called salting in, or decrease it, salting out. The addition of charged molecules, ions of polyelectrolytes to the solution of charged proteins modifies the electrostatic interactions between the proteins. Many models to account for the effect of simple ions have been developed such as the Debye-Hückel theory [46]. The main effect of the presence of mobile ions in solution is that they will tend to be found near opposite charges on the surface of the proteins. This has the effect of screening the surface charge and therefore the Coulomb interactions are greatly reduced.

At all times in a solution at any concentration, there are thermal fluctuations leading to Brownian motion. This creates a certain probability of collisions between proteins. When such a collision occurs, if the relative orientation of the proteins and the local environment

around them, such as the presence of ions, are in a certain state, there is a probability that the two proteins colliding will attach and form a dimer. The same can occur for multiple proteins and oligomers can form at any time in any solution. The probability of collisions increases with protein concentration. The stability of the clusters formed depends on the solution conditions and the level of supersaturation of the solution. With the equilibrium solubility of the protein for given conditions C_s, the supersaturation is defined as

$$\sigma = \ln\left(\frac{C}{C_s}\right) \tag{1.1}$$

where C is the protein concentration of the protein. This is a measure of how far from equilibrium the solution is. A supersaturated system will eventually return to equilibrium by removing the excess proteins from solution. The time scale on which this happens depends on the level of supersaturation. As proteins leave the aqueous phase, they form aggregates which can be crystalline or amorphous.

1.1.2 Nucleation

The thermal fluctuations in the system leading to the formation of stable clusters is called the nucleation step of crystallization. For a given supersaturated solution, there is a critical cluster size which is unstable [71]. Smaller clusters will tend to redissolve and larger one will be more likely to grow than to redissolve. This step is critical in obtaining crystals. Often, the level of supersaturation required to yield stable nuclei in a reasonable amount of time is different from the level of supersaturation needed for slow orderly growth. It sometimes proved necessary in order to grow large crystals to seed the solution with preexisting small crystals obtained from a different solution. Removing the seeds from a solution in which they would grow too fast and placing them is a less supersaturated solution slows the growth and improves the crystal quality.

The question of what is the critical size for given conditions is generally unanswered. A crystal is at equilibrium with the solution when the chemical potentials of the two phases are equal. The chemical potential of a phase is the work required to add 1 particle to this phase and is denoted by μ_x for the crystalline phase and μ_s for the solution. It is given by [10, 65]

$$\mu = u - Ts + Pv \tag{1.2}$$

where u, s and v are the potential energy, the entropy and the specific volume per particle respectively and T is the temperature and P is the pressure on the system. In the case of a large crystal when the surface contribution is small and can be neglected, $\mu_x = \mu_s$ is the equilibrium condition. However, for small crystals, the surface energy becomes significant and must accounted for.

We consider the case of homogeneous nucleation, which is the relevant situation for protein crystal growth in a bulk solution. Let's assume the cluster is spherical with no direction dependence to the specific surface free energy α. The surface energy is equal to $4\pi R^2\alpha$ where R is the radius of the sphere. Suppose a number δN of molecules leave the

solution and join the crystal phase. This leads to a change in the thermodynamic potential of $(\mu_s - \mu_x)\delta N$, but this increases the size of the sphere by an amount $\delta R = \frac{v\delta N}{4\pi R^2}$ where v is the volume of one molecule. The resulting increase in surface energy is $8\pi R\alpha\delta R$. If this increase in surface energy is smaller than $(\mu_s - \mu_x)\delta N$, then the crystal can grow. The equilibrium size is therefore reached when

$$\mu_s - \mu_x = \frac{2\pi\alpha}{R} \tag{1.3}$$

We are however more interested in the critical size than in the equilibrium size. Using similar arguments as above, we can calculate the change in thermodynamic potential ($\delta\Phi$) from the formation of a spherical cluster of radius R. Denoting $\Delta\mu = \mu_s - \mu_x$, we get

$$\delta\Phi = -\frac{4\pi R^3\Delta\mu}{3v} + 4\pi R^2\alpha \tag{1.4}$$

For small clusters, the surface term dominates and there is an increase with R for small values of R. For large sizes, the volume term dominates and there is a decrease in the potential with R. This function is therefore peaked and there is an unstable equilibrium at the R value corresponding to this peak. Smaller clusters tend to redissolve while larger ones tend to grow. This is the critical size and the condition for it is that the derivative of $\delta\Phi$ is zero, yielding a critical radius of

$$R_c = \frac{2v\alpha}{\Delta\mu} \tag{1.5}$$

The same calculation can be repeated for non-spherical clusters of known shape. The calculation is simple in the case where all the surface are equivalent as in the case of a cube for example or an octahedron with {111} faces. However, for general crystalline nuclei, there is a probability for each crystallographic face to be present, with the close-packed faces more likely to occur. To calculate the critical size, one must know the surface free energy of each surface and a shape of the cluster must be assumed. The logical assumption is the equilibrium crystal shape which can be calculated using the methodology of Wulff [90]. These calculations from classical nucleation theory breaks down when the critical size becomes so small that the bulk chemical potential of the crystal becomes ill-defined.

The nucleation rate can be calculated from fluctuation theory if we assume the system is close to equilibrium. The details of the calculation for a specific protein can be quite complicated. Many scientists are studying nucleation using different methods. It is generally very difficult to observe critical nuclei directly as they form. Yau *et al.* claim to have directly observed near-critical nuclei using AFM [92]. The difficulty in observing critical nuclei lies in the fact that they are first very small and second they are short-lived. Most studies of nucleation dealt with measurements of nucleation rates. A thorough review of recent results in the field are presented by Vekilov [85].

One important question about protein crystal nucleation is whether there is an amorphous step in crystallization. Is there a phase separation leading to a dense liquid state which eventually forms a crystalline cluster? A study on the formation of microcrystals of the brome mosaic virus indicates that there is no amorphous dense liquid state [9]. The

addition of polyethylene glycol (PEG) lead to the formation of small crystals which could be observed in real-time using synchrotron radiation. The small angle scattering pattern showed no signs of liquid-liquid phase separation.

1.1.3 Crystal Growth

The growth of crystals is much easier to measure experimentally than the nucleation and this leads to a much better understanding of the process. Many experimental techniques can be used to study the evolution of a crystal surface including atomic force microscopy and surface x-ray diffraction. Once a large facet of a crystal exists, it is possible to measure its evolution in real time. For critical nuclei, the difficulty in observing them lies in finding them in the first place.

The growth of crystals proceeds by attachment of single molecules or clusters of molecules to a preexisting surface. In hen egg-white lysozyme crystallization, there is clear evidence that only monomers of the protein attach to the preexisting surfaces [18]. In apoferritin however, there is direct evidence from AFM measurements that oligomers form in solution and get included in the crystal structure, leading to the formation of defects in the crystal [91].

The details of the growth process of large crystals are unimportant for the purpose of this thesis since no study of the growth process was performed. The growth generally proceeds layer by layer with step propagation along the surface. Single molecules are generally incorporated at the kink sites where the most new bonds are created [10].

The molecules or atoms at the surface of crystals are in a different environment than the bulk with dangling bonds. This often leads to surface relaxation which is often observed in inorganic crystals but was only recently observed in protein crystals using surface x-ray diffraction [6]. Inorganic crystal surfaces often undergo a reconstruction. This was also observed in lysozyme crystals using AFM [41].

Chapter 2

X-Ray Scattering

X-rays are a form of electromagnetic radiation discovered by Röntgen in 1895. As all forms of electromagnetic radiation, they can be thought of as either waves or massless particles. A beam of x-rays can be represented as a collection of particles called photons each possessing a certain energy ε. Alternatively, it can also be represented as a collection of waves with wavelength λ. These two descriptions are equivalent but one may prove more convenient than the other to explain a certain phenomena observed. For example, the absorption of x-rays is better explained using the particle nature of light while scattering is more easily represented using the wave description. The wavelength of the x-rays is related to the energy by $\varepsilon = \frac{hc}{\lambda}$, where h is Planck's constant and c is the speed of light.

What distinguishes x-ray photons from other types of radiation in the electromagnetic spectrum is simply their wavelength. Radiation with a wavelength in the range 0.1-100Å is considered to be x-ray radiation. A further distinction is made between hard x-rays ($\lambda \sim 2$Å or less) and soft x-rays (> 2Å). Hard x-rays have a wavelength closely matched to the size of atoms and atomic spacings of materials which makes them extremely useful in structural studies.

In this chapter, we will treat x-rays as waves. The derivation of the scattered intensity from a distribution of charges follows the treatments of Als-Nielsen [2] and Warren [86].

2.1 Elastic Scattering

Electromagnetic waves are transverse waves propagating in both space and time. They consist of two oscillating fields, an electric field ($\mathbf{E}$) and a magnetic field ($\mathbf{B}$) which are orthogonal to each other as well as orthogonal to the direction of propagation of the wave as required by Maxwell's equations of electrodynamics [35]. The two fields are given by

$$\mathbf{E_i} = E_0 e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}\hat{\epsilon} \tag{2.1}$$

$$\mathbf{B_i} = B_0 e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}(\hat{\epsilon}\times\hat{k}) \tag{2.2}$$

where $\mathbf{k}$ is known as the wavevector with magnitude given by $k = \frac{2\pi}{\lambda}$ and whose direction is the direction of propagation of the wave. The vector $\mathbf{r}$ is a spatial position, $\omega = \frac{2\pi c}{\lambda}$ is the cyclic frequency of the wave, t is time and $\hat{\epsilon}$ is the direction of the polarization, taken by definition as the direction of the electric field. The orthogonality of the $\mathbf{E}$ field to the direction of propagation of the wave implies that $\mathbf{k}\cdot\hat{\epsilon} = 0$. We shall discuss below

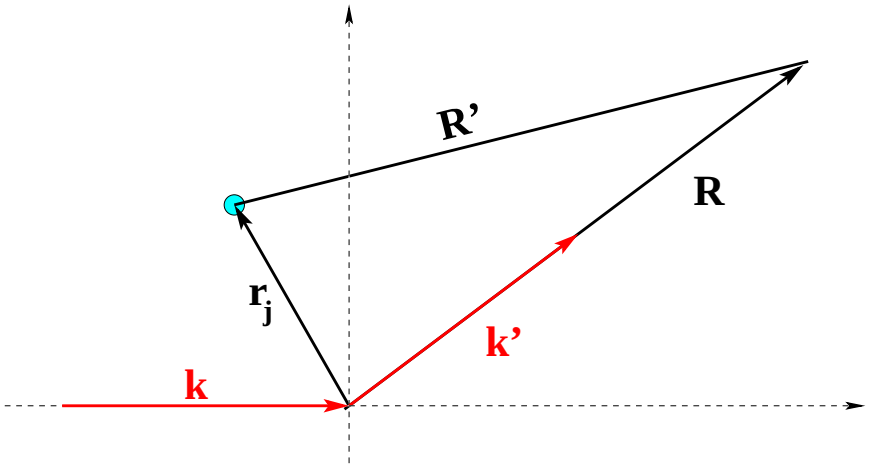


Figure 2.1: Scattering geometry of a single charged particle located at position r_j. The scattering is observed at a vector R from the origin with $R \gg r_j$.

only scattering events in which only the direction of $\mathbf{k}$ changes and its magnitude remains constant. That is we will consider scattering to be elastic. We will also assume there are no multiple scattering events, an approximation known as the kinematical approximation.

2.1.1 One Charge

When an electromagnetic wave impinges on a material, the electric field interacts with electric charges in the material and the magnetic field interacts with the spins present in the material. The interactions between the electric charges and the electric field are much stronger than the magnetic interactions, by a factor of $\frac{\hbar\omega}{mc}$. Since one can only measure the intensity of an electromagnetic wave, *i.e* the square of its amplitude, we get that for x-rays the magnetic scattering is roughly 10^4 times weaker than charge scattering. We will therefore neglect the magnetic interactions from this point on.

The interaction between the oscillating electric field of the x-ray and a charged particle located at position $\mathbf{r}_j$ has the effect of accelerating back and forth the charged particle in the direction of the electric field, *i.e* transverse to the direction of propagation of the x-rays. The scattering geometry is shown on Figure 2.1. From classical electrodynamics, we know that the force applied on a particle of charge e by an electric field is $\mathbf{F} = e\mathbf{E}$. Using Newton's law $\mathbf{F} = m\mathbf{a}$, we get the acceleration of the charged particle.

$$\mathbf{a}(t) = \frac{e}{m}\mathbf{E_i} = \frac{e}{m}E_0 e^{i(\mathbf{k}\cdot\mathbf{r_j}-\omega t)}\hat{\epsilon} \tag{2.3}$$

The charged particle is then set into an oscillatory motion in the $\hat{\epsilon}$ direction, which is equivalent to saying that it behaves as an oscillating electric dipole. The accelerated charge emits radiation with the electric field given by

$$\mathbf{E}' = -\frac{e\mathbf{a}(t')\sin\theta}{c^2 R'} = -\frac{e^2\sin\theta}{mc^2 R'}E_0 e^{i(\mathbf{k}\cdot\mathbf{r_j}-\omega t')}\hat{\epsilon} \tag{2.4}$$

where $\mathbf{R}'$ is the vector between the position of observation of the electric field and the charged particle and θ is the angle between $\hat{\epsilon}$ and $\mathbf{R}'$. The minus sign indicates a phase shift of π in the scattered radiation compared with the incident radiation. The $\sin\theta$ term implies that no scattering is observed in the direction of polarization. When evaluating the electric field measured at time t at a distance R' from the source, it is necessary to account for the fact that the radiation emitted travels at the speed of light. One must then use the acceleration at the time of emission $t' = t - R'/c$ when calculating the radiated electric field. Putting the value of t' into equation (2.4) and using the relation $\frac{\omega}{c} = \frac{2\pi}{\lambda} = k$, we get

$$\mathbf{E}' = -\frac{e^2 \sin\theta}{mc^2 R'} E_0 e^{i(\mathbf{k}\cdot\mathbf{r_j} + kR' - \omega t)} \hat{\epsilon} = -\frac{e^2 \sin\theta}{mc^2 R'} E_0 e^{i(\mathbf{k}\cdot\mathbf{r_j} + \mathbf{k}'\cdot\mathbf{R}' - \omega t)} \hat{\epsilon} \tag{2.5}$$

where we defined the wavevector of the scattered radiation $\mathbf{k}'$. We have explicitly assumed here that the magnitudes of $\mathbf{k}$ and $\mathbf{k}'$ are equal. We would like to write this equation in terms of the vector $\mathbf{R}$ from the origin of our coordinate system rather than in terms of $\mathbf{R}'$. We can write $\mathbf{R}' = \mathbf{R} - \mathbf{r_j}$ and replace in equation (2.5).

$$\mathbf{E}' = -\frac{e^2 \sin\theta}{mc^2 R} E_0 e^{i(\mathbf{k}\cdot\mathbf{r_j} + \mathbf{k}'\cdot\mathbf{R} - \mathbf{k}'\cdot\mathbf{r_j} - \omega t)} \hat{\epsilon} \tag{2.6}$$

where we have made the assumption the $r_j \ll R$ so that the angle θ measured from the origin or from $\mathbf{r_j}$ is approximately the same and $R \approx R'$ in the denominator. We can rewrite equation (2.6) into a simpler and more familiar form by introducing the vector $\mathbf{q} = \mathbf{k}' - \mathbf{k}$ known as the momentum transfer.

$$\begin{aligned} \mathbf{E}' &= -\frac{e^2 \sin\theta}{mc^2 R} E_0 e^{i(\mathbf{k}'\cdot\mathbf{R} - \omega t)} e^{i(\mathbf{k}\cdot\mathbf{r_j} - \mathbf{k}'\cdot\mathbf{r_j})} \hat{\epsilon} \\ &= -\frac{e^2 \sin\theta}{mc^2 R} E_0 e^{i(\mathbf{k}'\cdot\mathbf{R} - \omega t)} e^{-i\mathbf{q}\cdot\mathbf{r_j}} \hat{\epsilon} \end{aligned} \tag{2.7}$$

It is straightforward to extend equation (2.7) to the case of a distribution of N identical charges located at position $\mathbf{r}_j$.

$$\mathbf{E}' = \sum_{j=1}^{N} -\frac{e^2 \sin\theta}{mc^2 R} E_0 e^{i(\mathbf{k}'\cdot\mathbf{R} - \omega t)} e^{-i\mathbf{q}\cdot\mathbf{r_j}} \hat{\epsilon} \tag{2.8}$$

The prefactor before $e^{-i\mathbf{q}\cdot\mathbf{r_j}}$ is the same for all charges and $e^{-i\mathbf{q}\cdot\mathbf{r_j}}$ determines the scattering profile.

$$\mathbf{E}' = -\frac{e^2 \sin\theta}{mc^2 R} E_0 e^{i(\mathbf{k}'\cdot\mathbf{R} - \omega t)} \sum_{j=1}^{N} e^{-i\mathbf{q}\cdot\mathbf{r_j}} \hat{\epsilon} \tag{2.9}$$

Detectors of x-rays are incapable of measuring the electric field itself. The quantity measured is the intensity of the wave given by the square of the amplitude of the wave, or the square of the electric field, $I = E^2$. The exponential term before the summation in equation (2.9) is then only a constant phase factor which disappears when we measure the intensity.

$$I = E^2 = \Big(\frac{e^2}{mc^2}\Big)^2 \frac{\sin^2\theta E_0^2}{R^2} \Big(\sum_{j=1}^{N} e^{-i\mathbf{q}\cdot\mathbf{r_j}}\Big)^2 \tag{2.10}$$

The scattered intensity depends only on the distribution of the charges.

So far we have discussed charges in a general way. However, in real materials, only two types of charges will be present, the protons and the electrons. From equation (2.10), we can see that the intensity is proportional to the inverse square of the mass of the charged particle accelerated by the x-ray beam. This implies that the intensity scattered from a proton will be 6 orders of magnitude weaker than that from an electron since the proton is roughly 2000 times more massive. X-rays are then scattered primarily by electrons and we need only be concerned with the electron density of the material. Using e and m as the charge and the mass of the electron respectively, we get $\frac{e}{mc^2} = r_0$, where r_0=2.82×10^{-5}Å is known as the classical radius of the electron.

The polarization factor $\sin^2\theta$ depends on the polarization characteristics of the incident beam. The work presented here was performed at a synchrotron source which produces x-ray beam with horizontal polarization as will be discussed in Chapter 3. This implies that the polarization factor is 1 if the scattering plane is vertical and $\sin^2\theta$ if the scattering plane is horizontal. As will be discussed further in Chapter 3, the samples studied here produced scattering only at small angles <3° leading to $\sin^2\theta = 0.997$, very close to 1. We shall therefore ignore the polarization factor from this point forward. The scattered amplitude from a distribution of point charges is then

$$E(\mathbf{q}) = -\frac{r_0}{R}\sum_{j=1}^{N} e^{-i\mathbf{q}\cdot\mathbf{r_j}} \tag{2.11}$$

2.1.2 Continuous Charge Distributions

The formalism derived for a distribution of point charges can be extended to continuous distributions of charges which more accurately represent the real physical world. Electrons are quantum mechanical particles better described by a wavefunction than a classical particle. In materials, the electrons are generally bound to atoms and are described by orbitals or a charge density. Given any electron density $\rho(\mathbf{r})$ such that $\int \rho(\mathbf{r})\mathrm{d}\mathbf{r} = N_e$, where N_e is the total number of electrons illuminated by the x-ray beam, then one can write the scattered amplitude as

$$E(\mathbf{q}) = -\frac{r_0}{R}\int \rho(\mathbf{r})e^{-i\mathbf{q}\cdot\mathbf{r}}\mathrm{d}\mathbf{r} \tag{2.12}$$

Equation (2.12) is recognized as a Fourier transform. Thus, within the approximations made above, the scattered amplitude is the Fourier transform of the electron density of the illuminated sample.

The scattered amplitude for a single atom with atomic number Z can be calculated if the quantum mechanical electron distribution $\rho_Z(\mathbf{r})$ is known. We define here a scalar known as the atomic form factor $f_Z(\mathbf{q})$ given by

$$f_Z(\mathbf{q}) = \int \rho_Z(\mathbf{r})e^{-i\mathbf{q}\cdot\mathbf{r}}\mathrm{d}\mathbf{r} \tag{2.13}$$

such that $E_Z(\mathbf{q}) = -\frac{r_0}{R}f_Z(\mathbf{q})$ is the scattered electric field from this single atom. The form

factor is related to the scattered electric field by a multiplicative constant. In this thesis, we are not dealing with absolute values of the scattered intensity and thus we shall drop the prefactor $-\frac{r_0}{R}$ and deal only with form factors. From this point on, when discussing scattered intensity, it will be implied that the prefactor has been dropped so that

$$I = \Big[\int \rho_Z(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{r}} \mathrm{d}\mathbf{r} \Big]^2 \tag{2.14}$$

The atomic form factors were calculated and tabulated in the International Tables for Crystallography [63]. An analytical approximation is given by

$$f_Z(\mathbf{q}) = \sum_{j=1}^{4} a_j \exp(-\frac{b_j \mathbf{q}^2}{16\pi^2}) + c \tag{2.15}$$

where the a_j's, b_j's and c are fitted parameters.

2.1.3 One Molecule

Given these parameters, one can readily calculate the molecular form factor $F(\mathbf{q})$ of a given molecule of known structure. This can be done using the convolution theorem relating the products of Fourier transforms. Denoting the Fourier transform of two functions $g(\mathbf{r})$ and $h(\mathbf{r})$ as $\mathcal{F}(g(\mathbf{r})) = G(\mathbf{q})$ and $\mathcal{F}(h(\mathbf{r})) = H(\mathbf{q})$, the convolution theorem states that

$$\mathcal{F}(g(\mathbf{r})) \times \mathcal{F}(h(\mathbf{r})) = G(\mathbf{q}) \times H(\mathbf{q}) = \mathcal{F}(g(\mathbf{r}) \otimes h(\mathbf{r})) \tag{2.16}$$

where the $\otimes$ operator represents the convolution of two functions. Let us define here the inverse Fourier transform

$$\mathcal{F}^{-1}(G(\mathbf{q})) = \int G(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{r}} \mathrm{d}\mathbf{q} = \mathcal{F}^{-1}[\mathcal{F}(g(\mathbf{r}))] = g(\mathbf{r}) \tag{2.17}$$

It follows from equations (2.16) and (2.17) that

$$\mathcal{F}^{-1}[G(\mathbf{q}) \times H(\mathbf{q})] = \mathcal{F}^{-1}[\mathcal{F}(g(\mathbf{r}) \otimes h(\mathbf{r}))] = g(\mathbf{r}) \otimes h(\mathbf{r}) \tag{2.18}$$

Therefore, since Fourier transforms and inverse Fourier transforms have the same properties, we conclude that the Fourier transform of a convolution of two functions is the product of the Fourier transforms of the two functions and the Fourier transform of a product is the convolution of the Fourier transforms. These two facts often prove very useful when trying to calculate the scattered intensity.

Going back to the molecular form factor $F(\mathbf{q})$ defined above, suppose that the atomic structure of a certain molecule containing N atoms is known such that the electron density is given by

$$\rho(\mathbf{r}) = \sum_{j=1}^{N} \rho_j(\mathbf{r}) \otimes \rho_{\mathrm{mol}_j}(\mathbf{r}) \tag{2.19}$$

where $\rho_j(\mathbf{r})$ is the electron density of atom type j and $\rho_{\mathrm{Mol}_j}(\mathbf{r})$ is the distribution of atoms of

type j in the molecule. In this case, $\rho_{\mathrm{mol}_j}(\mathbf{r})$ is a sum of delta function $\delta(\mathbf{r}-\mathbf{r_j})$ representing the location of each atom $\mathbf{r_j}$ in the molecule. The electron density of the molecule is written as a convolution and one can then write the molecular form factor as the product of two functions

$$F_{\mathrm{mol}}(\mathbf{q}) = \sum_{j=1}^{N} \mathcal{F}[\rho_j(\mathbf{r})] \times \mathcal{F}[\rho_{\mathrm{mol}_j}(\mathbf{r})] = \sum_{j=1}^{N} f_j(\mathbf{q}) e^{-i\mathbf{q}\cdot\mathbf{r_j}} \tag{2.20}$$

where $f_j(\mathbf{r})$ is the atomic form factor defined in equations (2.13) and (2.15) and we used the fact that the Fourier transform of a delta function is $\mathcal{F}[\delta(\mathbf{r}-\mathbf{r_j})] = e^{-i\mathbf{q}\cdot\mathbf{r_j}}$. The convolution theorem allowed us to write the molecular form factor or scattered amplitude as a sum of waves coming from different points as in equation (2.11) but this time with the amplitude of each wave modulated by the corresponding atomic form factor.

In the case of large biological molecules, it can prove simpler and sufficient to approximate the electron density of the molecule by a well-defined shape of constant density. For example, the molecule can be known to have a spherical shape or in the case of many viruses and some proteins, it may have the shape of a spherical shell. The large size of these molecules makes it acceptable to neglect the atomic nature of matter and treat it as a continuum, provided scattering is measured only at values of $q \ll 2\pi/d$, where $d \approx 3$Å is a typical atomic spacing in biological material. When q approaches $2\pi/d$ the atomic nature of the sample becomes significant in the scattering. The electron density is then written as a $\rho_{\mathrm{mol}}(\mathbf{r}) = \rho_{\mathrm{avg}} s(\mathbf{r})$ where ρ_{avg} is the average electron density of the molecule and $s(\mathbf{r})$ is the shape function of the molecule given by

$$s(\mathbf{r}) = \begin{cases} 1 & \text{Inside the molecule} \\ 0 & \text{Outside the molecule} \end{cases} \tag{2.21}$$

The exact molecular form factors calculated this way for a sphere of radius R and a spherical shell of inner radius R_1 and outer radius R_2 are given respectively by

$$F(\mathbf{q}) = 3\frac{\sin(qR) - qR\cos(qR)}{(qR)^3} \tag{2.22}$$

$$F(\mathbf{q}) = 3\frac{[\sin(qR_2) - qR\cos(qR_2)] - [\sin(qR_1) - qR\cos(qR_1)]}{(q(R_2-R_1))^3} \tag{2.23}$$

2.1.4 One Supermolecule

In Chapter 4, we will discuss the scattering from small clusters of identical proteins related to critical nucleation. We will refer to these as supermolecules since they are molecules whose building blocks are not atoms, but molecules themselves. Given a cluster of a small number of proteins, the electron density is given by the convolution of 1 protein and a sum of delta functions representing the position in real space of each protein. This is analogous the case of a single molecule discussed in Section 2.1.3. In that case, one molecule was made up of atoms while now one supermolecule is made up of molecules. In the case of large arrangements such as these, it is common to define a new quantity called the structure factor denoted by $S(\mathbf{q})$. The scattered amplitude is then given by

$$
\begin{aligned}
E(\mathbf{q}) =& F(\mathbf{q})S(\mathbf{q}) \\
=& \mathcal{F}[\rho_{\text{mol}}(\mathbf{r})]\mathcal{F}[\rho_{\text{cluster}}(\mathbf{r})] \\
=& F(\mathbf{q})\sum_{j=1}^{N} e^{-i\mathbf{q}\cdot\mathbf{r_j}}
\end{aligned}
\tag{2.24}
$$

where we again made use of the convolution theorem. The molecular positions in the cluster $\rho_{\text{cluster}}(\mathbf{r})$ are again a sum of delta functions and the structure factor depends solely on this distribution. We can factor the scattered amplitude into a form factor which depends solely on the electron density of one molecule or one building block and a structure factor which depends only on the arrangement of the molecules. If the molecular form factor of the molecule is known either by measurement or calculation using equation (2.20), the scattered amplitude of any cluster or supermolecule can be calculated readily from a known distribution of the molecules.

2.1.5 A Collection of Supermolecules

Given an ordered cluster of proteins of size < 500Å containing on the order of 100 molecules, the scattered intensity from such a cluster is very weak and with present technology difficult to measure. The standard way of obtaining information about such a structure is to produce many of them and measure the scattering from the ensemble illuminated. Suppose we have an ensemble of many identical supermolecules all possessing the same orientation. In order to calculate the scattered amplitude, we must first discuss the coherence characteristics of the x-ray beam.

2.1.6 Coherence

Up to this point, the x-ray beam impinging on the sample has been assumed to be perfectly coherent, that is the phase of the wave is uniquely determined over all space. However, this is an unrealistic approximation. First of all the beam has a finite size and this obviously sets an upper bound on the coherence length. We define a coherence length as the distance between two points of the beam whose phase is related by a shift of π. That is the two points are out of phase with each other. The are two distinct coherence lengths we need to worry about : the transverse(L_T) and the longitudinal(L_L) coherence lengths. The longitudinal coherence length is determined by the monochromaticity of the beam. Suppose the beam is made up of two waves of different wavelengths λ_1 and λ_2. If both waves are in phase at the origin, *i.e* both phases are $\phi = 0$, at a distance d from the origin, their phases are respectively $\phi_1 = 2\pi d/\lambda_1$ and $\phi_2 = 2\pi d/\lambda_2$. The two waves are out of phase when

$$
\phi_2 - \phi_1 = \pi = 2\pi d\frac{\lambda_1 - \lambda_2}{\lambda_1\lambda_2} \tag{2.25}
$$

From equation (2.25) we solve for $d = L_L$ and get

$$
L_L = \frac{\lambda_1\lambda_2}{2(\lambda_1 - \lambda_2)} \approx \frac{\lambda^2}{2(\Delta\lambda)} \tag{2.26}
$$

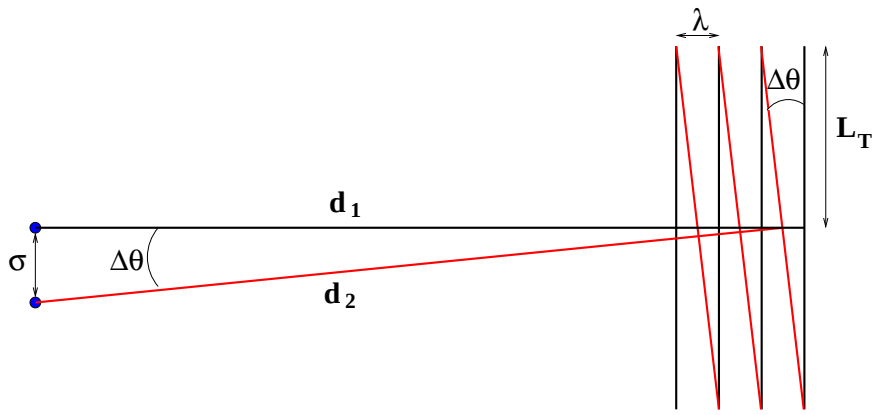


Figure 2.2: Diagram of the transverse coherence length. Two waves generated by a finite size source are in phase a distance L_T away from the axis of propagation of the beam but are out of phase on-axis

where we have defined $\Delta\lambda = \lambda_1 - \lambda_2$ and assumed $\Delta\lambda \ll \lambda$. For a distribution of wavelengths with width $\Delta\lambda$, the longitudinal coherence length is therefore related to the spread of wavelengths in the beam.

The transverse coherence length is determined by the size of the source (σ) and the distance of the source to the sample(D), assuming perfect optics between the source and the sample. Given two point sources a distance σ apart, each source emits a spherical wave at a wavelength λ as shown on Figure 2.2. The two waves are in phase at a distance L_T away from the axis of the beam meaning both waves traveled the exact same path length. The source size is assumed to be small compared with the source-sample distance so that $d_1 \approx d_2 = D$ and both waves can be approximated as plane waves. On the beam axis, the two waves are exactly out of phase so that one wave traveled an extra distance equal to $\lambda/2$. From the similarity of the triangles, we get $\sigma/R = \lambda/2L_T$. Hence

$$L_T = \frac{\lambda R}{2\sigma} \tag{2.27}$$

In is important to realize that two components of a sample separated by a distance larger than the coherence length cannot interfere coherently since the relative phases of the impinging beam at both locations is random. The interference varies over time and averages to zero. The intensity from each coherence volume adds incoherently. This means only structures smaller than the coherence volume can be resolved. It is also worth noting that even though the field of Coherent Scattering is a well-defined field on its own, all scattering is at some level coherent. For example the atomic form factors can only be measured because the coherence volume of the beam is larger than the atom. What is meant by Coherent Scattering is that the incoherent averaging is reduced or completely eliminated by closely matching the size of the sample to the coherence volume of the beam.

Given two atoms or two molecules, they will only scatter coherently if they are within one coherence volume of each other. If that is the case, then they scatter as one single object and the relative phases of the scattered beams must be carefully accounted for and

the scattered amplitudes must be added. If however the two molecules are far apart, then they scatter incoherently and not the amplitudes but the intensities must now be summed. The scattered amplitude of a molecule at $q = 0$ equals the number of electrons Z in the molecule, so that the intensity is $I(q = 0) = Z^2$. If N molecules are present, we get

$$I(q=0) = \begin{cases} N^2Z^2 & \text{Coherent Scattering} \\ NZ^2 & \text{Incoherent Scattering} \end{cases} \tag{2.28}$$

The coherence lengths at the synchrotron sources used for the experiments described in this thesis are on the order of a few microns so that small clusters of molecules are always coherently illuminated. However, these clusters are sufficiently sparsely distributed in a sample that we can assume that they scatter incoherently. So we do not need to calculate the Fourier transform of the distribution of supermolecules, but rather calculate the intensity of each possible supermolecule and sum these intensities.

Due to the random processes involved in the nucleation process of crystallization leading to the formation of small clusters of proteins, we do not expect all of them to have the same structure or the same size. For each structure containing N molecules, there is a probability $p(N)$ that it will occur such that $\sum p(N) = 1$. Therefore the scattered amplitude for each supermolecule is given by equation (2.24) as before and the intensity of a distribution of supermolecules is given by

$$I(\mathbf{q}) = \sum_N p(N) I_N(\mathbf{q}) = \sum_N p(N)[F(\mathbf{q})S(\mathbf{q})]^2 \tag{2.29}$$

2.1.7 Orientational Average

Not only do we expect the critical clusters to have a size distribution, there is also no reason to expect them to have the same orientation in space. The measured scattered intensity from a distribution of clusters will therefore be the orientational average of the intensity. We go back to equation (2.24) describing the scattered amplitude from one supermolecule. The orientational average of the square of the structure factor is

$$S^2(q) = \left\langle S^2(\mathbf{q}) \right\rangle = \left\langle \left| \sum_{j=1}^{N} e^{-i\mathbf{q}\cdot\mathbf{r_j}} \right|^2 \right\rangle = \sum_{j=1}^{N}\sum_{k=1}^{N} \frac{\sin(qr_{jk})}{qr_{jk}} \tag{2.30}$$

where the $\langle\rangle$ denotes orientational averaging and the r_{jk} are the distance between molecules j and k. It is simple to extend this to the situation where not all molecules in the cluster are the same so that the intensity is no longer factorable into a form and a structure factor. We then obtain a formula first derived by Debye in 1915 [12]

$$I(q) = \left\langle \left| \sum_{j=1}^{N} F_j(\mathbf{q}) e^{-i\mathbf{q}\cdot\mathbf{r_j}} \right|^2 \right\rangle = \sum_{j=1}^{N}\sum_{k=1}^{N} F_j(q)F_k(q) \frac{\sin(qr_{jk})}{qr_{jk}} \tag{2.31}$$

In equation (2.31), it is assumed that the molecular form factor is isotropic, $F_j(\mathbf{q}) = F_j(q)$. In that situation, with all $F_j = F$ we get

$$I(q) = \left\langle F^2(\mathbf{q}) S^2(\mathbf{q}) \right\rangle = \left\langle F^2(\mathbf{q}) \right\rangle \left\langle S^2(\mathbf{q}) \right\rangle \tag{2.32}$$

However, the molecular form factor is isotropic only if the molecule itself is isotropic and it is generally not the case for proteins. Even when they are closely spherical, they are still not perfectly isotropic. In this case, the orientational average of the intensity is no longer factorable and equation (2.32) no longer holds. The orientational average of the intensity can no longer be computed using the simple relation of equation (2.31) if the molecule is too strongly anisotropic but rather must be computed 3-dimensionally and then spherically averaged.

2.1.8 Thermal Fluctuations and Disorder

The spatial arrangements of real materials is not fixed as in the idealized discussion above. Thermal fluctuations will set single atoms and molecules as a whole into oscillatory motions. It is necessary to calculate how these motions will affect the scattering of x-rays from the sample. We consider here individual atoms and later extend the result to the motions of entire molecules. We start by writing the time dependent position vector of atom j as $\mathbf{r}_j(t) = \mathbf{R}_j(t) + \mathbf{u}_j(t)$ where the time averaged position is $\langle \mathbf{r}_j(t) \rangle_t = \mathbf{R}_j$, which implies that $\langle \mathbf{u}_j(t) \rangle_t = 0$. Here the $\langle \rangle_t$ indicates time averaging. Therefore we identify $\mathbf{R}_j$ as the average position used in all previous calculations and $\mathbf{u}_j$ as the fluctuation. The scattered intensity from any object containing N atoms is given by

$$\begin{aligned} I(\mathbf{q}) &= \left\langle \Big| \sum_{j=1}^{N} f_j(\mathbf{q}) e^{-i\mathbf{q}\cdot\mathbf{r}_j(t)} \Big|^2 \right\rangle_t \\ &= \left\langle \sum_j f_j(\mathbf{q}) e^{-i\mathbf{q}\cdot(\mathbf{R}_j + \mathbf{u}_j(t))} \sum_k f_k^*(\mathbf{q}) e^{i\mathbf{q}\cdot(\mathbf{R}_k + \mathbf{u}_k(t))} \right\rangle_t \\ &= \sum_j \sum_k f_j(\mathbf{q}) f_k^*(\mathbf{q}) e^{i\mathbf{q}\cdot(\mathbf{R}_k - \mathbf{R}_j)} \left\langle e^{i\mathbf{q}\cdot(\mathbf{u}_k(t) - \mathbf{u}_j(t))} \right\rangle_t \end{aligned} \tag{2.33}$$

Here we make the assumption that the atomic displacements $\mathbf{u}$ are harmonic, and described by gaussian statistics. In that case, it can be shown that

$$\begin{aligned} \left\langle e^{i\mathbf{q}\cdot(\mathbf{u}_k(t) - \mathbf{u}_j(t))} \right\rangle_t &= e^{-\frac{1}{2}\langle (\mathbf{q}\cdot(\mathbf{u}_k(t) - \mathbf{u}_j(t)))^2 \rangle_t} \\ &= e^{-\frac{1}{2}q^2\langle u_{qk}^2(t)\rangle_t} e^{-\frac{1}{2}q^2\langle u_{qj}^2(t)\rangle_t} e^{-\frac{1}{2}q^2\langle u_{qj}(t)u_{qk}(t)\rangle_t} \\ &= e^{-\frac{1}{2}q^2\langle u_{qk}^2(t)\rangle_t} e^{-\frac{1}{2}q^2\langle u_{qj}^2(t)\rangle_t} \\ &\quad \times \left(1 + \left\{ e^{-\frac{1}{2}q^2\langle u_{qj}(t)u_{qk}(t)\rangle_t} - 1 \right\}\right) \end{aligned} \tag{2.34}$$

where the new variable u_{qj} is the projection of $\mathbf{u}_j$ onto the $\mathbf{q}$ vector and was used to simplify the algebra. The last line of equation (2.34) was rewritten for reasons that will be made

clear in what follows. The scattered intensity is then obtained as

$$\begin{aligned} I(\mathbf{q}) = &\sum_j \sum_k f_j(\mathbf{q}) e^{-\frac{1}{2}q^2\langle u_{qj}^2(t)\rangle_t} e^{-i\mathbf{q}\cdot\mathbf{R}_j)} f_k^*(\mathbf{q}) e^{-\frac{1}{2}q^2\langle u_{qk}^2(t)\rangle_t} e^{i\mathbf{q}\cdot\mathbf{R}_k)} \\ &+ \sum_j \sum_k f_j(\mathbf{q}) e^{-i\mathbf{q}\cdot\mathbf{R}_j} f_k^*(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{R}_k} \left\{ e^{-\frac{1}{2}q^2\langle u_{qj}(t)u_{qk}(t)\rangle_t} - 1 \right\} \end{aligned} \tag{2.35}$$

The first term of equation (2.35) looks like our standard elastic scattering with a damping term if we replace the scattering form factor with $f(\mathbf{q})e^{-\frac{1}{2}q^2\langle u_{qj}^2(t)\rangle_t}$, where $\langle u_{qj}^2(t)\rangle_t$ is the root mean square (r.m.s.) displacement of atom j. The first effect of the harmonic thermal fluctuations is to reduce the intensities measured by the Debye-Waller factor (DWF) $e^{-q^2\langle u_q^2(t)\rangle_t} = e^{-2M}$. The second term of equation (2.35) represents what is known as thermal diffuse scattering of TDS. It is generally weak and has broad features and for the purpose of this thesis, represents a background intensity which needs to be subtracted.

The derivation above of the Debye-Waller factor was done considering individual atoms. However, the same formalism applies if the building blocks of the sample are molecules and a Debye-Waller factor for each molecule can be defined. In this case, the intensity from a supermolecule made up of proteins with r.m.s displacement $\langle u_{\text{mol}}^2\rangle$ is given by

$$I(\mathbf{q}) = F_{\text{mol}}^2(\mathbf{q}) e^{-2M_{\text{mol}}} S^2(\mathbf{q}) \tag{2.36}$$

where the molecular form factor is also attenuated by the atomic Debye-Waller factors

$$F_{\text{mol}}(\mathbf{q}) = \sum_j f_j(\mathbf{q}) e^{-M_j} e^{-i\mathbf{q}\cdot\mathbf{r}_j} \tag{2.37}$$

The next question to ask is what happens to the scattered intensity when we have a distribution of different structures scattering incoherently. Let us assume a sample is composed of many small clusters of a certain protein. The clusters are separated on average by a distance larger than one coherence length of the beam so that we can consider them to scatter incoherently. Each cluster contains the same amount of proteins but each one can have a slightly different structure due to small packing errors, that is each protein is represented by an average position it assumes in all clusters and a small displacement from that average position characteristic to each cluster. As long as the disorder in the position of each protein from cluster to cluster can be accurately described by gaussian statistics, than the same result as above applies. That is the overall sum of the intensities from the distribution of clusters is reduced by the Debye-Waller factor $e^{-2M} = e^{-q^2\langle u^2\rangle}$ where now the r.m.s. displacement is no longer a time averaged value but rather an average over the distribution.

With this, we are now in position to calculate the scattered intensity from a distribution of disordered supermolecules of different sizes undergoing Brownian motion. Denoting the temporal Debye-Waller factor as e^{-M_t} and the configurational Debye-Waller factor as e^{-M_C}, we get

$$
\begin{aligned}
I(\mathbf{q}) =& \text{TDS} + \text{Noise} + \sum_n p_n \left[F_{\text{mol}}(\mathbf{q}) S_n(\mathbf{q})\right]^2 \\
=& \text{Constant} + \sum_N p(N) \left[\sum_j \left\{ f_j(\mathbf{q}) e^{-(M_t)_j} e^{-i\mathbf{q}\cdot\mathbf{r}_j} \right\}\right]^2 \\
& \times \left[\left\{ e^{-(M_C)_N} \langle S_N(\mathbf{q}) \rangle \right\}\right]^2
\end{aligned}
\tag{2.38}
$$

where p_n is the probability of having a supermolecule of size N and configuration n, $p(N) = \langle p_n(N) \rangle$ is the probability of having a supermolecule of size N independent of configuration n, $S_n(\mathbf{q})$ is the structure factor of configuration n and $\langle S_N(\mathbf{q}) \rangle$ is the structure factor of an ideal,*i.e.* without thermal fluctuations, average supermolecule of size N. The intensity can be calculated from the average structures times a Debye-Waller factor with some background noise and thermal diffuse scattering which can be lumped into a constant factor. Once again, the expression of equation (2.38) can be factored if $F_{\text{mol}}(\mathbf{q}) = F_{\text{mol}}(q)$, *i.e.* the form factor is isotropic. In this case, the form factor can be pulled out of the summation over N and one only needs to calculate the structure factors of different supermolecules, multiply them by a Debye-Waller factor to dampen the higher q intensity and average over a certain distribution of sizes.

2.2 Small Angle X-Ray Scattering

In the case of an ensemble of identical spherically symmetric particles randomly oriented in a sample, the formalism developed above can be used directly since the random orientation of the particles does not affect the scattering pattern. However, for particles of general shapes not possessing well-defined symmetries, it can be difficult to extract the electron distribution from an orientationally averaged scattering curve. The lack of symmetry leads to scattering profiles with little or no significant features such as peaks in the intensity and many possible structures can give rise to similar scattering curves. It is still possible however to obtain some important information about a sample by studying the scattering of x-rays at small angles [26, 14].

2.2.1 Solution Scattering

Small angle x-ray scattering is a useful experimental technique for studying proteins which cannot be crystallized or studying protein interactions in solution. Proteins are generally large molecules, much larger than the typical solvent molecules like water, in which they are maintained. For this reason, the scattered intensity from proteins differs from zero only in a small range of angles near zero. The scattering from water on the other hand extends to wide angles due to the small size of the water molecule. The scattering curve of water even displays a characteristic broad peak at $q \sim 2.15\text{Å}^{-1}$ representative of the oxygen-oxygen distance of adjacent water molecules. The O-O distance is the characteristic distance because hydrogen atoms (Z=1) are essentially invisible to x-rays compared with oxygen atoms (Z=8) which scatter 64 times more. The scattering at angles below that of

the water peak is therefore dominated by the scattering of the proteins in solution. The small contribution of the solvent can be measured using a sample containing only solvent and then subtracted from the sample containing proteins.

The presence of the solvent also has the effect of effectively reducing the electron density of the protein. That's because the scattering depends on the difference in density. Considering the solvent to be a homogeneous medium of a certain density ρ_0, then the effective electron density of a protein in solution is $\Delta\rho(\mathbf{r}) = \rho(\mathbf{r}) - \rho_0$. The contrast of density between the medium and the sample is the relevant quantity determining the overall intensity. The contrast however does not affect the shape of the scattering curve but only its scale.

Proteins in solution are expected to be randomly oriented and therefore the scattered intensity is given by the orientational average of the square of the Fourier transform of a single protein in the dilute limit where the structure factor is equal to unity. The electron density itself may prove difficult to retrieve due to the fact that the phase of the scattered waves is lost in the measurement and therefore inversion to real space cannot be directly performed. This will be discussed further in Section 2.3. A more useful expression for the intensity is

$$\begin{aligned}\langle I(\mathbf{q})\rangle =& \left\langle \int\int \rho(\mathbf{r})\rho(\mathbf{r}')e^{-i\mathbf{q}\cdot\mathbf{r}}e^{-i\mathbf{q}\cdot\mathbf{r}'}\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \right\rangle \\ =& \left\langle V\int \gamma(\mathbf{R})e^{-i\mathbf{q}\cdot\mathbf{R}}\mathrm{d}\mathbf{R} \right\rangle \\ =& 4\pi\int g(R)\frac{\sin qR}{qR}\mathrm{d}R\end{aligned} \tag{2.39}$$

where $\mathbf{R} = \mathbf{r} - \mathbf{r}'$, V is the volume of the particle and the autocorrelation function $\gamma(\mathbf{R})$ and the pair distribution function $g(R)$ have been defined

$$\begin{aligned}\gamma(\mathbf{R}) =& \frac{1}{V}\int \rho(\mathbf{r})\rho(\mathbf{r}')\mathrm{d}\mathbf{r} \\ g(R) =& R^2V\gamma(R) = R^2V\left\langle\gamma(\mathbf{R})\right\rangle\end{aligned} \tag{2.40}$$

The intensity is then related to the pair distribution function by a simple integral and no knowledge of the phases of the scattered x-rays is required. The usefulness of equation (2.39) stems from the fact that the autocorrelation function can be calculated from first principles using interaction potentials such as the hard sphere and adhesive hard sphere potentials [74, 37, 62].

2.2.2 Low Q Approximation

For particles with low symmetry, the orientational averaging has the effects of smearing the shape of the particle and washes out the features of the scattering pattern by averaging them out. Useful information can be obtained in the low q limit which is calculated by expanding the $\frac{\sin qR}{qR}$ in equation (2.31), keeping only the first term.

$$\begin{aligned} I(q) =& I(0)\left[1 - \frac{q^2}{3}\frac{\int \rho(\mathbf{r}) r^2 \mathrm{d}\mathbf{r}}{\int \rho(\mathbf{r})\mathrm{d}\mathbf{r}} + \cdots\right] \\ =& I(0)\left[1 - \frac{q^2 R_g^2}{3} + \cdots\right] \\ \approx& I(0) e^{-\frac{q^2 R_g^2}{3}} \end{aligned} \tag{2.41}$$

The result in equation (2.41) was first obtained by Guinier [26] and is known as the exponential approximation or as Guinier's law. R_g is the radius of gyration of the scattering particle and is defined in equation (2.41). The interesting consequence of Guinier's law is that two objects of very different shape that happen to have the same radius of gyration will have the same SAXS curve at low q. The radius of gyration is readily obtained from the slope of the $\ln(I)$ vs q^2 plot, known as a Guinier plot.

2.2.3 High Q Approximation

The asymptotic shape of the SAXS curve can be obtained by rewriting equation (2.39) using the fact that $I(q)$ and $g(R)$ are Fourier transforms of each other.

$$g(R) = R^2 V \gamma(R) = \frac{1}{4\pi^3}\int qRI(q)\sin qR \mathrm{d}\mathbf{R} \tag{2.42}$$

Since they are related by a Fourier transform, the high q value of $I(q)$ will depend on the $R \to 0$ limit of $g(R)$, hence on $\gamma(0)$. We expand $\gamma(r)$ near zero [26]

$$\gamma(r) \approx 1 - \frac{SR}{4V} \tag{2.43}$$

where S is the surface area of the particle. Putting this back into equation (2.39) and integrating by parts, we get that that asymptotic limit of the intensity is

$$\begin{aligned} I(q) \approx& \frac{2\pi S}{q^4} \\ \approx& \frac{A}{q^4} + B \end{aligned} \tag{2.44}$$

where B is a constant signal level arising from noise and incoherent scattering.

2.3 Coherent X-Ray Diffraction

Chapter 3 of this thesis will discuss the results obtained when performing Coherent X-ray Diffraction (CXD) on crystals of proteins. The concepts needed to understand these diffraction results are introduced here. Some of the developmental and technical details of the technique can be found in the doctoral dissertations of Pitney [61], Williams [88] and Pfeifer [60].

2.3.1 Crystal Diffraction

We start by considering a perfect infinite crystal of proteins, that is proteins are arranged on a periodic infinite Bravais lattice [3]. The scattered intensity from such an infinite crystal of proteins can be calculated by going back to the equation derived above for one supermolecule, *i.e* equation (2.24). The electron density is given by the convolution of the electron density of a single protein and the infinite series of delta functions representing the location of every lattice point. We define the lattice positions or direct lattice vectors as

$$\mathbf{R}_n = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c} \tag{2.45}$$

where the n_i's can take on any integer value from $-\infty$ to $+\infty$ and the vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are known as the primitive lattice vectors. The primitive vectors must not all lie in the same plane for a 3-dimensional crystal. The Fourier transform of an infinite set of delta functions is another infinite set of delta functions or in other words, the Fourier transform of the direct lattice is itself a Bravais lattice known as the reciprocal lattice. The reciprocal lattice vectors are defined as

$$\mathbf{G}_n = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \tag{2.46}$$

where the primitive reciprocal lattice vectors are given by

$$\mathbf{a}^* = 2\pi\frac{\mathbf{b}\times\mathbf{c}}{\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}} \quad \mathbf{b}^* = 2\pi\frac{\mathbf{c}\times\mathbf{a}}{\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}} \quad \mathbf{c}^* = 2\pi\frac{\mathbf{a}\times\mathbf{b}}{\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}}. \tag{2.47}$$

It follows directly from equation (2.47) that $\mathbf{a}\cdot\mathbf{a}^* = 0$, *i.e.* a direct lattice vector is perpendicular to its corresponding reciprocal lattice vector. In a perfect infinite crystal, the only points in reciprocal space where intensity is non-zero are the reciprocal lattice points. The scattered amplitude is then given by, using the convolution theorem

$$E(\mathbf{q}) = F_{\text{mol}}(\mathbf{q}) \sum_{\{h,k,l\}=-\infty}^{\infty} \delta(\mathbf{q} - \mathbf{G}_{\{h,k,l\}}) \tag{2.48}$$

The intensity is zero everywhere except on the reciprocal lattice points where the intensity is proportional to the molecular form factor at the given $\mathbf{q}$. The peaks in intensity observed are called Bragg peaks.

Every lattice point of a crystal does not necessarily need to be occupied by a single atom or single molecule. We define as a unit cell the arrangement of atoms which is reproduced at every lattice point. A primitive unit cell contains a single atom.

2.3.2 Miller Indices

The diffraction from crystals can be in a rather simplistic fashion interpreted as the reflection of x-rays from lattice planes. This provides an intuitive but incomplete picture of x-ray diffraction. The diffraction from a set of lattice planes gives rise to a Bragg peak at a $\mathbf{q}$ vector normal to these lattice planes. Bragg peaks are generally labeled using their h, k, l values defined above. A Bragg peak with h, k, l indices is identified as coming from a set of

lattice planes perpendicular to the vector (h, k, l). These indices are called the Miller indices. It is necessary here to define some conventions used when describing crystal directions.

Lattice planes are are labeled with parentheses. The notation (hkl) denotes a lattice plane normal to the (h, k, l) vector. Negative values of the indices are denoted as $-h = \bar{h}$. Due the the perpendicularity of direct lattice planes and the reciprocal lattice, (hkl) indicates directions in the reciprocal lattice.

When one wishes to describe directions in the direct lattice using the values of n_1, n_2 and n_3 defined above, a similar notation is used but this time with square brackets to avoid confusion. Therefore, $[n_1 n_2 n_3]$ indicates a real lattice direction.

Due to the symmetry of a given lattice, there may be multiple symmetry related Bragg directions. For example, in a cubic lattice every permutation of the Miller indices as well as a change of sign describes an equivalent reflection. Therefore, when one refers to the family of equivalent reflections, the following notation is used $\{hkl\}$. When referring to the $\{111\}$ Bragg peaks, one then means the whole family of Bragg peaks including all the following : the (111), $(\bar{1}11)$, $(1\bar{1}1)$, $(11\bar{1})$, $(\bar{1}\bar{1}1)$, $(\bar{1}1\bar{1})$, $(1\bar{1}\bar{1})$, $(\bar{1}\bar{1}\bar{1})$.

Similarly, a family of direct lattice directions are labeled as $\langle n_1 n_2 n_3 \rangle$.

2.3.3 Finite Size Crystal

We now turn to the problem of calculating the scattered intensity from a finite crystal. More specifically, a crystal whose volume is smaller than a coherence volume of the x-ray beam so that it is coherently illuminated. For this discussion, we are still treating the scattering as elastic and within the kinematical approximation. We can write the electron density of a small crystal as the electron density of an infinite crystal multiplied by the shape function of the crystal denoted as $s(\mathbf{r})$ which is equal to one inside the crystal and zero outside. Using the convolution theorem once again, the scattered amplitude is then written as a convolution of the Fourier transform of the infinite crystal and the Fourier transform of the shape function.

$$\begin{aligned} E(\mathbf{q}) =& \int [\rho_{\text{mol}}(\mathbf{r}) \otimes \rho_{\text{lat}}(\mathbf{r})] s(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{r}} \mathrm{d}\mathbf{r} \\ =& S(\mathbf{q}) \otimes [F_{\text{mol}}(\mathbf{q}) \sum_n \delta(\mathbf{q} - \mathbf{G}_n)] \\ =& F_{\text{mol}}(\mathbf{q}) S(\mathbf{q}) \otimes \sum_n \delta(\mathbf{q} - \mathbf{G}_n) \\ =& F_{\text{mol}}(\mathbf{q}) \sum_n S(\mathbf{q} - \mathbf{G}_n) \end{aligned} \tag{2.49}$$

In equation (2.49), $\rho_{\text{mol}}(\mathbf{r})$ is the electron density of the molecule making up the unit cell of the crystal, $\rho_{\text{lat}}(\mathbf{r})$ is the lattice positions and $S(\mathbf{q})$ is the Fourier transform of the shape function. In the last step of equation (2.49) we explicitly wrote out the convolution of $S(\mathbf{q})$ with the series of delta functions of the reciprocal lattice. The physical consequence of equation (2.49) is that the Fourier transform of the shape of the crystal which is expected, due to the large size of the crystal to be measured in a small region of reciprocal space around the origin is now reproduced at every Bragg peak of the underlying crystal. The

crystalline nature of the large particle coherently illuminated reproduces the small angle scattering pattern at every Bragg peak.

The size of the crystals studied is in the micron range so that it matches closely the coherence of the x-ray beam and the signal level is measurable. The crystal size is therefore very large compared with the size of a unit cell of inorganic crystals ($\sim$ 5 Å). This means that the Fourier transform of the shape extends in reciprocal space only to a fraction of the size of the reciprocal unit cell and intensity is measurable only over a small range of $\mathbf{q}$ near a Bragg peak.

However, when performing CXD experiments on crystals of proteins, the unit cell of the crystal is in the range of 100 Å. This means that Bragg peaks are roughly 20 times closer in reciprocal space, but for a crystal of the same size, the Fourier transform of the shape extends over the same $\mathbf{q}$ range as that of the inorganic crystal. Even though one cannot measure the CXD pattern at atomic resolution, for protein crystals it is possible to measure the diffraction at molecular resolution and get information about the structure at the unit cell level.

In order to calculate the scattered intensity near a Bragg peak $\mathbf{G}_n$, we define the vector $\mathbf{Q} = \mathbf{q} - \mathbf{G}_n$. The intensity is given by the square of the amplitude, approximating the molecular form factor as having the constant value it has at the Bragg position, $F_{\mathrm{mol}}(\mathbf{Q}) = F_{\mathrm{mol}}(\mathbf{G_n})$

$$I(\mathbf{Q}) = F^2_{\mathrm{mol}}(\mathbf{G_n}) \left| \int s(\mathbf{r}) e^{-i\mathbf{Q}\cdot\mathbf{r}} \mathrm{d}\mathbf{r} \right|^2 \tag{2.50}$$

For any real object, the scattering around the origin must be centrosymmetric, *i.e* $I(\mathbf{q}) = I(-\mathbf{q})$. Therefore, the CXD pattern near a Bragg peak is also centrosymmetric around $\mathbf{G}_n$ so that $I(\mathbf{Q}) = I(-\mathbf{Q})$ for a real object.

Given that for protein crystals we expect to be able to measure the intensity near a Bragg peak with near molecular resolution, it becomes interesting to ask what happens to the intensity in the presence of lattice defects or strain [69].

2.3.4 Strain

The presence of a strain field in the small crystal does not affect the shape of the crystal but changes the internal structure by introducing a small displacement $\mathbf{u}(\mathbf{R}_n)$ of the unit cells from their ideal lattice positions, $\mathbf{R}_n$. The scattered amplitude near a Bragg peak is then written as

$$\begin{aligned} E(\mathbf{q}) =& \int [\rho_{\mathrm{mol}}(\mathbf{r}) \otimes \rho_{\mathrm{lat}}(\mathbf{r})] s(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{r}} \mathrm{d}\mathbf{r} \\ =& F_{\mathrm{mol}}(\mathbf{q}) \int s(\mathbf{r}) \sum_{n=-\infty}^{\infty} \delta(\mathbf{r} - \mathbf{R}_n - \mathbf{u}(\mathbf{R}_n)) e^{-i\mathbf{q}\cdot\mathbf{r}} \mathrm{d}\mathbf{r} \end{aligned} \tag{2.51}$$

We first perform the integral and obtain using the picking property of the delta function

$$\begin{aligned} E(\mathbf{q}) =& F_{\mathrm{mol}}(\mathbf{q}) \sum_{n=-\infty}^{\infty} s(\mathbf{R}_n + \mathbf{u}(\mathbf{R}_n)) e^{-i\mathbf{q}\cdot(\mathbf{R}_n + \mathbf{u}(\mathbf{R}_n))} \\ \simeq& F_{\mathrm{mol}}(\mathbf{q}) \sum_{n=-\infty}^{\infty} s(\mathbf{R}_n) e^{-i\mathbf{q}\cdot\mathbf{R}_n} e^{-i\mathbf{q}\cdot\mathbf{u}(\mathbf{R}_n)} \end{aligned} \tag{2.52}$$

where the approximation that the shape function was the same at $\mathbf{R}_n$ and $\mathbf{R}_n + \mathbf{u}(\mathbf{R}_n)$ was made. We now reintroduce integrals over $\mathbf{r}$ of delta functions using again the picking property of the delta function.

$$\begin{aligned} E(\mathbf{q}) \simeq& F_{\mathrm{mol}}(\mathbf{q}) \int \sum_{n=-\infty}^{\infty} s(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{r}} e^{-i\mathbf{q}\cdot\mathbf{u}(\mathbf{r})} \delta(\mathbf{r} - \mathbf{R}_n) \mathrm{d}\mathbf{r} \\ \simeq& F_{\mathrm{mol}}(\mathbf{q}) \int s(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{u}(\mathbf{r})} e^{-i\mathbf{q}\cdot\mathbf{r}} \mathrm{d}\mathbf{r} \otimes \sum_{n=-\infty}^{\infty} \delta(\mathbf{q} - \mathbf{G}_n) \\ \simeq& F_{\mathrm{mol}}(\mathbf{q}) S'(\mathbf{q}) \otimes \sum_{n=-\infty}^{\infty} \delta(\mathbf{q} - \mathbf{G}_n) \\ \simeq& F(\mathbf{q}) \sum_{n=-\infty}^{\infty} S'(\mathbf{q} - \mathbf{G}_n) \end{aligned} \tag{2.53}$$

The form of equation (2.53) is similar to that of equation (2.49) except that the shape function of the particle is no longer real but no has an imaginary part dependent on the projection of the strain vector onto the reciprocal lattice vector. This implies that the presence of strain within the crystal can be modeled as an object with an imaginary component to the electron density provided the strain is small compared with the size of the unit cell. The CXD pattern measured is now the square of the Fourier transform of this complex shape function. It is worth noting that as before the intensity is centrosymmetric about the origin in reciprocal space but the presence of strain coupling to the $\mathbf{q}$ vector introduces asymmetries around the Bragg peaks away from the origin so that $I(\mathbf{Q}) \neq I(-\mathbf{Q})$ in the presence of strain.

$$I(\mathbf{Q}) = F_{\mathrm{mol}}^2(\mathbf{G_n}) \left| \int s(\mathbf{r}) e^{-i\mathbf{Q}\cdot\mathbf{u}(\mathbf{r})} e^{-i\mathbf{Q}\cdot\mathbf{r}} \mathrm{d}\mathbf{r} \right|^2 \tag{2.54}$$

2.3.5 2D Cuts Through Diffraction Pattern

It is impossible to measure an entire 3-dimensional coherent diffraction pattern all at once. The best one can do in a single image is to measure a 2D cut through some plane in the diffraction pattern using a 2D detector. A complete 3D diffraction pattern can be obtained by putting many of theses 2D slices together. Only a complete pattern can give the whole information about the shape of the illuminated particle. As will be discussed in Chapter 3, it may prove difficult or even impossible to measure a complete 3D pattern. The question to ask is whether useful information can be obtained from a single 2D slice through the CXD pattern.

The direction of the slice through reciprocal space is determined by the angle of the 2D detector. The direction normal to the plane of the detector is the same direction as the

direction of the scattered radiation hitting the center of the detector k' defined on Figure 2.1. We choose this direction as the Q_z direction with the in-plane directions as Q_x and Q_y. The scattered amplitude is written as

$$E(Q_x, Q_y, Q_z) = \int\int\int \rho(x, y, z) e^{-i(Q_x x + Q_y y + Q_z z)} \mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{2.55}$$

If we assume that the sample is perfectly aligned so that the 2D slice goes exactly through the center of the diffraction pattern, then $Q_z = 0$ and we get

$$\begin{aligned} E(Q_x, Q_y, 0) &= \int\int\int \rho(x, y, z) e^{-i(Q_x x + Q_y y)} \mathrm{d}x\mathrm{d}y\mathrm{d}z \\ &= \int\int \tilde{P}_z(\rho(x, y, z)) e^{-i(Q_x x + Q_y y)} \mathrm{d}x\mathrm{d}y \end{aligned} \tag{2.56}$$

where the projection operator $\tilde{P}_z$ has been defined as

$$\tilde{P}_z(\rho(x, y, z)) = \int \rho(x, y, z) \mathrm{d}z \tag{2.57}$$

The scattered intensity of a single 2D slice is then the square of the Fourier transform of the projection of the electron density onto a plane parallel to the detector plane. The effects of misalignment of the sample giving rise to a 2D slice which does not go exactly through the center of the pattern were investigated by Pfeifer [60]. If the slice if off center in the Q_z direction by an amount ΔQ_z, the scattered amplitude is given by

$$\begin{aligned} E(Q_x, Q_y, \Delta Q_z) &= \int\int\int \rho(x, y, z) e^{-i(Q_x x + Q_y y + \Delta Q_z)} \mathrm{d}x\mathrm{d}y\mathrm{d}z \\ &= \int\int \tilde{P}_z \left(\rho(x, y, z) e^{-i\Delta Q_z z}\right) e^{-i(Q_x x + Q_y y)} \mathrm{d}x\mathrm{d}y \end{aligned} \tag{2.58}$$

The result is that the intensity of an off center slide through the diffraction pattern is the square modulus of the projection of the electron density times a complex phase factor. The interpretation of the results from any off center slice can prove to be tricky. However, the effect of misalignment should be the introduction of a constant phase gradient across the recovered electron density so the real space projection of the electron density should be recoverable by flattening the phase across the reconstructed object.

2.3.6 Phase Retrieval

As previously mentioned the complex amplitude of the scattered radiation cannot be measured. Only the intensity is measured, that is the square modulus of the amplitude and all information about the phase of the x-rays is lost in the measurement. In order to retrieve the complex shape function $s(\mathbf{r})e^{-i\mathbf{G}_n \cdot \mathbf{u}(\mathbf{r})}$ of the illuminated particle or crystal, one must find a way to retrieve the phases of the scattered radiation. Once these phases ($\phi(\mathbf{Q})$) are known, a simple Fourier transform will yield the desired shape function.

$$s(\mathbf{r})e^{-i\mathbf{G}_n \cdot \mathbf{u}(\mathbf{r})} = \mathcal{F}^{-1}[\sqrt{I(\mathbf{Q})}e^{i\phi(\mathbf{Q})}] \tag{2.59}$$

The problem of retrieving the lost phases is known as the phase problem of crystallogra-

phy. Many methods exists for retrieving the phases and the method of choice depends of the experiment performed. For example, in protein crystallography where the intensities from thousands of Bragg peaks from a high quality large crystal of proteins are measured, the technique of Multiple Anomalous Dispersion (MAD) [14] is often used to retrieve the phases. The technique consists of measuring the diffraction using x-rays of energy just below and just above the absorption edge of a specific atom. The significant difference in scattering power at the two energies allows one to narrow down the phase of the measurements to only a small set of values out of which the real solution can be found.

In our case, we focus our attention on a class of iterative phasing methods known as the Gerchberg-Saxton(GS) method [23]. The method was first developed for phase retrieval of electron diffraction data but it was later recognized by Millane that it could be useful in crystallography [53]. Many generalizations of the original GS algorithm have been proposed over the years by Fienup [15, 16, 17] and Millane [54, 55] to name a few. These algorithms and their implementation were discussed and developed extensively by Williams in his doctoral dissertation [88]. The computer programs performing the iterative retrieval of the phases of the diffracted intensities used in the data analysis of Chapter 3 were developed primarily by Pitney [61] and Williams [88]. We will discuss here only the main ideas behind them without sinking too deeply into the details.

An initial requirement to make the phase retrieval possible is for the data to be oversampled, a fact realized by Sayre [72, 73]. Generally speaking oversampling implies that the data is sampled at a frequency which is higher than the highest frequency present in the signal. When dealing with x-ray scattering, the highest frequency in the scattered amplitude is determined by the size of the sample illuminated. Given a particle of size d, the corresponding frequency in reciprocal space is $f_0 = \frac{d}{2\pi}$. This implies, according to Shannon's theorem that the amplitude must be sampled at least at twice this frequency to obtain all the information in the signal. This sampling frequency is well-known as the Nyquist frequency. It implies that at least two measurements are made between successive peaks in the amplitude function.

However, when measuring the intensity and not the amplitude of the scattered wave, the measurement is no longer the Fourier transform of the electron density but rather the square modulus of the Fourier transform of the electron density, or similarly the Fourier transform of the autocorrelation function of the electron density. For a particle of size d, the autocorrelation function extends in real space to a size of $2d$, which implies the intensity must be sampled at twice the Nyquist frequency or $f_0 = \frac{4d}{2\pi}$. Rewriting this in terms of resolution in reciprocal space, we get that the intensity function must be sampled at a resolution of $\delta q = \frac{2\pi}{4d}$.

The basic idea behind the iterative phasing methods used, shown in Figure 2.3, is to start with the square root of the intensity data and then assign random phases to each data point to generate a set of scattered amplitudes consistent with the measured intensities but with random phases. An inverse Fourier transform is performed on this array of complex amplitudes yielding a real space complex density function which is not the actual solution seeked unless by some chance the random phases happened to all be correct. At this point some constraints are applied to the real space array. The possible choices of constraints are

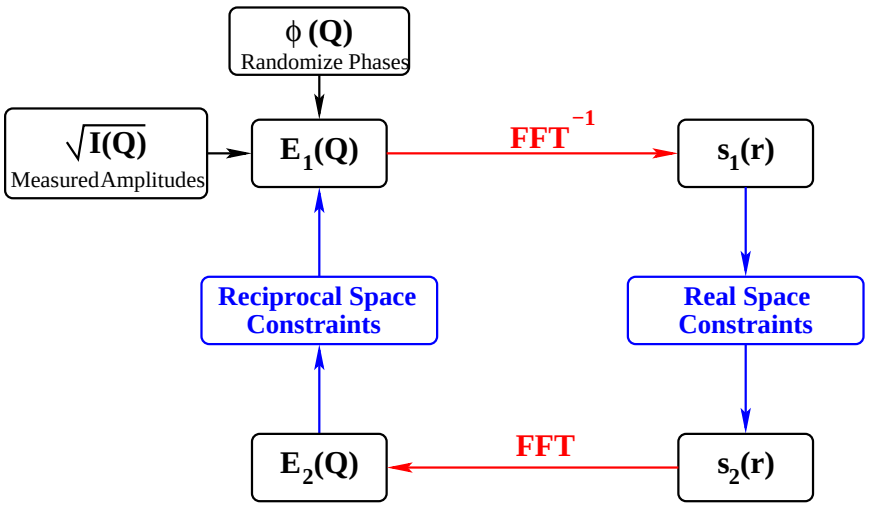


Figure 2.3:

numerous and this is where the differences between the original GS algorithm and those of Fienup and Millane are found. The basic constraint one may wish to use are positivity of the electron density, requirement that the density be real and restriction to a finite support. More complicated constraints are applied to speed up convergence or avoid stagnation but in general they consist primarily of the three main ones mentioned in some desired combination.

The use of a finite support requires some prior knowledge of the particle illuminated. The intensity pattern itself may provide the necessary information. The coherent diffraction of a small particle should lead to the presence of oscillations in the intensity function in a manner similar to the diffraction of visible light by a slit [29]. The spacing of the fringes is given by $\frac{2\pi}{d}$ where d is again the size of the particle. This allows one to set the real-space constraint to a finite support of size d, that is all the values of the density a set to zero outside a box of size d. Other constraints may or may not be applied at the same time but finite support is generally required due to the fact that that data is oversampled in reciprocal space. The extra number of variables in reciprocal space needs to be balanced by fixing a large number of points in the real space array.

After applying the real space constraints, the array is Fourier transformed back to reciprocal space, where another constraint must be enforced so that the inverse Fourier transform of the next cycle does not simply backtrack to the same electron density. The constraint applied here is to keep all the phases obtained during the last cycle but replacing all the amplitudes with the measured amplitudes. In theory, cycling through these steps many times should force convergence of the retrieved phases toward the real solution. The existence of a unique solution can be proven except for some rare pathological cases [5, 4] and was studied and discussed at length by Williams [88]. Once the algorithm is satisfied that a solution was obtained, the measured amplitudes are combined with the retrieved phases and the complex reconstructed object is calculated from the inverse Fourier transform of the complex scattered amplitude.

Chapter 3

Coherent X-Ray Diffractive Imaging from Protein Crystals

The basic idea at the start of the work described in this thesis was to use x-rays to study protein crystal formation. More precisely, to use the newly developed technique of Coherent X-Ray Diffraction (CXD) introduced in Section 2.3. The recent development of brilliant coherent sources of x-rays at synchrotron sources around the world has made such experiments possible.

Coherent x-ray scattering in general is a rapidly growing field as described in a review by Van der Veen [82]. In the more narrow field of coherent x-ray diffractive imaging, recent progress has been made. Early results on two dimensional samples were obtained by Vartaniants *et al.* using the coherence characteristics of the beam to reconstruct the surface morphology of the Si(111) crystal surface [83]. The technique was then extended to three dimensional samples. Robinson *et al.* [70] first reconstructed the shapes of gold nanocrystals in two dimensions using iterative reconstruction. The technique was later extended to three dimensions [89], again using gold crystals of size near 0.1 microns. Arrays of ordered small crystalline particles can also be imaged using this technique. Vartaniants and Robinson [84] published a paper in which they discuss simulations and real measurements of the diffraction pattern of quantum dot arrays under coherent and partially coherent illumination. Recently, the use of the phase retrieval algorithms was extended to electron diffraction by Zuo *et al.* [93]. Applying the same ideas to an electron diffraction pattern, they were able to retrieve the phases of the scattered waves and reconstructed in three dimensions a double-wall carbon nanotube at a resolution of 1Å.

The technique is not limited to crystalline samples however. The greatest promise of coherent x-ray imaging may lie in a so-called lensless x-ray microscope [77, 50] which could be used to image, at theoretically atomic resolution, non-crystalline samples [51, 49]. The phase retrieval in this case is done the same way as for crystalline samples. The difference only lies in the specifics of the experiment. In crystalline samples, the Fourier transform of the shape of the object is measured near a Bragg peak, at a large angle away from the direct beam. This makes the measurement simpler because it is easier to collect the entire 3D pattern and there is no contamination of the signal by the direct beam. Another advantage of measuring the CXD pattern near a Bragg peak is the possibility of measuring the strain field as discussed in Section 2.3.4.

In this chapter we discuss results on CXD experiments performed on crystalline biological samples. The technique used is similar to that used in gold nanocrystals experiments of Robinson *et al.* [70, 89]. Some extra difficulties are expected when dealing with biological

samples and they are discussed. Calculations by Marchesini *et al.* [43] indicate that exposure times required for CXD measurements of biological samples are near the limits of tolerance of such samples but the technique should still be applicable. The goal was to measure the complete diffraction pattern of a micron sized crystal of proteins and invert this data back to real-space using iterative phasing methods to determine the shape and size of the crystal.

3.1 Facility and Equipment

All the x-ray diffraction and scattering experiments discussed in this thesis were performed at Sector 34-ID-C of the Advanced Photon Source (APS) located on the site of Argonne National Laboratory (ANL) in Argonne, Illinois.

3.1.1 Source

The APS is what is known as a third generation synchrotron source of x-rays, though it technically is a storage ring and not a synchrotron. A beam of electrons is accelerated to energies of 7GeV and stored in the ring. The APS ring is made of 40 sectors, each of which is made of a bending magnet, *i.e.* a curve section, and of a straight section. The high energy electrons travel through a straight section at constant velocity but are accelerated toward the center of the ring when passing through a bending magnet. The magnetic fields are calibrated to maintain the electrons on a stable orbit in the ring. As discussed in Chapter 2 a charged particle like an electron, when accelerated will emit photons. At such high energy and at the operating current of roughly 100mA, the electrons emit a large number of x-rays. Detailed calculations of the emitted radiation from a synchrotron source can be found in Jackson [35]. First generation synchrotron sources were parasitic to large particle physics experiments. X-rays were produced as a by-product of the particle accelerators used in those experiments. Soon people realized that it would be useful to build dedicated storage rings for synchrotron radiation production. Those are known a second generation sources.

It was later realized that the straight sections between bending magnets could not only also be used to produce x-rays, but could also produce more brilliant beams. This was done by introducing undulators in the straight sections and synchrotron sources with undulators are known as third generation sources [13]. An undulator is an array of magnets with alternating polarity. When an electron passes through the undulator, it is then accelerated back and forth on a roughly sinusoidal track. The magnetic field, the size and the spacing of the magnets is adjusted to keep the orbit of the electrons stable while generating the maximum flux of x-rays at the desired energy. The undulator has two arrays of magnets, one above the beam and the second under it. The spacing of these two arrays can be adjusted. This changes the magnetic field, and thus changes the precession frequency and the energy of the x-ray beam generated.

The energy spectrum of the x-ray beam generated by a bending magnet is fairly broad and is fixed by the size of the ring, *i.e.* the radius of the orbit and the energy of the beam. Undulators, also known as insertion devices or ID's, are much more tunable by changing the gap between the two arrays of magnets. They also generate a much narrower range

of energies. This means that the beam is much more monochromatic and therefore more brilliant. When using the term monochromatic it is in analogy to visible light where white light contains all energies or wavelengths of the visible spectrum while monochromatic light contains a single wavelength or color. The brilliance of the beam is defined as photons per second per mm^2 per mrad per 0.1% bandwidth.

The usefulness of undulators stems from their tunability, high brilliance and small source size giving rise to a large coherence length of the x-ray beam, the latter two being necessary to perform the experiments discussed below.

3.1.2 Monochromator

Even though the undulator at Sector 34 generates a fairly monochromatic beam compared with the beam produced by a bending magnet, the bandwidth of the beam is still larger than desired and the use of a monochromator is necessary. The monochromator used was designed and built by Advanced Design Consulting and is described in detail in reference [38]. It consists of two roughly parallel Si(111) crystals and it is therefore known as a double-crystal monochromator. The incoming x-ray beam hits the first crystal and x-rays of different energies are reflected at different angles according to Bragg's law, $d_{111} = \frac{\lambda}{2\pi \sin\theta}$, where d_{111} is the spacing of the (111) crystal planes of silicon. The angle of the first crystal with respect to the beam ensures that the Laue condition is met only for a single wavelength. The second crystal is oriented so that the reflected beam from the first crystal meets the Laue condition on the second crystal. The doubly reflected beam is then steered into the experimental station. The scattering plane of the monochromator is vertical and thus the monochromatic beam ends up a few millimeters higher than the incoming beam. The monochromator is located 5 meters upstream of the sample position.

3.1.3 Experimental Hutch

The storage ring is inaccessible by the experimenter. Only the undulator gap could be controlled to maximize the flux at the energy desired. As far as we are concerned, the storage ring is a black box producing an x-ray beam at a certain energy and with defined properties such as intensity and coherence volume. The x-ray beam exits the vacuum of the storage ring and passes through a Be window into the vacuum system of the beamline. A complete schematic of the beamline can be seen on the UNICAT website [81]. Sector 34-ID is composed of five hutches labeled as Stations A through E going from upstream of the beam toward downstream. Each of these hutches is made of thick lead walls to shield personnel from the dangerous radiation. Station A contains some optical elements necessary to manipulate the beam as desired. The most important of these is a mirror which is partly inserted in the beam and used as a beam splitter. The undeflected part goes on straight through to the D and E Stations where a second set of independent experiments can be performed. The beam deflected by 5 mrad continues on into Station B where the monochromator is located. The monochromatic beam propagates to the experimental station, Station C, after passing through a second Be window located at the entrance of Station C.

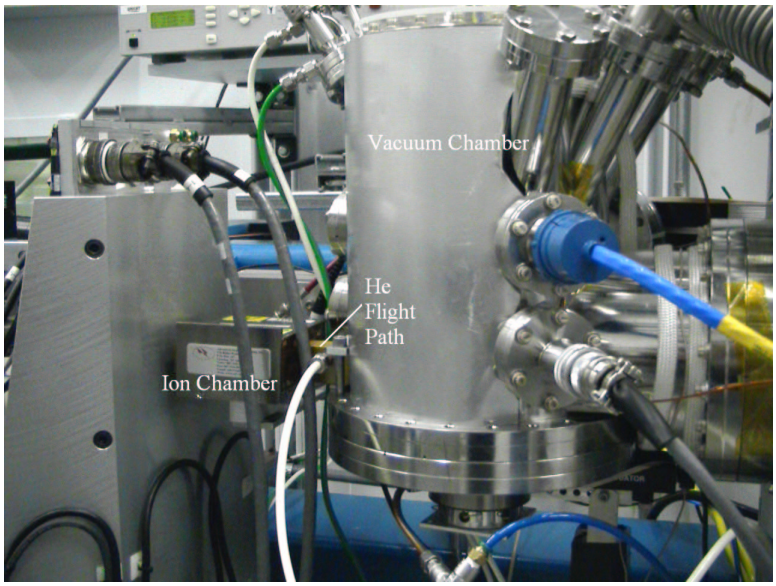


Figure 3.1: The vacuum chamber upstream of the sample where the roller blade slits are located is seen. Immediately after the vacuum chamber is a short flight path filled with helium gas pressed again the beryllium window of the chamber. The Be window is mostly covered by an aluminum cover for protection. An ion chamber is used as a monitor of the incoming beam intensity. The beam travels from right to left

A set of rough slits consisting of two L shaped edges located 3m upstream of the sample collimate the beam by absorbing the undesired tails of it. The beam then passes through a pair of roller blade slits, one in the horizontal direction and one in the vertical. Each roller blade slit consists of two Molybdenum cylinders 3/8" in diameter spaced by 200μm. The gap between the cylinders can be made to appear smaller by a simple rotation of the whole assembly around the center of the gap. This is accomplished by exerting force at the end of a flexpivot located outside of the vacuum chamber where the roller blades are located. A small translation of 0.042mm horizontally and 0.028mm vertically on the actuators exerting the force on the flexpivots was found to lead to a 1μm change in the opening of the slits. Therefore, this design permits very fine control of the size of the beam. The position of the roller blade slits could be controlled by a simple translation of the entire assembly. This allowed the experimenter to select the desired part of the beam. The rough slits were generally set at a 0.5mm gap while the roller-blades were set between 5μm and 50μm. Thus the *hottest* or *cleanest* part of the beam could be selected.

After passing through the roller blade slits, the x-ray beam goes through a large Be window. The roller blade slits are located in a large vacuum chamber which is used for experiments requiring ultra high vacuum (UHV) seen in Figure 3.1. In these experiments, the scattering coming from the sample inside the chamber needs to be measured at high angles. Beryllium is a good material to use as a window because of its low atomic number. It therefore scatters very little and it still is strong enough to hold UHV. This large Be window is a quarter of a cylinder of radius 8" and a height of 12" with the axis of the cylinder in

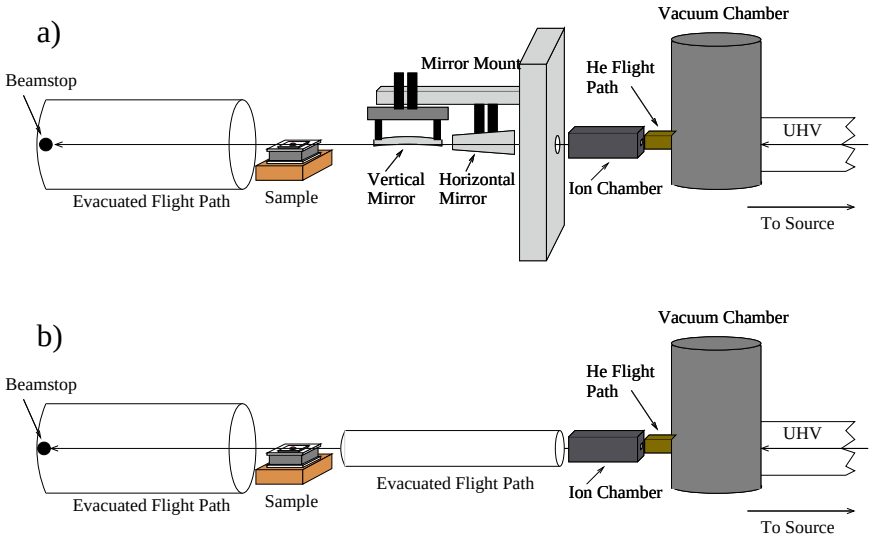


Figure 3.2: Representation of experimental setup with a) KB mirror focusing of the beam and b) with no focusing. The incoming beam direction is from the right toward the left. The direct beam hits the sample and then is absorbed by a beamstop.

the vertical direction. On the picture of Figure 3.1, the beryllium window is covered by a large aluminum shield to protect it.

The x-ray beam makes its first exit out of the UHV at this point. The intense beam in air causes the production of ozone which rapidly corrodes beryllium. Damage to the window is avoided by placing a 5cm helium-filled flight path immediately after the window as shown on Figure 3.1 and schematically on Figure 3.2. This flight path is pressure sealed onto the window. The other end of the flight path is sealed using a Kapton film. Kapton is a polyimide film which does not scatter x-rays very strongly and will also hold a decent vacuum.

The x-ray beam then exits into air where it goes through an ion chamber shown on Figure 3.1 which is used as an incident beam monitor. The ionization of the air by the beam is measured and related to the number of x-rays passing through it. The ion chamber is located 1m from the sample location.

The setup up to this point in the beamline was constant for all the data discussed below. However, some different setups were used for the last meter upstream of the sample depending on the experiment performed. There were 3 basic setups used. The first one was a standard evacuated flight path all the way up to the sample. The second one used focusing optics known as Kirkpatrick-Baez (KB) [32, 34] mirrors between the ion chamber and the sample. The third setup was used for SAXS experiments and will be discussed in Chapter 4. Figure 3.2 shows the first two of these setups. The x-ray beam is pictured as traveling from right to left. This is so because this is the direction of the beam at synchrotron source. The electron beam travels in the ring in a clockwise direction as seen from above so the radiation emitted always comes from the right side when looking at the ring from the outside. On Figure 3.2b) is shown the setup without focusing. In order to minimize air scatter and

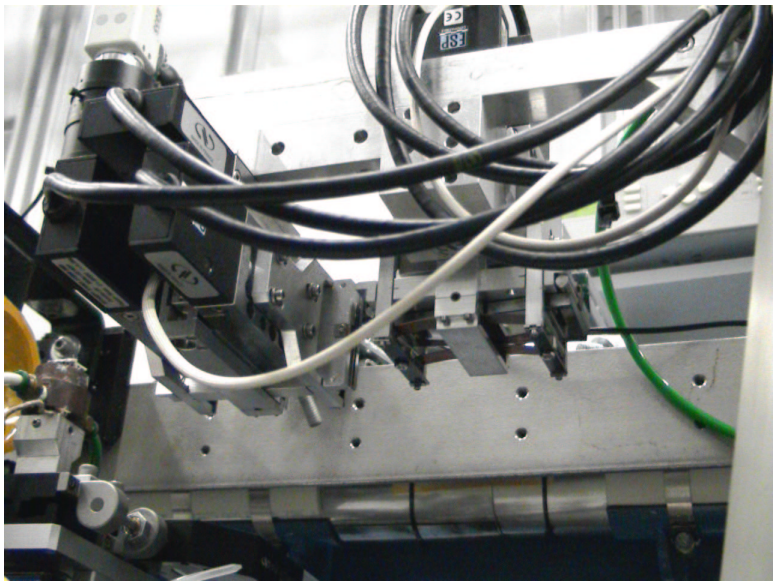


Figure 3.3: The KB mirror assembly is seen from the bottom. A motor controls the position of the mirror while a second one controls its orientation. Two more motors apply force at each end of the silicon mirrors to bend them in order to focus the reflected beam. The vertical focusing is done first (right) followed by the horizontal focusing (left).

also the loss of intensity of the incoming beam as it passes through a thick path of air, an evacuated beam pipe is put between the ion chamber and the sample. Each end of the flight path is seal with a piece of Kapton.

It is possible to increase the signal level by focusing the x-ray beam into a small volume. This was done with KB mirrors as shown on Figure 3.2a). In order to conserve the coherence properties of the beam, the size of the beam impinging on the mirror must be closely matched to the coherence length of the beam. This means a beam of roughly 25-50μm was focused down to size as low as 1.5μm. The KB mirrors are highly reflective silicon crystals bent by applying force at each end. The curvature of the highly polished crystals provides the focusing power. The focusing increases by a factor of close to 1000 the number of x-rays hitting a sample smaller than 1.5μm. This increase in intensity comes at the cost of an increased divergence of the beam and a reduction of the transverse coherence length. However, for samples smaller than 1μm, there is an excess coherence length which can be spared. The focusing is first done in the vertical direction and then in the horizontal. The focal lengths were 200mm and 100mm respectively so that the focal point was exactly at the sample. A picture of the KB mirror assembly is shown on Figure 3.3 as viewed at an angle of 45° below horizontal.

The sample is mounted on a goniometer capable of orienting in any desired direction so that any Bragg peak from a single crystal is accessible. The diffractometer shown on Figure 3.4 was built by Newport and has six angles which can be moved independently. The long detector arm is very strong as it was built to support a 15lb CCD camera on at 2-3 meters from the center of rotation. The length of the arm is necessary to achieve the desired high

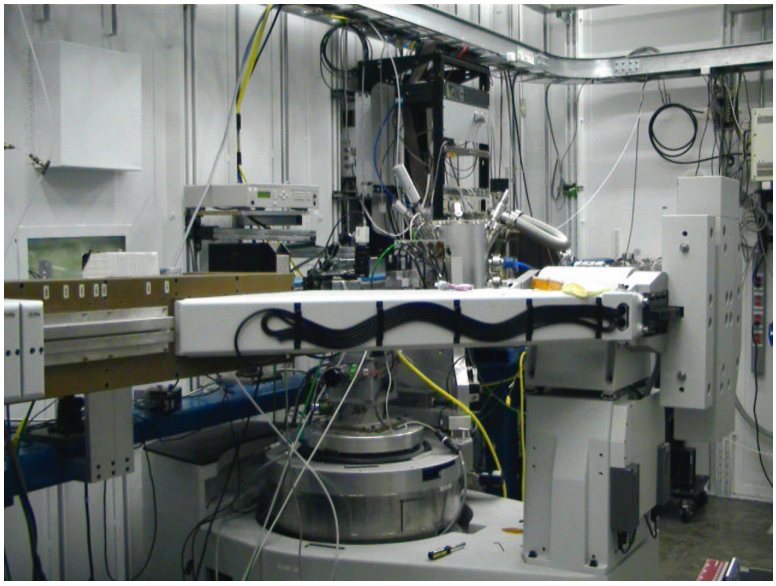


Figure 3.4: View from downstream of the Newport diffractometer. The long detector arm can be seen at the forefront and it shields the sample from view. The detector arm looks to be almost horizontal due to the large lattice spacing and thus small scattering angle of the protein crystals studied. The detector is also hidden from view.

resolution in q space. In the experiments described in the thesis there was rarely a need to move any of the motors of the diffractometer and they went mostly unused. The detector was set at a desired angle and rarely moved. The two motors controlling the position of the detector where the 2θ motion in the horizontal and the γ motion in the vertical.

The large lattice spacing of the protein crystals studied meant that the scattering angle was always small ($<3°$). This explains why the detector arm looks almost horizontal on Figure 3.4. The angles controlling the motions of the sample were θ, χ and ϕ. The axis of the θ rotation is fixed and vertical. Mounted on it is the χ motion whose axis is along the direction of the beam when θ=0. On top of that is the ϕ motion with an axis of rotation horizontal and perpendicular to the direct beam when θ=χ=0. A rotation of 90° in θ simply interchanges the orientations of χ and ϕ and so depending on how the zero of θ is defined, the two other motions may be interchanged. Only small motions of no more the 2° needed to be done with any of these three motors. In the case when the KB mirrors where mounted, they were actually mounted on a bracket attached to the θ motion in such a way that θ could not be used to control the sample position since it would have changed the alignment of the focusing mirrors. In any event, only sporadic use was made of the ϕ motion and the χ motion was hardly ever moved from its zero position.

Translation stages were mounted directly on top of the ϕ arc. These translations were called x and z with a positive increment in x being in the outboard direction, *i.e* away from the electron storage ring and a positive change in the z direction moving the sample up. These translation stages were the most important motors used in these experiments since they allowed for the sample to be moved in order to irradiate a fresh part of it, thus reducing

the effects of radiation damage.

3.1.4 Detector

In order to measure a complete CXD pattern in a manageable amount of time, we need to use a 2D detection device. The resolution of this detector or pixel size is very important since one hopes to measure very fine features of the diffraction pattern. Another key concern when selecting which detector to use is the speed at which the data can be read since many images need to be collected. In the case of protein samples, the speed is even more crucial due to unavoidable radiation damage inflicted to the sample. Combining these two characteristics of small pixel size and speed, the choice of detector was a direct read Charge Coupled Device (CCD) detector. The CCD was purchased from Roper Scientific and was a 1340×1300 pixels chip with a pixel size of 20μm. What is meant by direct read CCD is that the x-rays interact directly with the depletion layer of the chip. In comparison, many CCD's used as x-ray detectors work in the visible region of the spectrum. The x-rays interact with a phosphorescent screen which emits light which is then channeled to the CCD.

The CCD chip in the camera was kept under vacuum and maintained at a temperature of -50°C by a Peltier cooler. The heat from the Peltier was dissipated by a constant flow of cold water through the body of the camera. A beryllium window in from of the chip maintained the desired 10mTorr vacuum while still allowing the x-rays through. The cooling of the chip greatly reduces the background noise measured. The CCD was controlled via the ST133 controller from Roper Scientific which was connected to a computer and read out using the program WinView™provided by Roper Scientific. The efficiency of the detection of x-rays is found to be maximum at an x-ray energy of roughly 6keV where it is at 60% [1]. The efficiency quickly falls at energies below 5keV and above 15keV. The optimal working range is between 5 and 10keV outside of which the use of this CCD quickly becomes impractical.

The main shortcoming of this detector is its small dynamic range. Each detected photon creates a cascade of electron-hole pairs in a pixel. The signal is then digitized by the Analog to Digital Converter (ADC) and sent to the computer program. Each photon results in between 700 and 1000 AD Units or counts. Using a 16 bit ADC, the maximum number of AD Units per pixel is limit to 65,536, which corresponds to roughly 60-90 photons, for a dynamic range of only 2 orders of magnitude. This can be overcome by acquiring many accumulations and adding them up thus increasing the dynamic range to $90\times N_{\mathrm{acc}}$ where N_{acc} is the number of accumulations.

However, even though the CCD is relatively fast, it still takes 15 seconds to read out the data on the entire chip. This means that if the CCD saturates in 1 second, then most of the data collection time is spent on readout and not actual measurement. When dealing with radiation damage or unstable samples as we will discuss later, this becomes a big problem. It can be overcome slightly by switching from the slow ADC in the ST133 controller to the fast ADC also included. This cuts the readout time from 15 to 1 second but at a great cost in noise and dynamic range. The saturation level of the fast ADC was roughly 24,000 counts instead of 65,536 and many ghosting effects are introduced. However, this mode needed to be used in some cases as it was the only way to obtain the data in the limited amount of

time allowed by the lifetime of the sample as will be discussed later.

Another way to speed up the readout of the chip is to bin pixels together and reject unnecessary pixels on the edges on the chip by defining a Region Of Interest (ROI) in WinView™. Doing this effectively reduces the number of pixels that are read and speeds up the readout accordingly. This may not be very useful if the diffraction pattern is not much oversampled without the binning, that is if the reciprocal space resolution is already low without binning. In this case, binning the data will wash away the fine features of the diffraction pattern by averaging them out.

3.2 Sample and Preparation

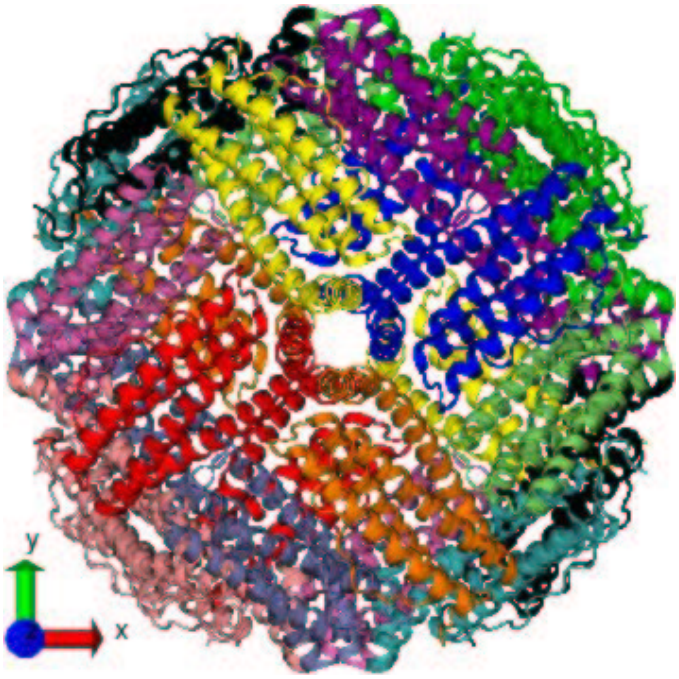
3.2.1 Ferritin

The experiments described in this thesis were all performed using a model protein called ferritin obtained from horse spleen which was chosen as a model system for the following reasons. First, it is a readily available protein and was purchased from the Sigma-Aldrich company. Second, it is known to crystallize easily by the simple addition of cadmium salt to the solution. It was found as early as 1937 that a simple addition of 10% $CdSO_4$ directly to spleen tissue from a horse immediately produces crystals of ferritin whose growth can be watched in real time [52]. Easy crystallization makes it possible to obtain the large quantity of crystals needed with minimal effort.

The horse spleen ferritin molecule is a spherical shell of inner diameter $\sim$ 80Å and outer diameter $\sim$130Å. It has a molecular weight of 474kDa. It is shown on Figure 3.5. The structure can be obtained from the Brookhaven Protein Data Bank (PDB) and was published by Granier *et. al* [25]. Ferritin molecules are found in most living organisms from mammals to plants to bacteria. Despite the fact that the amino acid sequence differs greatly for ferritins from different species (sometimes by as much as 80%) the secondary, tertiary and quaternary structures of all ferritins is conserved [24]. That is they all possess the same spherical shell structure. They are composed of 24 subunits consisting of four long helices and a fifth shorter one as seen on Figure 3.5. The 24 subunits can be all identical or can be of two different types, the heavy H-chain and light L-chain ferritin. The L and H chains can be present in varying ratios. In the case of ferritin from horse spleen, the L-chain is present 90% of the time. In some cases, when talking about ferritin, the authors refer only to one of these subunits and the whole assembled spherical shell protein is referred to as a 24-mer. In this thesis, when we refer to ferritin, the whole 24 subunits together is implied. The 24 subunits are arranged with 432 point symmetry [8], which will be discussed further in Section 3.2.2. The subunits in Figure 3.5 have been color coded according to this symmetry of the molecule. Only L-chains are shown since the PDB structure obtained was for the L-chain. The image of the structure of the protein were produced using the VMD software [33]. The axes shown in the bottom left of every image are kept consistent with each other and indicate the orientation of the molecule.

The biological function of ferritins is iron storage. The iron is stored in ferrihydrite ($5Fe_2O_3{\cdot}9H_2O$) form inside the protein shell. A full review of the function and structure

a)



b)

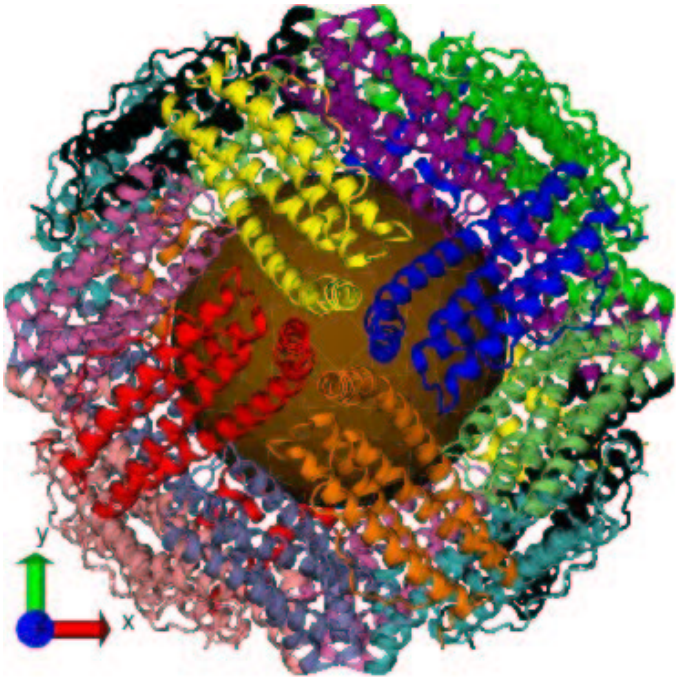


Figure 3.5: a) Atomic structure of the apoferritin molecule. The backbone of the protein is represented as ribbons. The molecule is made off 24 identical units which are displayed using different colors. Each color is used twice for 2 symmetry related subunits. b) Structure of holoferritin, *i.e.* apoferitin with an iron core. The images were generated using the VMD software [33] and the structure available on the Protein Data Bank (PDB) [25]

of the iron core is found in the article by Harrison *et al.* [28]. The iron core can be as large as 4500 iron atoms. The core possesses varying degrees of crystallinity and magnetic properties. The ferritin molecule without iron in the core is called apoferritin (Figure 3.5a)) while the iron-containing version is referred to as holoferritin (Figure 3.5b)). The presence of the iron core can greatly enhance the scattering from the protein and is a further reason why holoferritin was used for these experiments. If fully loaded with iron, the molecular weight of the protein increases by 53% from 474kDa to 726kDa. The scattering measured in the experiments described below was limited to a small region near the origin of reciprocal space. To first order, the intensity is therefore that of the origin and scales as N^2 where N is the total number of electrons in 1 molecule. To first order, with 53% more mass, there are 53% more electrons present. Taking into account that the density of proteins is roughly 1.3 times that of water, and that the relevant quantity is the density difference between the scatterer and the solution, we get that the holoferritin molecule scatters roughly 6 times more than the apoferritin molecule at small angles. This increase in intensity makes the measurements a lot easier.

3.2.2 High Quality Single Crystal

The apoferritin protein has been found to crystallize primarily in the face centered cubic structure (fcc) or the F432 space group. This is not surprising since the molecule itself possesses 432 point symmetry. Thus the crystal structure arranges itself following the symmetry axes of the protein. It has been found that holoferritin also crystallizes in the fcc structure under the same conditions. We therefore conclude that the presence of the iron core does not affect in an important manner the crystallization of ferritin. This is not surprising since the crystallization is found to be mediated by the formation of cadmium salt bridges between neighbor proteins and therefore the presence of an iron core buried deep into the protein should not affect the strong covalent interaction at the surface of the molecule. The use of the iron-filled protein to increase the signal is then justified as it does not affect significantly the crystallization kinetics. The presence of the iron core will affect the kinetics slightly by the simple fact that it makes the protein much heavier and therefore will diffuse at a different rate, but it does not affect the structure.

Protein crystals, in order to grow to a large size, must be grown at a slow rate, over a period of days. This is generally accomplished by slowly driving the system out of equilibrium toward a state of supersaturation. A solution is supersaturated when the number of protein molecules dissolved is higher than the equilibrium value. At such point the system tries to return to equilibrium by expelling proteins from solution. These proteins can then either sediment in an amorphous state or, if the conditions are right, slowly come together in an ordered fashion to form a crystal.

Producing large high quality crystals can prove difficult and is done with great care. If a highly supersaturated solution is prepared immediately, there will be a large driving force toward equilibrium and the protein molecules will leave the solution very quickly and most likely form amorphous aggregates. The art of protein crystal growth consists of starting with a non-saturated solution and slowly pushing the system toward saturation and beyond. This

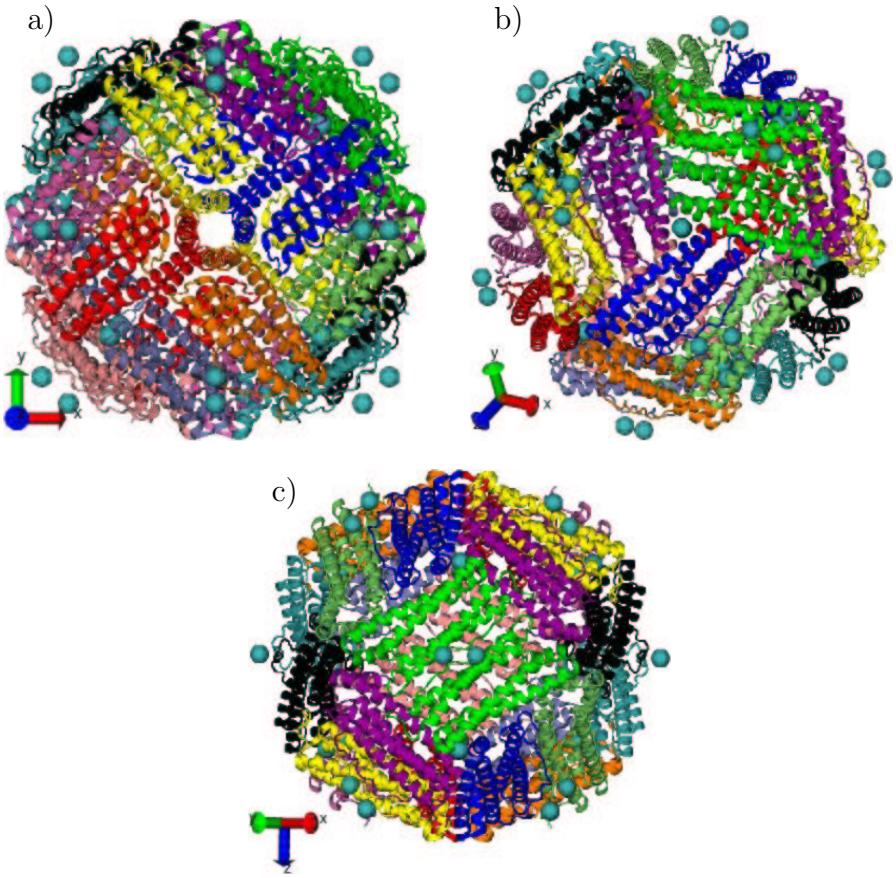


Figure 3.6: Different axes of symmetry of the apoferritin molecule and the fcc crystal of ferritin. Both belong to the 432 space group [8]. Cyan spheres are Cd^{2+} ions either adsorbed at specific locations or acting as salt bridges between adjacent molecules in the crystal structure. a) 4-fold axis corresponding to the edges of the cube in the fcc crystal. b) 3-fold axis corresponding to the body diagonal of the cube. Two cadmium ions are adsorbed at the 3-fold axis. c) 2-fold axis corresponding to the face diagonal of the cube. This is the close-packed direction and the direction of the two cadmium salt bridges.

can be accomplished in many ways as discussed by McPherson [45]. Since many parameters such as protein and salt concentration, *p*H and temperature affect the solubility of the protein, each of these parameters can be manipulated in order to try to produce crystals. The parameters usually controlled in order to drive the system toward supersaturation are the protein and salt concentration. This is done generally using vapor-diffusion. A large reservoir of a desired solvent and precipitant is prepared in a well. Then a small quantity of this solvent is mixed with a small quantity of protein solution in a drop on a glass slide. The slide is then flipped over and used as a cover on the reservoir well. Surface tension keeps the drop on the slide even while inverted and the well is sealed using vacuum grease. At this point, the whole system including the reservoir and the drop containing the protein is sealed. The reservoir has a higher concentration of the precipitant since the added protein diluted it in the drop and this means that the concentrations in the drop and the reservoir

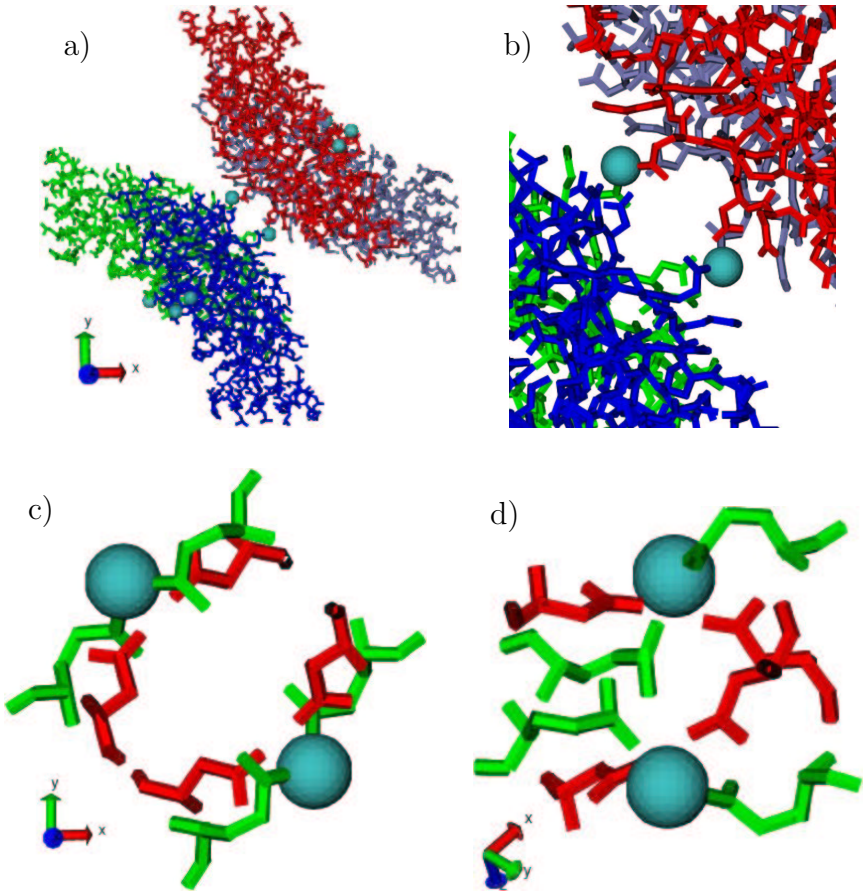


Figure 3.7: Different views of the cadmium salt bridges present at every 2-fold contact in the fcc crystal of ferritin. a) Two Cd^{2+} ions (cyan spheres) connect four subunits (differently colored) of ferritin from two adjacent proteins in the crystal. b) Closer view of the contact between the four subunits. c) The salt bridge is between two glutamines (in green) and two aspartic acids (in red), one of each from each of the two proteins. d) Different view of the salt bridge.

will slowly equilibrate through vapor diffusion. If the initial conditions were right, the final equilibrium state of the system will lead the drop to be in a supersaturated state leading to proteins leaving the solution. The slow pace of the vapor diffusion assures that this supersaturated state is reached slowly so there is no massive rapid precipitation of proteins but rather a slow orderly crystal formation.

In the case of ferritin, a standard condition leading to the formation of large high quality crystals is a reservoir of 0.2M sodium acetate at pH 5.0 and cadmium sulfate as the precipitant. However, the crystal used to obtain the structure shown here obtained from the PDB was grown using a reservoir of 1M ammonium acetate, 80mM cadmium sulfate and 3mM sodium azide used as a preservative [25]. Many different recipes will lead to good quality crystals of ferritin, almost all of them having in common the presence of Cd^{2+} ions creating salt bridges in the crystals. The locations of cadmium ions in the face centered crystal, as well as the symmetry of the crystal and the protein itself are shown on Figure 3.6.

The 432 symmetry refers to the presence of three 4-fold, four 3-fold and six 2 fold symmetry axes [8]. In a standard cartesian system, we align one of the 4-fold axes in the z direction. The two other 4-fold axes are then in the x and in the y direction. In the cubic system of the fcc crystal, this means the 4-fold axes are along the edges of the cube shown on Figure 3.6a). The 3-fold axes shown on Figure 3.6b) are found to lie in the body-diagonal of the cube. Two Cd^{2+} ions (shown as cyan spheres) are found to be bound on or near the 3-fold axes. These two ions are not salt bridges involved in the crystal formation. They are adsorbed on the surface of the molecule. It has been shown that the iron atoms enter and exit the shell at the 3-fold axis [78]. The cadmium ions in the crystal structure merely plug the holes.

The 2-fold axes of the molecules and the fcc crystal are along the face diagonal of the cube. This means that for spheres the crystal contacts must be along this axis since they are found along the face-diagonal in fcc crystals. The nearest neighbors are found in these directions, the $\langle 110 \rangle$ directions. The contacts between adjacent proteins are mediated by two cadmium salt bridges, each of them between a glutamine and an aspartic acid residue from each protein as shown of Figure 3.7. Four subunits, two from each protein are involved in the crystal contact. The different subunits are individually colored on Figure 3.7a) and b), while panels c) and d) show the glutamines in green and the aspartic acids in red. With a rough size of 130Å for the ferritin molecule, the size of the lattice parameter is $130\times\sqrt{2}$=183.8Å. The lattice spacing was found by Granier *et al.* to be a=182.91568Å, which indicates that the size of the protein is slightly smaller than 130Å. A fit to the small angle scattering profile of apoferritin was performed by Häußler *et al.* using the spherical shell approximation of equation (2.23) and they found the inner diameter to be on average 75.5Å and outer diameter 122.2Å. The diameters were found to have a polydispersity of 3.8% for the outer and 2.9% for the inner. This polydispersity is reflective of the fact that the shell is not exactly spherical and also some possible breathing of the proteins, *i.e.* volume conserving periodic deformations that are thermally activated [40].

The reciprocal lattice of the fcc lattice is itself a lattice, but it has the body centered cubic structure. The reciprocal lattice spacing is $a^* = \frac{2\pi}{a} = \frac{2\pi}{182.91568} = 0.03435\text{Å}^{-1}$. A cut through the reciprocal space is shown on Figure 3.8. The cut is along the plane spanned by the (110) and the (002) vectors. For an infinite crystal, the Bragg peaks would be delta functions. However, the reciprocal space is shown for a finite crystal of octahedral shape. The diffraction pattern for a finite coherently illuminated crystal can be calculated from equation (2.50). The scattered amplitude is shown as an isosurface for each Bragg peak. The amplitude at each Bragg peak is the Fourier transform of the shape of the crystal as described in Chapter 2. The CXD patterns shown correspond to a crystal size of 0.16μm. Comparing the spacing of the fringes of the CXD pattern with the spacing of the Bragg peaks, we can deduce the size of the crystal in the simulation. It this case, the fringe spacing is 15 times smaller than the reciprocal lattice spacing in the (111) direction. Therefore the size of the crystal is 15 times larger than the direct lattice spacing in the (111) direction, which is 106Å. A more detailed discussion of the simulations of CXD patterns is found in Section 3.11.

Other crystal forms have also been observed for apoferritin and holoferritin. They have

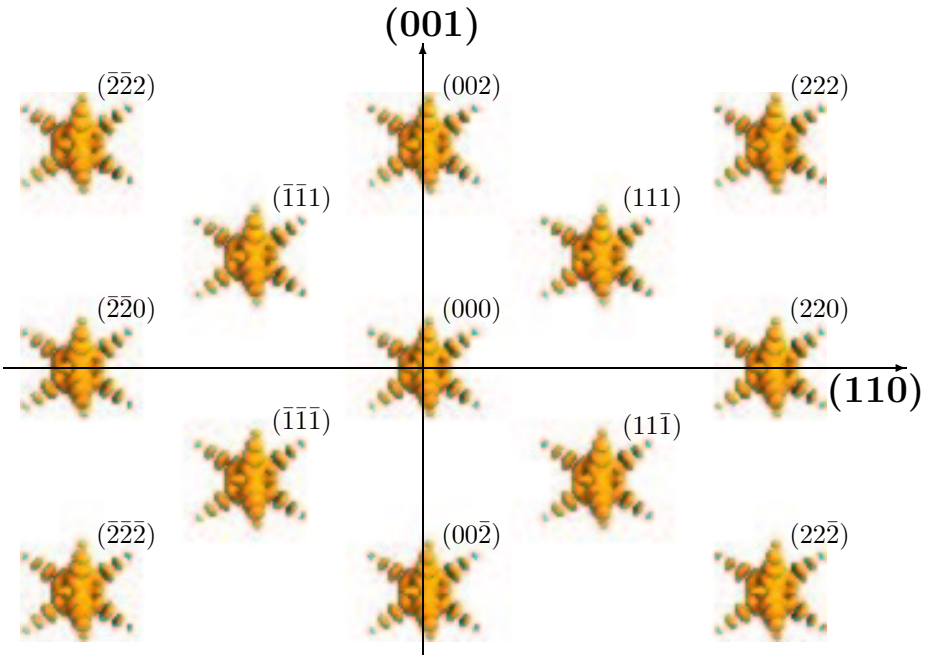


Figure 3.8: Reciprocal space of a perfect micron sized FCC crystal with octahedral habit. The reciprocal space is shown as a cut in the plane spanned by the (110) and (002) vectors. The streaks of intensity and fringes are perpendicular to the faces of the crystal and are what CXD aims to measure. The coherent pattern is identically reproduced at each Bragg peak, including the origin.

both been crystallized in the tetragonal and orthorhombic structures [24, 25]. These crystals were also obtained using cadmium as a precipitant, using slightly different concentrations. They were however found to grow much slower than the cubic crystals. While the cubic crystals grew in a large number over a scale of a few days, the other two structures were found to take over 1 month to grow and they did so in a much smaller number.

3.2.3 Protein Solutions

The protein solutions of apoferritin and holoferritin were purchased from Sigma. The apoferritin solution was provided in 100mM NaCl at a concentration of 50mg/ml. The holoferritin solutions were provided at 100mg/ml in 150mM (8.75mg/ml) NaCl. Both were sterile filtered using 0.2μm filters. The apoferritin solution has a hazy yellow appearance while the holoferritin solution appears blood red due to the presence of iron.

The characteristics of the apoferritin solution as purchased was studied by Thomas *et al.* [80]. They found it was free of foreign proteins at better than 99.9% w/w. They also found a large proportion of oligomers present in the solution (>45% w/w), most of which consisted of dimers (35% w/w) and trimers (10% w/w). They also found a small fraction (<1%) of higher molecular weight oligomers ranging from 10 molecules to 180 molecules. They were able to separate the monomers and obtain a solution containing 5% w/w oligomers. They found that the presence of dimers disrupted the crystal habit and reduced the crystal growth.

However, it did not affect the nucleation. The addition of trimers was found to increase nucleation and reduce growth. In the work presented in this thesis, the fractions of oligomers were not separated. The solutions were used as purchased either directly or they were diluted or concentrated without purification. This has little effect for our present purpose, which is to grow a large number of small crystals ($\sim 1\mu$m), discussed more thoroughly in Section 3.2.4. The presence of the oligomers reduces the quality and size of large crystals ($\sim$1mm), but does not seem to prevent the formation of smaller crystals.

In order to increase the concentration of the solutions obtained, centrifugation through Amicon Ultra 10,000 Molecular Weight Cut Off (MWCO) centrifugal filters from Millipore. MWCO is defined as the molecular weight at which 90% of the solute is retained by the filter. Everything heavier is mostly retained and lower molecular weight molecules pass through more easily. The centrifugation is necessary to speed up the filtering process. Thomas *et al.* [80] found that concentrating the separated monomer fraction using centrifugal filters lead to the formation of oligomers whose distribution is similar to that of the purchased product. Therefore, separation of monomers would prove useless if a higher final concentration was needed in the end requiring centrifugation.

The scattering obtained from solutions of apoferritin and holoferritin is shown in Section 3.4 and Chapter 4.

3.2.4 Crystal Shower

In order to be able to measure the diffraction from a single crystal, one requires the ability to orient the crystal in the desired direction so that the Bragg condition is met. This is generally easily achieved with large macroscopic crystals and can even be done for crystals that are too small to be seen with visible light. This was done for the coherent diffraction study of gold nanocrystals [70, 89]. However, it requires the crystal studied to be fixed in some way to the diffractometer so that its orientation with respect to the angles of the diffractometer can be known and fixed.

There is however no simple way to fix in space the position and orientation of a micron sized crystal of protein. It must be kept in solution at all times and its small size allows it to move around. There is no *a priori* expected orientation of the crystal in solution. If one was to produce only one micron size crystal, it would be required to search all the possible orientations in order to find the desired Bragg peak. The probability of having the crystal in the proper orientation so that a Bragg peak can be seen is given by

$$\text{Prob} = M \frac{\phi_{\text{rock}}}{2\pi} \frac{\Omega_{\text{det}}}{2\pi} \tag{3.1}$$

where ϕ_{rock} is the rocking solid angle of the Bragg peak and Ω_{det} is the angle along a ring of constant q spanned by the detector. The rocking width of the Bragg peak is measured as the Full Width at Half the Maximum intensity (FWHM) in the angle ϕ scan around a Bragg peak. The angle ϕ is shown on Figure 3.9 and corresponds to a rotation axis perpendicular to the beam. This width can be pretty narrow for large high quality crystals. For the ferritin crystals discussed below it was estimated to be between 0.1 and 0.4°, or 2 and 7 mrad. The angle ϕ can take any value between 0 and 2π. The probability that the crystal will meet

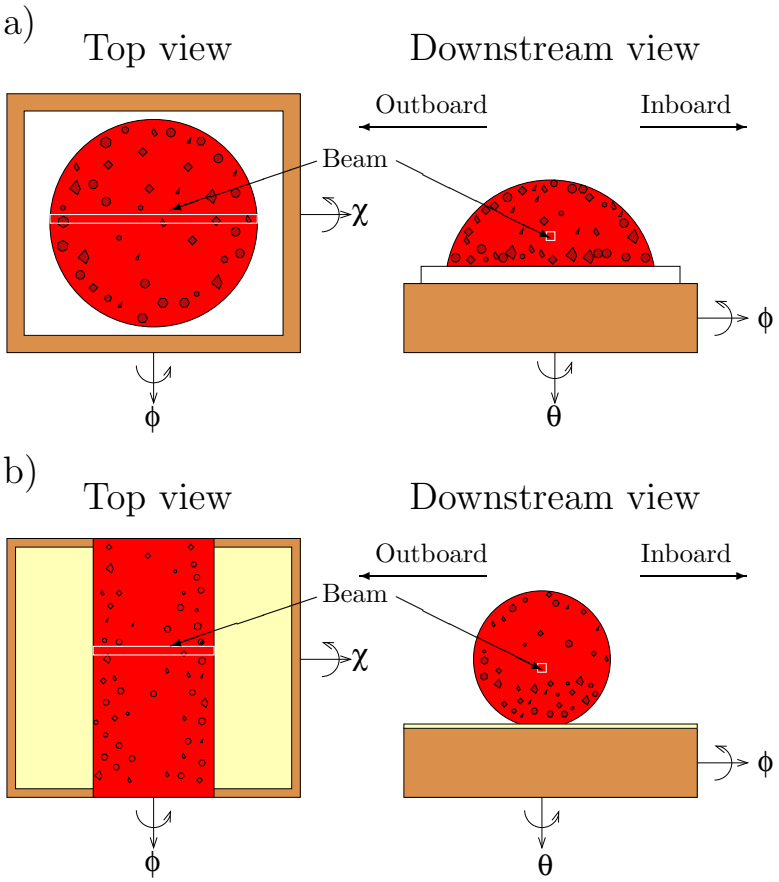


Figure 3.9: a) Drop method. The mixing of the protein solution and the precipitant is done in a drop on a temperature controlled glass slide. A large number of small crystals of varying size and shape is present. They tend to be more numerous on the surface of the drop and at the bottom. The size of the x-ray beam is shown. b) Capillary method. The mixing is done inside a quartz capillary. More crystals are seen at the bottom and at the sides of the capillary. The rotation and translation axes of the sample are shown.

the Bragg condition is therefore $\frac{M\phi_{\text{rock}}}{2\pi}$ where M is the multiplicity of the Bragg reflection equal to 8 for the $\{111\}$. The angle θ only rotates the crystal around the q vector and does not affect whether the Bragg condition is met. The angle χ determines which direction the scattered beam will take. χ can also take any value from 0 to 2π. The detector can only catch a certain angle on the ring of constant intensity defined by a rotation in χ. This angle Ω_{det} is for the CXD experiments detailed below roughly 25°. Therefore, the probability of seeing a single crystal by pure fortuitous alignment is $\sim 6\times 10^{-4}$.

To circumvent this low probability of observation, it is necessary to produce a large quantity of randomly oriented crystals in order to bring the total probability of observing at least 1 Bragg peak up to 1. It is then necessary to create a sample containing at least 1,500 crystals within the volume illuminated by the beam.

Such a shower of crystals is prepared by preparing a highly supersaturated solution of ferritin leading to rapid nucleation, hence a large number of crystals. A recipe which was

found to achieve just that was to mix an equal amount of ferritin solution as purchased and $CdCl_2$ solution ranging from 30 to 300 mM. At the upper end of that range, the number of crystals gets too large and it becomes impossible to separate each individual Bragg peak. They merge into a single continuous ring of intensity in what is known as a powder diffraction pattern shown on Figure 3.15. At the lower end of the range of $CdCl_2$ concentration, the number of crystals gets too small and the probability of observing at least one Bragg peak falls below a reasonably measurable value. There are still crystals present, they just are too sparse which makes it difficult to find one.

It has been found [24, 25], as discussed above that crystal forms of ferritin other than fcc exist. However, these occur rarely and grow very slowly. It is therefore expected the large number of small crystals produced rapidly will be in the fcc structure.

On top of the number of crystals produced depending on cadmium concentration, other aspects of the sample are tied to this variable and they are all dependent on each other. High supersaturation will create lots of crystals rapidly, which means the fast growth will lead to lower quality crystals, which will only grow to a small size. On the other hand, lower supersaturation leads to slower growth and fewer crystals of higher quality and larger size. The size, quality and number of crystals are related in a way that makes it impossible to obtain a large number of large crystals rapidly. Small crystals will always tend to be of low quality because of the way they are prepared.

The mixing of the protein and the precipitant, the cadmium salt, was performed in two main ways. It was done in air directly onto a cooled glass slide which will be referred from now on as the drop method shown on Figures 3.9a) and 3.10a). The second method, involved mixing the sample inside a quart capillary shown on Figures 3.9b) and 3.10b). The advantages of the drop method include a quick sample change and the possibility of testing different conditions with a single sample. The evaporation of the solvent over time would lead to a change in the concentration of the sample in the drop leading to an increasing supersaturation over time. The rate of evaporation could be controlled using the Peltier cooler and setting the temperature at or near the dew point in the experimental hutch on that particular day. The slow evaporation over time off the sample is also the main drawback of the drop method since it limits the total time the sample can be studied. This however was generally not an issue since radiation damage would occur in a much shorter time scale than the evaporation of the sample.

The capillary method solves the evaporation problem but does not allow the slight changes in supersaturation over time evaporation provides. Both methods were found to give identical results.

3.3 Experiment

CXD experiments on ferritin protein crystals have been performed multiple times at the APS with small variations in the experimental setup and sample preparation. All the experiments however did follow a standard methodology common to all of them. A sample was prepared by adding cadmium salt to the ferritin solution. This sample was placed on a temperature control stage. The experimental hutch was then exited as fast as possible

in order to try to minimize the time for the sample to evolve. This was done in order to try to measure the crystals at the earliest possible stage of their development. The original goal was to measure crystals at various stages of the crystallization process. This however proved difficult because of the following reasons. First, in order to produce enough crystals so that they can be seen as described in Section 3.2.4, they must grow very rapidly, on the scale of seconds. It is possible to slow this down to a few minutes and still obtain enough crystals. However, these prove to be larger than the coherence volume of the x-ray beam, making it impossible to measure their CXD pattern. This kind of samples obtained with slower crystallization could only be obtained with the x-ray beam off. It was found that irradiating the sample damaged the proteins in such a way that it would prevent the formation of crystals. Two identical samples, one placed in the beam and one outside of the hutch were prepared. The one in the beam would not yield crystals over a period of hours while the one outside the hutch would produce many crystals, which could be easily observed when this sample was mounted in the beam. This therefore made it impossible to observe different steps of crystallization with a single sample.

Many attempts were made to try to control the speed of the crystallization in order to observe many different stages. This can be achieved using many samples evolving at the same time but only irradiating them one at a time. A injection device was also built to insert microliter quantities of cadmium precipitant slowly in a capillary in order to slowly produce crystals. This device and other methods used to control the speed of crystallization were eventually not pursued very much because it proved difficult to observe CXD patterns from any crystal and we were not able to push the experiments to the point of studying the time evolution of growing crystals.

The mindset of the original experiments, measuring single crystals growing in time, quickly changed to simply finding a way to observe CXD from small ferritin crystals in any way at all.

3.3.1 Sample Mount

The sample mount used in the CXD experiments is shown schematically on Figure 3.10 for both the drop and capillary methods. The x-ray beam comes from the right side on both pictures. A water chilled copper block was machined to be mounted on top of the sample rotation and translation stages. A Peltier cooler was glued on top of the copper block using thermal paste. The water cooling of the copper block was necessary to remove the heat dumped into it by the hot side of the Peltier cooler. The cool side of the Peltier was also covered in thermal paste to get a good contact between it and the sample, either a glass slide or a capillary. A thermocouple was used to measure the temperature of the sample. The thermocouple was place on the Peltier as close to the sample as possible and the temperature it measured was assumed to be the same as that of the entire sample. There was a small time lag between the temperature readout and the equilibration of the sample at that temperature. This was determined using a sample of pure water and watching it freeze with an optical camera. The thermocouple would read temperatures below the freezing point before the actual freezing of the sample. This is not surprising due to the extended

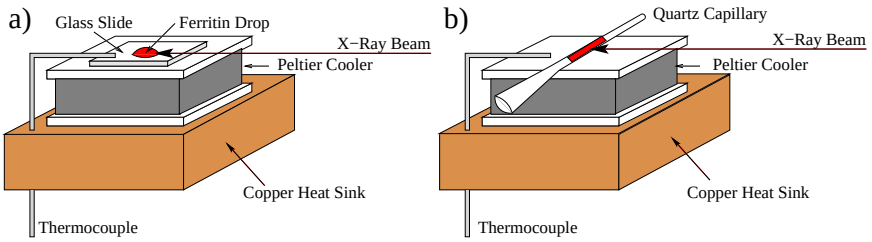


Figure 3.10: A copper block serving as heat sink was mounted directly on top of the x and z translation stages. A Peltier cooler is glued on top with thermal paste. a) Sample in a drop in air. A glass slide is glued with thermal paste on the Peltier and a drop of ferritin solution is placed on it. b) Sample in a capillary. The capillary is glued directly onto the Peltier using thermal paste.

size of the sample and the cooling on only one surface of it. Therefore proper equilibration time was given to the sample whenever the temperature was changed.

The sample was leveled so that the direct beam could go through without striking the Peltier device. The direct beam, attenuated by absorption continued on straight past the sample and the scattered beam went on in all directions, with different intensities. The CCD x-ray detector was placed at an angle away from the direct beam corresponding to the {111} Bragg peak of fcc crystals of ferritin. At the energies used, between 8 and 11 keV, this angle (2θ) was between 0.6° and 0.9°. Due to the extended size of the CCD, the setting of the 2θ was not overly critical, provided the Bragg angle was somewhere near the center of the CCD.

In the case where the sample temperature was set either very high of very low, some problems could be encountered with the sample. At high temperature, the solvent evaporation would be too rapid and the sample would be lost. At lower temperatures below the dew point of that particular day, there would be condensation of water of the sample. Below the freezing point of water, ice would start to form which could eventually be so thick that the x-ray beam would not get through it. This could be significantly mitigated by enclosing the sample in a small finite volume of air. This is shown on Figure 3.11, which shows a picture of the sample environment. A fitted Kapton skirt was placed over the sample and enclosed a small volume of air. This greatly limited condensation at low temperature and evaporation at high temperature. The presence of the two Kapton layers only slightly increased the absorption of the beam. A pipe fitting was installed on the bottom plate on which the Kapton skirt was mounted. This allowed a gas cylinder to be hooked up to replace the atmosphere with helium, which scatters a lot less than air. The PVC pipe is evacuated and is large enough that the whole range of scattering angles we are interested easily fitted inside of it. The detector arm of the diffractometer could move freely behind it.

3.3.2 Ewald Construction

A simple graphical method to better understand what is being measured in a scattering experiment was proposed by Ewald. It is known as the Ewald construction and is shown on Figure 3.12. Suppose the incidence direction of the x-ray beam is fixed in space and

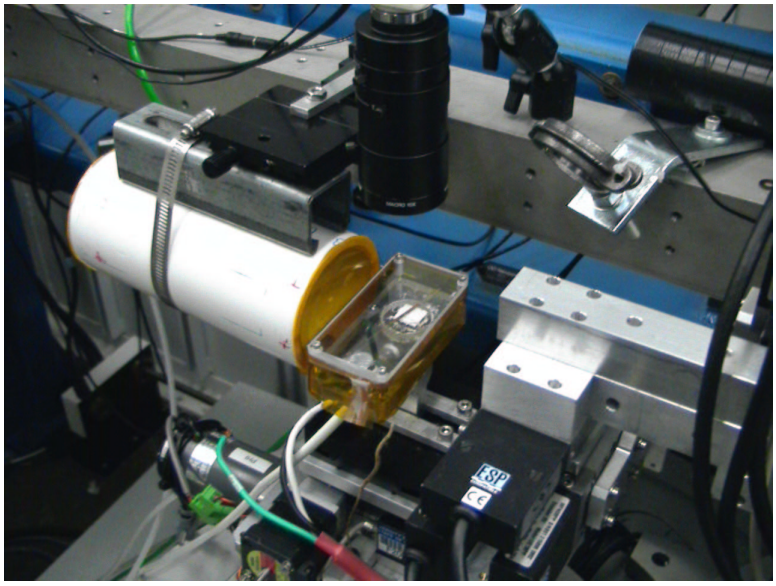


Figure 3.11: The sample mount is shown. The beam comes in from the right. The quartz capillary is glued to a Peltier cooler using thermal paste. The heat from the Peltier is dissipated through a water cooled copper block. A Kapton skirt is placed over the sample to restrict air flow and prevent the formation of ice when the sample is cooled below 0°C. A video camera used to monitor the location of the sample can be seen at the top. On the left is an evacuated flight path.

the beam is monochromatic so that the whole beam can be described by the wavevector $\mathbf{k}$ which is fixed. Then in the case of elastic scattering where the scattered beam has the same wavelength as the incident beam, the scattered beam must have a wavevector $\mathbf{k}'$ with magnitude $|\mathbf{k}'| = |\mathbf{k}|$. In the case of forward scattering, the two wavevectors have the same magnitude and direction and the $\mathbf{q}$ vector is zero. We therefore draw the end of the vector $\mathbf{k}$ at the origin of reciprocal space so that when representing $\mathbf{q}$ graphically as $\mathbf{k}' - \mathbf{k}$, the momentum transfer is zero and the $\mathbf{k}'$ vector also has its tip at the origin of reciprocal space.

With $\mathbf{k}$ fixed in space, the direction of $\mathbf{k}'$ can be changed to all possible angles and this will trace out a sphere. The values of $\mathbf{q}$ which can be measured in the experiment correspond to this sphere.

Keeping the vectors fixed in space, one can rotate the sample to any desired orientation. If the sample is a crystal with well-defined Bragg peaks as shown in Figure 3.12, then the reciprocal space of the sample is a set of delta functions. A Bragg peak can be measured when it lies on the surface of the Ewald sphere. With infinitely narrow Bragg peaks, it would be impossible to measure them since it would be impossible to have any single point lie exactly on the surface of the Ewald sphere. This is in practice circumvented by the fact that the Bragg peaks are not infinitely narrow and also the x-ray beam is not perfectly monochromatic. The spread of wavelengths in the beam transforms the Ewald construction from the surface of a sphere into a small volume contained between two spheres of different radius whose surfaces have a common point at the origin which is at the end of the $\mathbf{k}$ vector. With the beam used at the APS, the spread in wavelengths $\frac{\Delta\lambda}{\lambda}$ is on the order of 10^{-4} so

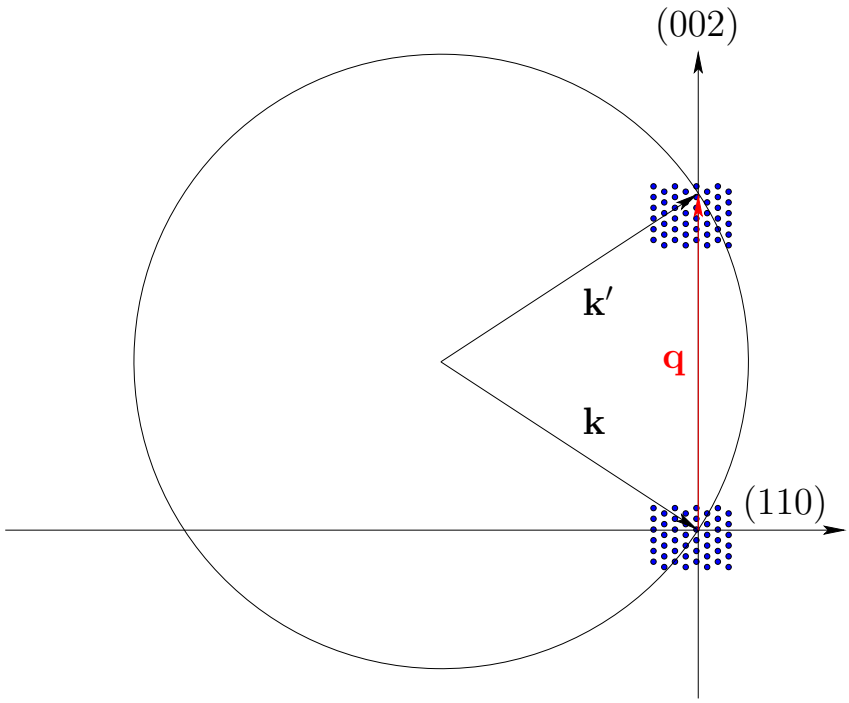


Figure 3.12: The Ewald construction for an fcc crystal drawn to scale for ferritin with 9keV x-rays. Each point represents a Bragg peak or a reciprocal lattice vector. Only parts of the reciprocal space lattice are shown for clarity. The end of the **k** vector lies on the origin of reciprocal space. The end of the $\mathbf{k}'$ vector must lie on a sphere in reciprocal space. The magnitude of both wave vector is equal for elastic scattering.

graphically, the Ewald construction still looks like a single sphere.

Setting a detector at a certain angle specifies which point on the sphere is being probed. Then the experimentalist must orient the single crystal so as to rotate the set of Bragg peaks in order to have one of them lie on the point which is being probed. For single crystals, only a discrete set of angles meet this condition. In the case of a large set of randomly oriented small crystals, also called a powder crystalline sample, the Bragg points are turned into Bragg shells by effectively rotating the reciprocal space through all possible orientation. Each Bragg point then traces out a path at a constant distance from the origin, *i.e* a shell. Then each possible Bragg reflection will have a circle of intersection with the Ewald sphere and they are all measured simultaneously.

The use of a 2-dimensional detector such as a CCD camera allows one to probe many orientations of the $\mathbf{k}'$ vector at the same time. The size of the detector and its distance from the sample determine the range of angles measured and this selects a small area on the Ewald sphere.

On Figure 3.12, the Ewald construction is shown roughly to scale for a ferritin fcc crystal. It is shown as a 2D cut through the 3D reciprocal space, so the sphere is shown as a circle. One direction on the CCD is pointing along the circumference of the circle while the second direction is perpendicular to the page. The large unit cell of the crystal compared with the wavelength of the x-ray beam means that the reciprocal lattice spacing is much smaller

than the magnitude of the **k** vector. This means that very high orders of diffraction can be measured but also means that the first Bragg peaks are very close to the origin. Therefore, when measuring near the (111) Bragg peak, which is the first Bragg peak, the direction of the Ewald circle is then very close to radial from the origin. This means that when using a 2D detector, one direction of the 2D slice through reciprocal space near a Bragg peak will always be nearly radial for low order Bragg reflections. This is not usually the case for inorganic crystals with smaller unit cells or similarly for high order reflections of a ferritin crystal where determining the direction of the cut through reciprocal space defined by the detector depends on the geometry of the experiment.

When dealing with a finite small crystal coherently illuminated, each Bragg peak is an extended function in reciprocal space determined by the crystal size and shape. A 2D cut through this pattern is measured and the direction of the cut needs to be understood. As mentioned, for low order peaks of large unit cell crystals, one of the two directions of the 2D cut is known. The second direction depends on the orientation of the crystal around the **q** vector. As mentioned before, a rotation around the reciprocal lattice vector **q** does not move the crystal out of the Bragg condition and the Bragg peak is still measured. However, since the whole reciprocal space rotates with the crystals, the off-axis volume of space adjacent to the Bragg peak rotates with it. This orientation is harder to determine experimentally since it is free to take on any value.

Figure 3.13 shows schematically these cuts through a model CXD pattern for an octahedral fcc crystal with (111) faces. The locations of the Bragg peaks with respect to the crystal are fixed. The (002) direction corresponds to the direction from the center of the crystal to the top apex if the crystal. The CXD pattern near the (002) Bragg peak is shown with intensity flares perpendicular to the crystal facets. A rotation of the crystal around the (002) vector still allows the Bragg peak to be measured and the whole CXD pattern is rotated along with the crystal. Therefore, the direction of the cut horizontally is determined by this rotation around the (002) vector while the vertical direction of the cut is nearly radial, *i.e.* in the (002) direction as discussed above due to the small angle scattering geometry.

A series of 2D cuts as shown on Figure 3.13 can be measured and put together to obtain a 3D CXD pattern. It is more efficient to measure multiple orientations by rocking the crystal slightly around a vector perpendicular to both the **q** vector and the **k** vector. This rotates the sample out of the Bragg condition and allows to measure slices away from the center of the diffraction pattern. All the slices are roughly in the same direction although small geometrical corrections need to be applied to each slice individually [60]. A small rotation of the crystal of less than 2° allows for the whole CXD pattern to be sampled. On the right side of Figure 3.13, the diffracted amplitude measured by the CCD for the corresponding slice on the left is shown. The 3D pattern is centrosymmetric but only the the center slice is symmetric in 2D, in general.

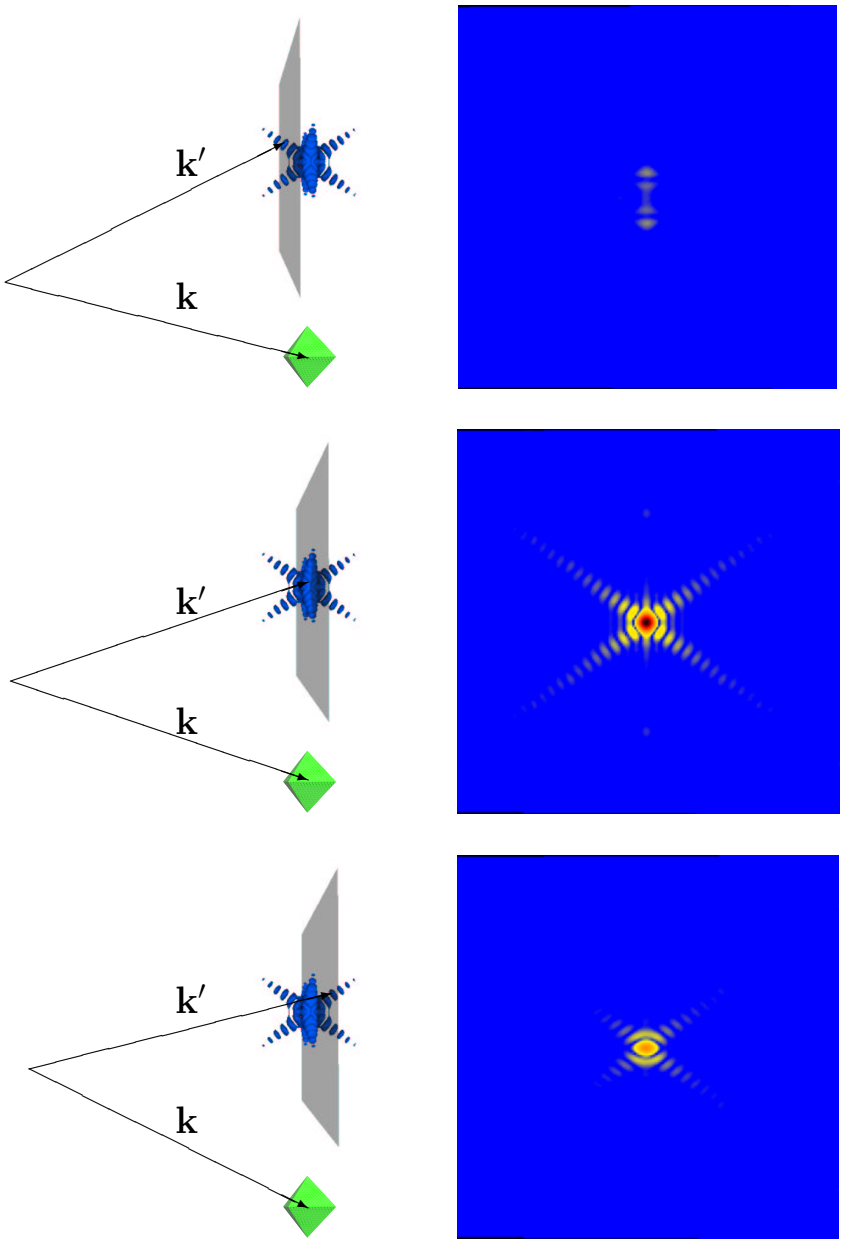


Figure 3.13: The orientation of the CXD pattern from an octahedral fcc crystal is fixed to the orientation of the crystal. Rotation of the crystal is equivalent to a rotation of the x-ray beam. A 2D detector probes slices through the CXD pattern and different slices are measured for different sample orientations. The scattered amplitude for the corresponding 2D slice is shown on the right side.

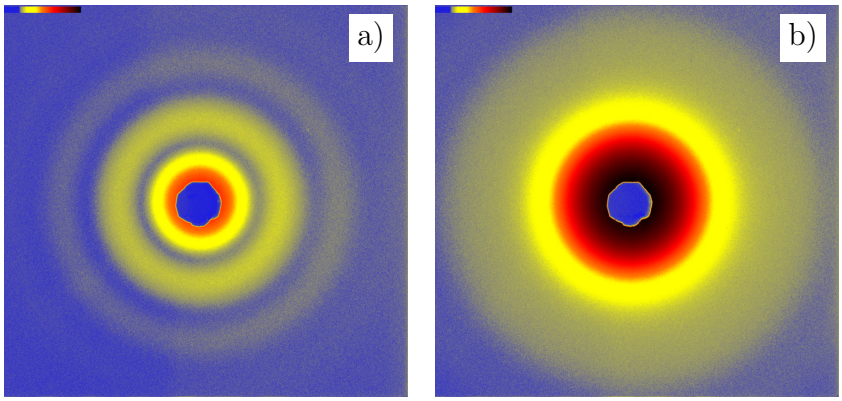


Figure 3.14: a) Small angle scattering pattern (shown on a log scale) of a solution of apoferritin at 50mg/ml. The central part of the detector is blocked by a beam stop. The oscillations in intensity expected for a spherical shell are seen. b) SAXS pattern (log scale) of a solution of holoferritin at 100mg/ml. The oscillations observed with apoferritin are washed out by the presence of an iron core which varies in size and shape from molecule to molecule. The intensity of holoferritin is higher than that of apoferritin.

3.4 Results

The small angle scattering pattern from both apoferritin at 50mg/ml and holoferritin at 100mg/ml are shown on Figure 3.14 a) and b) respectively. These patterns were collected with the detector located at 322mm from the sample and at an energy of 10keV. The difference between the empty and iron-loaded protein are clearly highlighted. The oscillations in the intensity for apoferritin are clearly indicative of the spherical structure of the protein. On the other hand, the oscillations are washed away by the presence of the iron core. This is because unlike the protein shell, the iron core does not have a well-defined spherical shape. The two images are on the same intensity scale and the stronger scattering from holoferritin is clearly highlighted. The dark spot in the middle of the images is the shadow cast by the beam stop placed there to absorb the direct beam and protect the CCD camera. Except for the low q part of the image where correlations lead to a significant contribution to the structure factor of the solution, the images correspond roughly to the square of the form factor of the protein.

When 100mM $CdCl_2$ is added to both solutions, a large quantity of small crystals is produced as shown on Figure 3.15 a) and b). The crystals are randomly oriented and so there is a Bragg peak in every direction and a set of continuous rings typical of a powder diffraction pattern is clearly visible. Indexing of the q values of these rings reveals that they correspond to the expected locations of the fcc lattice of ferritin with lattice spacing $a \approx$183Å. A few differences are seen between the two patterns. The first one is the simple difference in overall intensity from the fact that holoferritin scatters significantly more than apoferritin. But also, the different molecular form factors for the two proteins lead to different relative intensities of the Bragg peaks. Some of the peaks lie at or near a minimum of the form factor of apoferritin and this causes the intensity of these peaks to be lower relative to peaks near a maximum in the form factor.

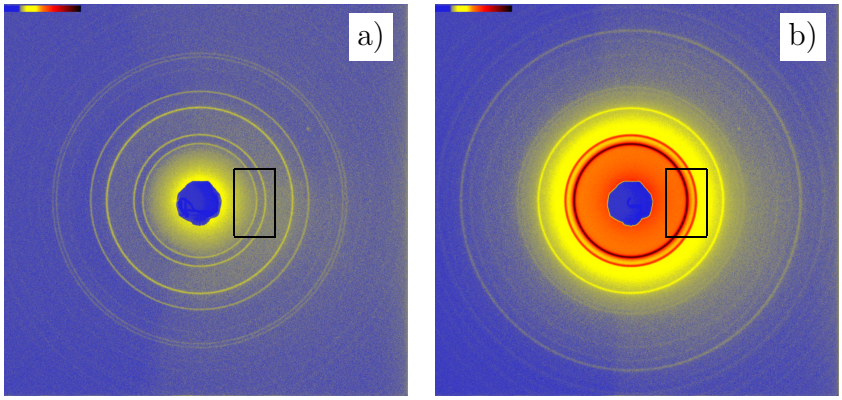


Figure 3.15: a) Powder diffraction (shown on a log scale) of apoferritin with added cadmium salt. b) Powder diffraction (log scale) of holoferritin with the same amount of cadmium salt. The intensity is higher for holoferritin due to the presence of the iron core. The box indicates the area of the image probed when zooming in for CXD measurements.

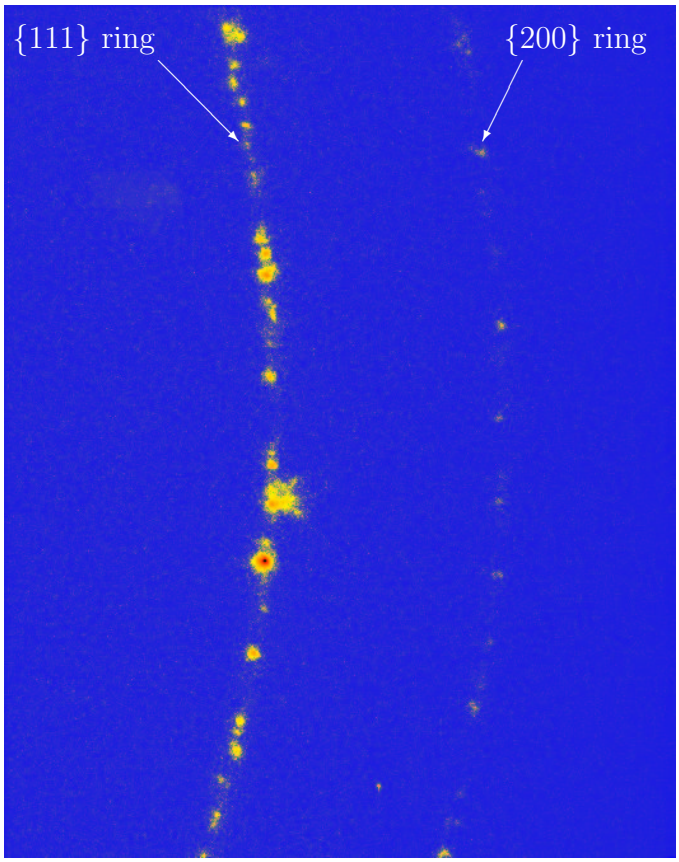


Figure 3.16: Individual Bragg peaks from a sample of many small holoferritin crystals. Each peak comes from the diffraction of a single crystal. Peaks lie on the {111} and {200} Bragg rings.

The intensity on Figures 3.14 and 3.15 are on the same scale. Comparing the two, an important feature of aggregation can be seen. The total diffracted intensity integrated over all reciprocal space is fixed by the number of scatterers illuminated by the beam. If proteins aggregate and form crystals, Bragg peaks appear leading to high scattered intensity at localized spots. But since the total intensity must be conserved, this means the intensity has to go down somewhere else, *i.e* at low q. The intensity near the origin goes down when crystals are formed as can clearly be seen when comparing Figures 3.14 and 3.15.

If the number of crystals in the sample is reduced or alternatively if the x-ray beam is made smaller so that fewer crystals are illuminated, the powder diffraction pattern starts to break up from a continuous ring into a set of distinct Bragg peaks coming from individual crystals. In order to measure the coherent diffraction pattern near one such Bragg peak, it is needed to zoom in on it to get the desired q resolution. This is done by moving the detector back from 800mm to 2000mm. The zoom in area is outlined on Figure 3.15 and shown on Figure 3.16, which was taken during a different experimental run and with a different sample of holoferritin. Only a small section of two Bragg rings is now visible. These two rings are the first two fcc rings, the {111} and {200}. Individual Bragg peaks are clearly identified.

With the setup used to obtain Figures 3.14 and 3.15, which we will refer as the SAXS setup described more thoroughly in Chapter 4, the detector is moved very close to the sample in order to be able to measure at higher q. This has the effect of reducing the q resolution and moving the rings close together. In this setup, the width of the rings is smaller or comparable to the size of the beam and therefore the real width of the powder rings cannot be accurately measured. Increasing the q resolution can however be done by moving the detector back.

This was done for a few $CdCl_2$ concentrations added to a solution of holoferritin. A volume of 6μl of $CdCl_2$ salt was added to 6μl of holoferritin at 100mg/ml. The detector was set 1185mm behind the sample. Only parts of the entire powder rings were collected, the detector spanning roughly 90° or 1/4 of the pattern. The patterns obtained were circularly integrated using the Fit2D program developed at the ESRF [27]. Figure 3.17 shows the integrated 2D patterns for 5 different $CdCl_2$ concentrations near the range yielding micron size crystals.

The expected behavior discussed in Section 3.2.4 is seen with the number of crystals increasing with cadmium concentration. This appears in Figure 3.17 as a larger intensity at the Bragg peaks. However when the cadmium concentration gets too large, there is no longer any crystal formation but rather an amorphous precipitation. There is indication of the beginning of this process on Figure 3.17 where the intensity measured is lower for 98mM than for 80mM. The number of crystals could be counted directly for 3 of the concentrations while for the other two, the powder rings were continuous and individual Bragg peaks could not be resolved and therefore could not be counted directly. The number of individual Bragg peaks or crystals seen in the samples is shown on Figure 3.18.

The size of the crystals in the sample can be obtained from the width of the peaks using the Scherrer formula given by

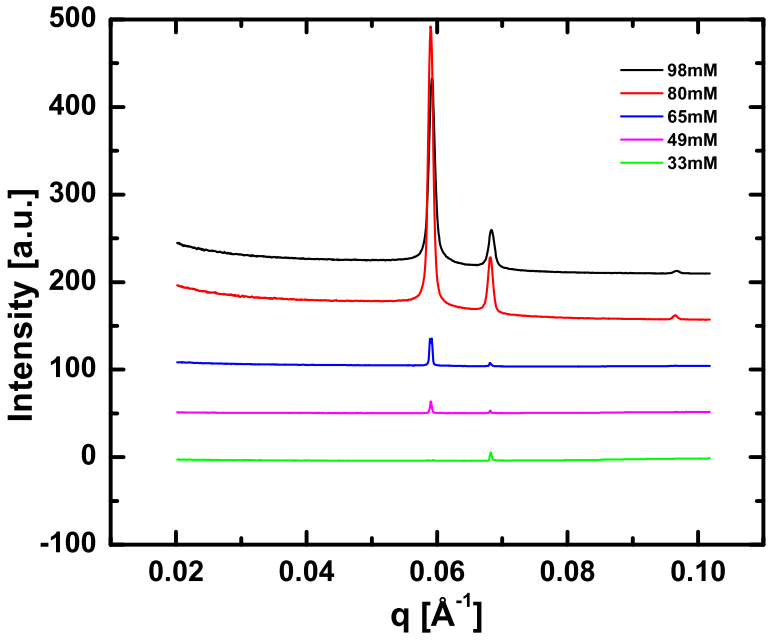


Figure 3.17: Circularly integrated powder diffraction patterns for 5 different $CdCl_2$ concentrations added to 100mg/ml holoferritin solution. The width of the Bragg peaks increases with cadmium concentration and the intensity goes up due to the larger number of crystals present.

$$\text{Size} = F\frac{2\pi}{\text{FWHM}} \tag{3.2}$$

where FWHM is the full width at half the maximum of the Bragg peak and F is a prefactor which depends on the model used to derive the expression. The value of F is 0.94 if a Gaussian is used to model the central diffraction peak while a more naive derivation yields F=1. Using the {111} Bragg peak, we obtain crystal sizes ranging from 0.5μm to 4.1μm, seemingly ideally matched to the coherence length of the x-ray beam. The crystal size versus cadmium concentration is shown on Figure 3.19.

The size of the crystals gets smaller with cadmium concentration, while at the same time the number of crystals increases, indicating that these two variables are linked. To obtain more crystals, one needs a high nucleation rate which leads to rapid growth which limits the size of the crystals.

In the next few sections, we analyze the time evolution of individual crystals by tracking over time the individual Bragg peaks as those shown on Figure 3.16. We analyze patterns as these for different conditions as well as the effect of radiation. We then analyze the shape of the CXD patterns measured for single Bragg peaks.

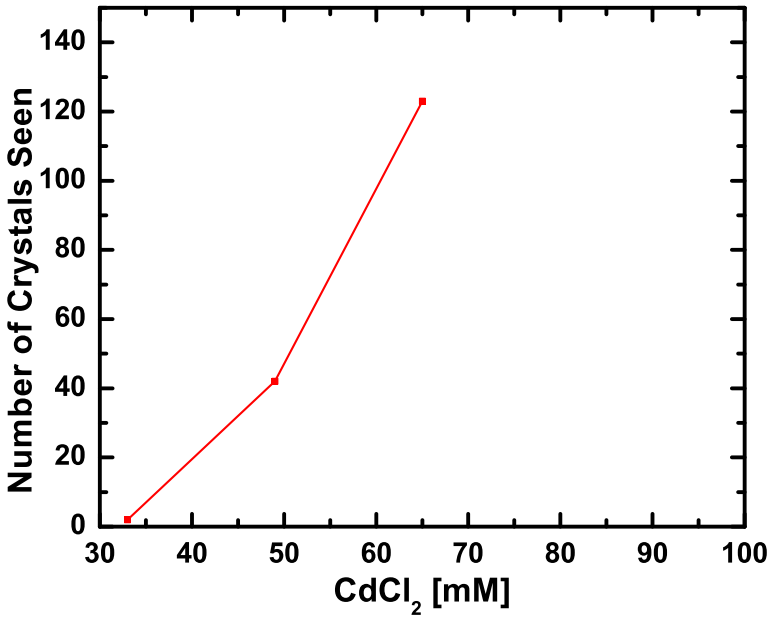


Figure 3.18: Number of individual Bragg peaks resolved with a 25×25μm beam on a 12μl drop sample versus $CdCl_2$ concentration. The number of crystals produced increases with cadmium concentration.

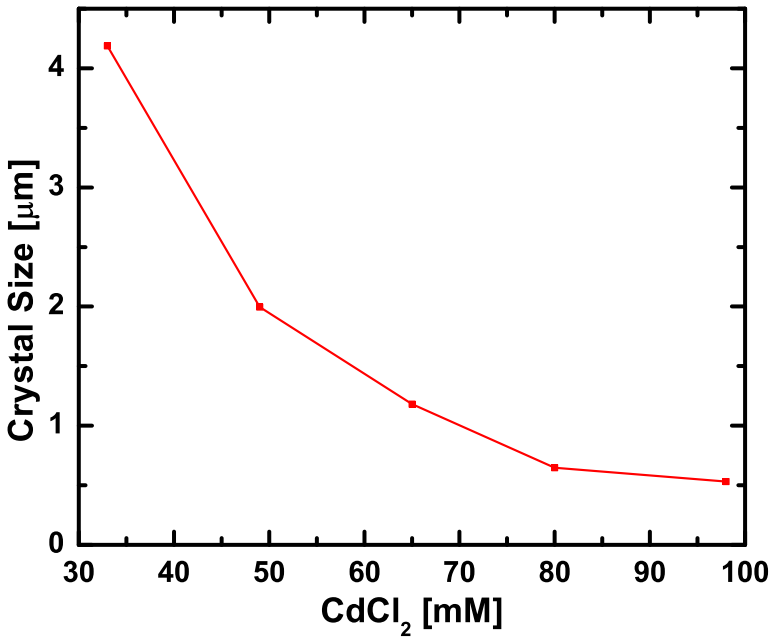


Figure 3.19: Average crystal size produced with a 6μl drop of holoferritin to which 6μl of $CdCl_2$ salt at different concentrations is added. The crystals are smaller for higher cadmium concentrations.

3.5 Thermal Fluctuations

With the small dynamic range of the CCD detector used in these experiments, it is not possible to measure in a single image a complete CXD pattern from a small crystal. This is because the central part of the diffraction pattern, *i.e* the Bragg peak is much more intense than the fine features around it arising from the shape of the crystal. With a dynamic range of roughly 60, only the central part can be resolved. It is therefore needed to accumulate many exposures and add them together in order to measure the complete diffraction pattern. The procedure relies on a simple fact to hold true : the sample does not change over time such that successive diffraction images are of the same crystal, in the same orientation. However, it was found that this is not the case. For samples at room temperature, changes over time were observed during the measurement of time series These fluctuations were observed in every single sample. A typical time series is seen in Appendix A.

There were five types of changes over time observed in the many time series that were done. The first one was a fluctuation in the intensity of each pixel by a small random amount. This is obviously associated with counting statistics error with a typical error of order $\sqrt{N}$ for N counts.

The second change over time observed was a fluctuation in the intensity of a certain Bragg peak from image to image by an amount larger than the typical counting statistics fluctuations. This effect can be associated with random Brownian motion. A crystal will diffract only if it is oriented so that it meets the Bragg condition. There is only a small range of angle around that orientation that will lead to diffraction. If the crystal orientation changes slightly in time due to rotational Brownian motion, successive images can be taken with the crystal at the peak of the rocking curve and far enough away from the peak that there is a significant difference in intensity. Such rapid oscillations in the crystal orientation leads to a Bragg peak flashing in and out on successive diffraction images.

The third change over time is also due to the free rotation of the small crystals in the sample. In some cases, Bragg peaks are seen to appear, then slowly get more intense, until they reach a peak intensity, a which point they slowly decrease and disappear in roughly the same way as they previously appeared. This is exactly what a rocking scan does. That is, as you rotate the crystal around an axis perpendicular to the scattering plane, the crystal is rocked through the Bragg condition at a constant speed as described in Figure 3.13. Typical rocking curves for a few of the small ferritin crystals are shown on Figure 3.20.

In some cases, the crystals happened to be rotating freely in the solution exactly around the proper axis leading to an accidental rocking curve measurement. The crystal rotates into the Bragg condition then out of it at a roughly constant but fairly slow speed. Intensity versus time for all the peaks found for a single time series are shown on Figure 3.21. Highlighted in blue is a freely rotating crystal that moves in and out of the Bragg condition.

A fourth fluctuation of the data over time was very similar to the one just discussed. In some case, a Bragg peak was seen to rotate around the powder ring at a roughly constant speed. A single Bragg peak could be tracked over time as it moves around the ring. This also corresponds to a free rotation of the crystal in solution, but this time the rotation is around the incident beam direction. Such a rotation does not change the fact that the crystal meets

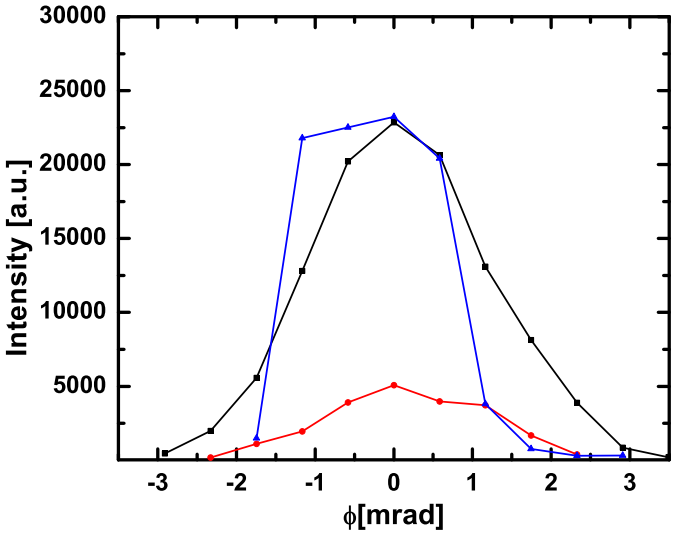


Figure 3.20: Rocking curves of three individual crystals of holoferritin from a single sample. The FWHM of the rocking curve is 3mrad. The intensity was obtained by integrating the diffraction spot on the CCD detector.

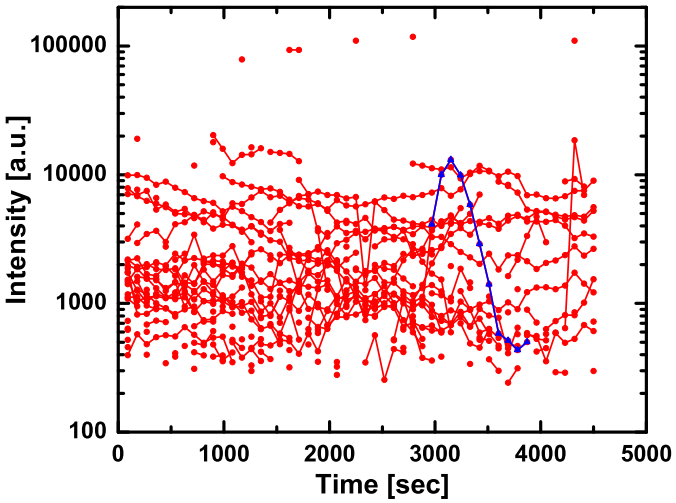


Figure 3.21: Intensity of single Bragg peaks measured versus time. Lines joining points indicate the same crystal diffracts over multiple frames with the intensity changing over time. The curve highlighted in blue is a good example of a crystal freely rotating in and out of the Bragg condition.

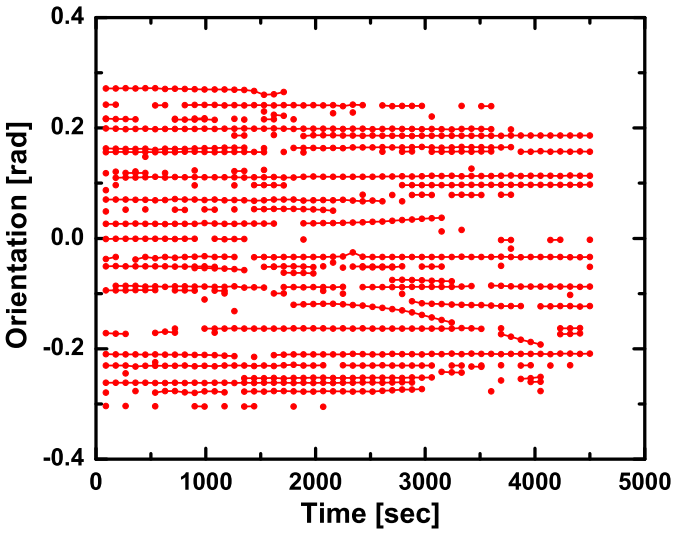


Figure 3.22: Angular trajectories of individual Bragg peaks. Most peaks are seen to be fairly stable over time, with some appearing and disappearing. A few are seen to move significantly over time, indicating a rotation of the crystal around the direction of the incoming beam.

the Bragg condition. It simply rotates the scattering plane. Using a 2D detector allows us to keep track of this changing scattering plane. The angular position on the detector of each Bragg peak was extracted and trajectories are plotted on Figure 3.22.

Since it can be seen that the crystals are free to rotate around two distinct perpendicular axes, it stands to reason that then can also rotate around a third axis perpendicular to both of these axes. This is not directly observed because this third axis is the direction of the reciprocal lattice vector and a rotation around this axis does not change the central part of the Bragg peak in any significant way since it keeps the crystal diffracting. Therefore, rotations around this axis are not directly observed but must be present. In fact, with no preferred directions in the sample, rotations should occur around any possible axis.

Finally, the last type of change seen over time is a destruction of the the Bragg peak in a way best described as an explosion of the Bragg peak. A Bragg peak which is seen to be stable over a certain period of time eventually changes shape and size gradually until it disappears completely. We associate this with radiation damage and discuss it further in Section 3.6.

For the moment, let us focus our attention on the motions of the crystals within the sample, or the thermal fluctuations. The level of motions in the sample was characterized by performing a statistical analysis on the angular velocity of the peaks moving around the Bragg ring. Each image of a data series was run through a peak search algorithm which identified the position in pixels of each *good* peak. A peak made of a certain number of pixels was deemed good if that number of pixels was within a set lower limit (5 pixels) and an upper limit (51 pixels). A background box was determined around the peak and if the total integrated intensity per pixel in the peak was smaller than 5 times the background level, it was rejected. Then finally, every saturated peak, that is the peaks with 1 or more

pixels reaching the saturation value of the detector, was also rejected.

Having found all these peak positions, the list of positions in pixels was output to a second program called **fit_multiple_2D_gaussians** which was written to go to these locations, extract a 41×41 pixel box around that location and finally perform circular averaging of the data around the center point. This produced a curve of intensity versus Q which was then fitted to a Gaussian. We recall here that $\mathbf{Q}$ was defined as $\mathbf{Q} = \mathbf{q} - \mathbf{G}_n$, with $\mathbf{G}_n$ being the reciprocal lattice vector. Knowing the position of the direct beam on the original image, the q value of the peak over time could be extracted, as well as the intensity, the width of the peak and its angular position around the Bragg ring. The width of the peak could be used to get a rough size of the crystal and track its changes over time. The circular averaging makes the implicit assumption that the width of the peak is the same in both directions. In cases where that was not the case, the fitted value of the width is an average of the two widths.

Another short computer program called **merge_powder** was written to identify peaks that are present on consecutive images and reorder the list in the data file so that peaks are no longer listed in order of the image number on which they are found in the time series, but rather list them as individual trajectories for each crystal seen over many images. The program identified peaks in an image that were within a certain distance from peaks in the previous image and then checked that they were not assigned to multiple peaks from the previous image. The trajectories shown in Figure 3.22 were obtained in this way. The programs written allowed for many thousands of images to be processed in a single week, during which upward of 300,000 Gaussians were fitted to individual peaks and the multi-dimensional trajectories (intensity, width, q, angular position, angular velocity versus time) were obtained for a few thousand crystals.

Using the angular velocity thus obtained, a distribution was plotted and is shown on Figure 3.23. At room temperature, where most of the experiments were performed, the distribution was found to have a FWHM of roughly 0.2mrad/sec. Even though this is the angular velocity only measured around the axis parallel to the incident beam direction, there are no reason to expect that rotations around other axes would be significantly different. Therefore, a crystal has a non negligible probability to be rotating at a speed of 0.2mrad/sec around the direction perpendicular to the scattering plane. That is the crystal can be rocking in or out of the Bragg condition at such a velocity and there is therefore a non negligible probability that after 60 seconds, the crystal will have rotated by an amount close to the width of the rocking curve, which is ∼3mrad as seen on Figure 3.20. This means there is a decent chance a Bragg peak seen at a certain instant will be gone 1 minute later. There is also a good chance that some of the crystals will be seen for very long periods of time.

3.5.1 Freezing

In order to try to minimize the effect of radiation, and also believing the motions of the crystals could be slowed down significantly, it was attempted to freeze the sample to liquid nitrogen temperature using what is called a "cryostream" cooler obtained from Oxford Instruments. A cryostream generates a stream of nitrogen gas at a temperature just above

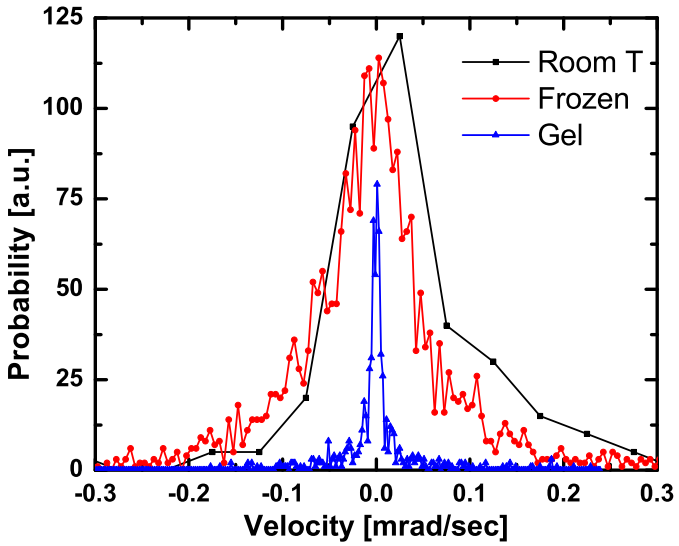


Figure 3.23: The distribution of angular velocity around the direct beam is peak at zero. Crystals are seen to move significantly at room temperature. Freezing to liquid nitrogen has surprisingly little effect while trapping the small crystals in an agarose gel greatly reduces the observed rotations. The distributions have non exponential tails and are better fitted by a Lorentzian function than a Gaussian or Maxwellian distribution.

the liquid nitrogen boiling point which is flowed from a nozzle onto the sample. The sample immediately freezes while the constant flow of nitrogen on it prevents the formation of ice over the sample.

Though it was found that a sample frozen in such a way significantly more resistant to radiation, it was shocking to find out that the small crystals embedded in the frozen solution still moved around significantly. This is shown on Figure 3.23. When the distribution of angular velocities in the cryocooled sample is compared with the room temperature distribution, very little difference is seen. The width of the distribution is roughly the same. This was certainly a surprising result since one would expect that the frozen solution would trap the crystals and they would be unable to move very much. A possible explanation as to why this happens might be that there is significant heating of the sample by the intense beam and the ice where the beam hits gets melted by the beam and therefore the frozen and room temperature samples are the same as seen by the x-ray beam. No visible indication of melting by the beam was seen, but with a very small beam, some local melting could occur on the 25 micron scale and not have any visible effects, especially with the constant cooling from the cryostream. It is however doubtful that the heat transfer from the beam causes the temperature to increase from 80K all the way up to 273K. Computational modeling of heating of cryogenically cooled crystals have indicated that melting due to the beam should not occur [47].

Figure 3.24 shows an image of the two typical Bragg rings with individual crystals when the sample was frozen to liquid nitrogen temperature. The individual peaks are present as in the room temperature case, but a brand new feature can be seen. A broad ring, as

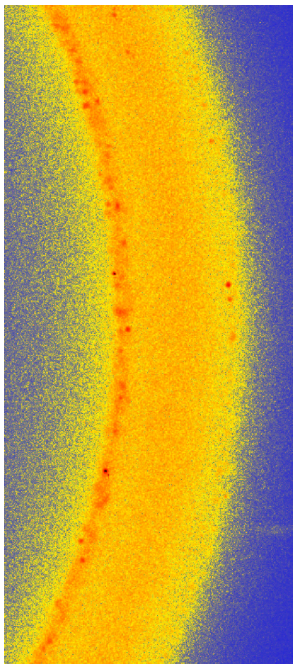


Figure 3.24: Powder diffraction pattern from holoferritin cryocooled to liquid nitrogen temperature. Individual Bragg peaks are seen as in the room temperature case but a new broad ring of intensity is now present between the {111} and {200} rings.

wide as the spacing between the two rings appears when the sample is frozen. Chapter 4 is dedicated to the study of this broad ring.

Since freezing the sample did not reduce much the motions of the crystals in the sample, it was not much help in improving the measurements of coherent diffraction patterns from single crystals.

One last comment about freezing a sample with many small crystals present. When rapidly freezing to cryogenic temperatures, the small crystals are preserved, but when freezing only slightly below the freezing point, the small crystals completely disappear. This is somewhat expected since freezing is known to damage protein crystals by the formation of ice crystals. If the cooling is rapid enough to low temperature, the formation of ice crystals is impeded and a more amorphous ice formation occurs, which does not damage the crystals present before freezing.

3.5.2 Gels

In a further effort to try to damp out the motions observed in the sample, after freezing proved unsuccessful, it was attempted to trap the small crystals into a highly viscous medium. The crystals were therefore grown in an agarose gel. A 1% agarose gel was prepared by mixing 1g of agarose dried powder with 99ml of water. The mixture was heated to the boiling point until the agarose powder fully dissolved and the solution appeared clear. It was then allowed to cool and just before it solidified into a gel, a solution of ferritin was

added to the agarose gel. The protein solution was then included in the gel inside a quartz capillary at room temperature.

A solution of 60mM $CdCl_2$ was added on top of the gel in the open end of the capillary. The cadmium salt slowly diffused into the gel and crystals formed over a period of a few hours. The concentration of cadmium varied with position along the capillary due to the slow diffusion of the salt into the gel. With a higher concentration near the injection spot, a larger number of smaller crystals is seen while as the distance from the injection end increases, the number of crystals decreases and they become larger. A wide range of crystal sizes from 1 micron to roughly 0.5mm in size was observed. Since we are interested in the smaller crystals, a location on the sample with a good number of crystals of size near 1μm was found and attempts to measure CXD patterns were made.

It was found, again by tracking the motions of individual Bragg peaks that the crystals were stabilized significantly by the gel. The distribution of angular velocities measured is shown on Figure 3.23. The width of the distribution is seen to be much smaller than in the room temperature and frozen cases. This allowed measurements to be performed for a much longer period of time on individual Bragg peaks. The gel however did not provide any relief from the radiation damage.

3.6 Radiation Damage

In almost every x-ray experiment on biological samples, especially at synchrotron sources, the damage caused to the sample by the intense beam becomes a significant issue. The study of the effects of radiation on protein crystals has become recently a research field of its own. A short review of recent development was published by Nave and Garman [58]. The effect of radiation damage in protein crystallography is generally observed as a loss of resolution, that is a loss of the higher order Bragg peaks which would be necessary to obtain a structure at high resolution [79]. As radiation slowly damages the molecules by breaking bonds and destroying crystal contacts, the higher order Bragg peaks slowly disappear because this effectively increases disorder which leads to a Debye-Waller factor reduction of the intensity at high q. This is so-called "spot fading". Most current studies are performed on cryocooled crystals since this is more relevant to protein crystallographers. Evidence of specific damage has been shown by many authors [87, 7]. Some amino acids as well as some quaternary structures are more sensitive to radiation and get damaged more rapidly than others, as can be visualized in the atomic structure.

The effect of radiation on a single small crystal of holoferritin is shown on Figure 3.25. Each image corresponds to an exposure time of 95 seconds and the data shown is from a certain time after the beginning of the exposure to x-rays. For the time prior to that shown, the Bragg peak was stable and seen for a few minutes. However, after a period of roughly 15 minutes, the Bragg peak underwent a significant transformation as shown on the images of Figure 3.25. The peak starts to break up into 2 or more peaks. As time goes by, these peaks are seen to move to the right side of the image, which in this particular case corresponds to a higher value of q. The origin in reciprocal space was located on the left side of the image and the movement of the intensity toward higher q was radial.

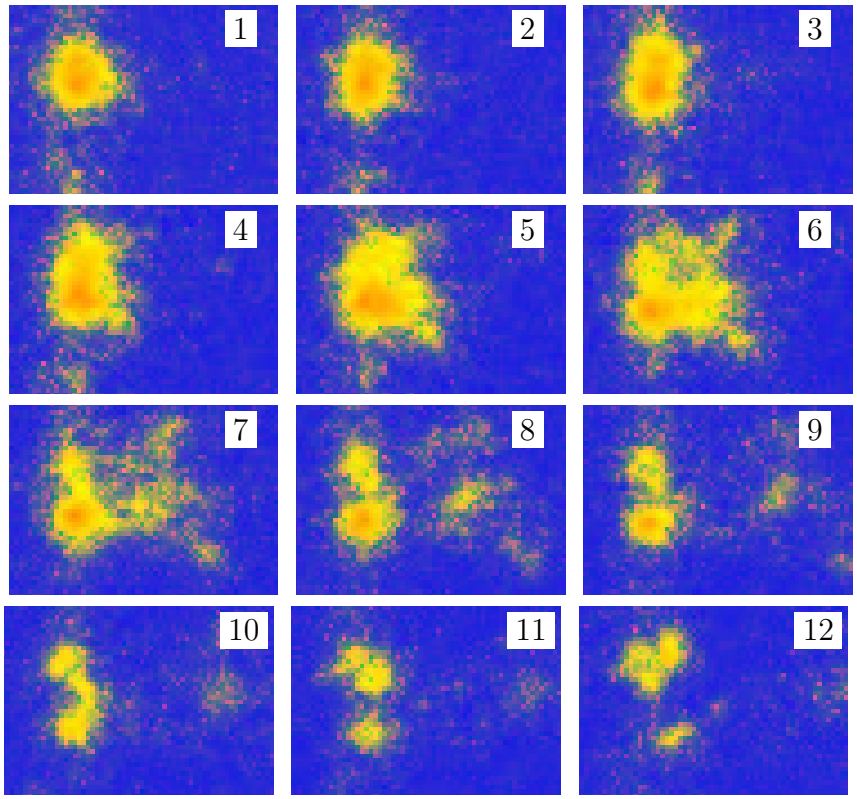


Figure 3.25: Images of the logarithm of the intensity of a single Bragg peak over time. Each image is taken 95 seconds apart. After a period where the peak was stable, it breaks up into multiple peaks which move radially to higher q with time.

Many similar Bragg peak explosions have been observed multiple times with different samples. In all cases, the Bragg peak was seen to be stable for a period of time varying between 10 and 30 minutes before the peak started to change. The intensity of a few of these exploding Bragg peaks was integrated radially and turned into a intensity curve versus q. This is shown on Figure 3.26 for a (111) Bragg peak and on Figure 3.27 for a (200) Bragg peak.

Figure 3.26 is the integrated data from Figure 3.25 and is a typical example of a Bragg peak which was seen from the start of the measurement. The peak eventually breaks up into many parts. One part remains at the expected q value of the ferritin crystal while other parts slowly move to higher q with time.

In some case, the Bragg peak did not break up in many distinguishable parts and it was stable for a period of time before the entire intensity shifted to a higher value of q. This is shown on Figure 3.27 for a (200) peak. Compared with Figure 3.26, the entire crystal now undergoes changes with time and not just parts of it.

A change in the q position of a peak is indicative of a change in the lattice spacing of the crystal. Such changes in lattice spacing have been observed before for protein crystals and in particular ferritin crystals [64]. These authors showed that the lattice of ferritin crystals tends to expand when exposed to x-rays. However, all of these measurements were

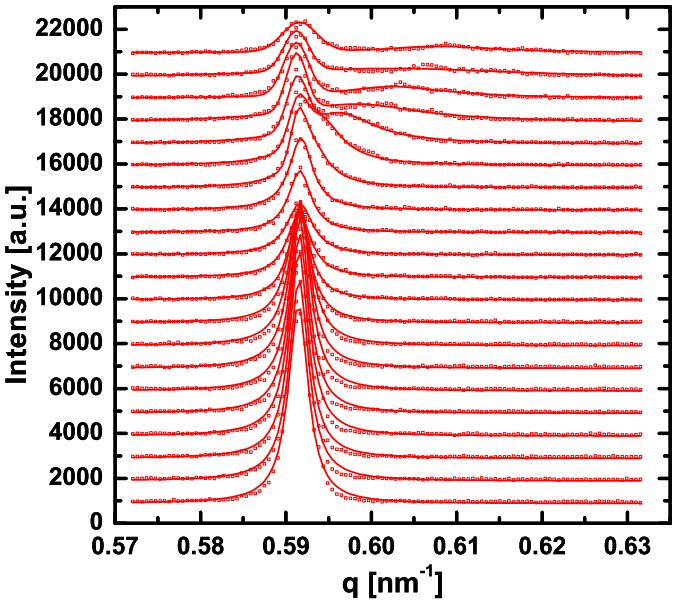


Figure 3.26: Radially integrated intensity from a (111) Bragg peak versus time. Each curve corresponds to a different time, with time moving from the bottom curve to the top curve in increments of 95 seconds. The circles represent the data with the solid curve corresponding to a fit using a simple Hendricks-Teller model.

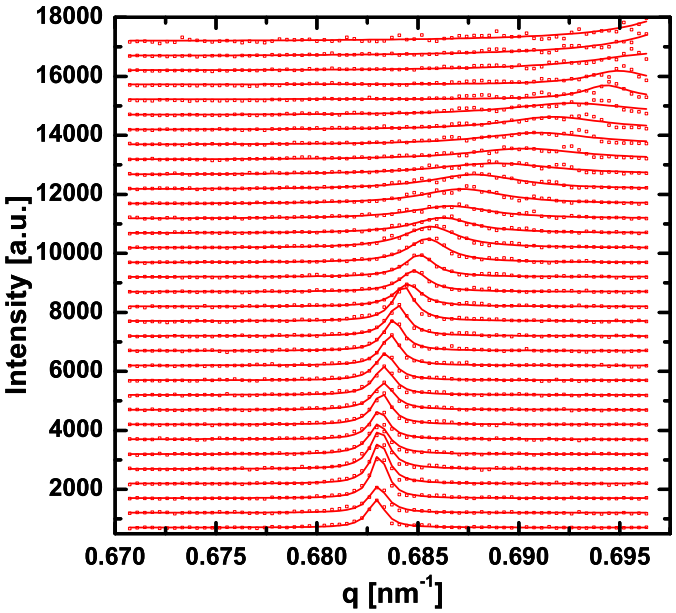


Figure 3.27: Radially integrated intensity from a (200) Bragg peak versus time. Each curve corresponds to a different time, with time moving from the bottom curve to the top curve in increments of 117 seconds. The circles represent the data with the solid curve corresponding to a fit using a simple Hendricks-Teller model.

performed at temperature between 80K and 180K while our measurements were all at room temperature.

While changes in lattice spacing are not surprising with radiation exposure, what is very surprising about the results shown in Figures 3.26 and 3.27 is that we observe a change toward higher q, which corresponds to a smaller spacing in real space. Therefore we observe a contraction rather than an expansion which is contrary to what most studies on the effect of radiation on the lattice spacing of protein crystals have revealed. However, some cases of contraction with radiation dose have been published by Ravelli *et al.* [64]. While most of their results indicate an expansion of the lattice with dose, all of these were at low temperature. The highest temperature measurement they performed was at 180K and this single measurement actually showed a lattice contraction.

Radiation damage to protein crystals has been shown to be dependent on dose and not dose rate [58]. The relevant quantity here is the dose and not the exposure time. We therefore calculated the dose absorbed by the small crystals per second. The density of the ferrihydrite in the iron core is roughly half the density of iron. Using that fact, we can calculate the x-ray absorption coefficient of the iron core. We also calculate the absorption coefficient of the water in the crystal as well as the protein shell itself. The absorbed fraction is given by

$$\text{Abs} = 1 - e^{-\mu_{\text{Fe}} l} e^{-\mu_{\text{Prot}} l} e^{-\mu_{\text{H}_2\text{O}} l} \tag{3.3}$$

where μ_j is the absorption coefficient from component j and l is the path length through the crystal. The main contribution to total absorption is from the iron. We have fitted to width of the peaks measured and deduced the size of the crystal from it. The crystal size is $3\pm0.25\mu$m and from that we get that $\sim$8% of the x-rays hitting the crystal are absorbed by it.

The number of x-ray hitting the crystal per second can be obtained using the measured current in the ion chamber beam monitor. For a beam of $25\times25\mu\text{m}^2$ at 9keV, the measured flux is 1.64×10^8 photons/sec. Therefore the number of photons hitting a $3\times3\mu\text{m}^2$ crystal is 2.37×10^8 photons/sec. The standard unit of dose is the Gray (Gy=Joules/kg). We can turn the number of photons per second into an energy per second using the energy of the photon, which was 9keV. We then get 3.4×10^{-9} J/sec crossing the surface of the crystal. We know the mass if the protein and the structure of the crystal and we can then calculate the total mass of the crystal and get 2.1×10^{-14} kg. The absorbed dose in gray per second is then

$$\begin{aligned} \text{Dose} =& \text{Abs} \times 3.4 \times 10^{-9} \frac{\text{J}}{\text{sec}} \times \frac{1}{2.1 \times 10^{-14}\text{kg}} \\ & 1.3 \times 10^4 \frac{\text{Gy}}{\text{sec}} \end{aligned} \tag{3.4}$$

This is our conversion factor from time to dose. This number is only an approximation since we do not know for sure the size of the crystal and the incident beam intensity could fluctuate a lot, especially if the crystal was close to the edge of the beam. The number could easily fluctuate by 10% but it is a good ballpark figure. Henderson has shown that the dose

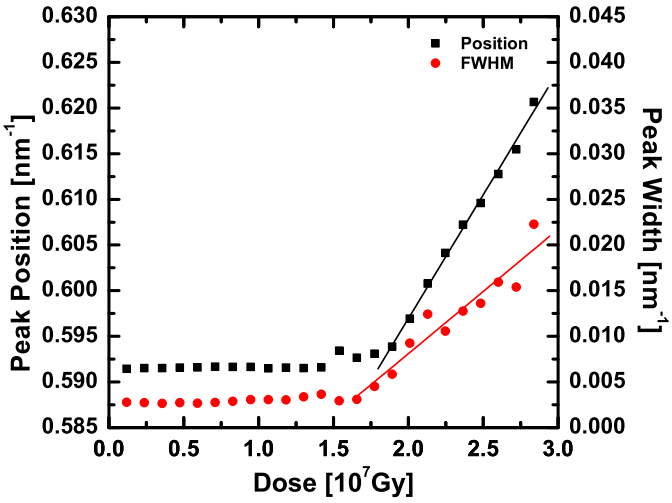


Figure 3.28: Position and width of the (111) Bragg peak shown in Figure 3.26 versus dose. There is a stable latent period followed by a roughly linear in time increase with dose of both parameters.

limit where spot fading becomes significant for protein crystals at cryogenic temperature is 2×10^7 Gy [30]. The limit is significantly lower by a factor of 70 for room temperature [58]. Nevertheless, with such a dose rate, we expect to reach the Henderson limit in roughly 30 minutes.

To quantify the lattice spacing change with dose, each curve was fitted to a Gaussian and the position of the peak was extracted as well as its width. The two quantities are plotted versus dose on Figure 3.28 for a (111) peak and on Figure 3.29 for a (200) peak. In both cases, there is a latent period where there is no change to the lattice spacing and the width of the peak. After a certain time or dose which varies from crystal to crystal, there is a close to linear increase in the position of the peak and the width of the peak.

The data shown are surprising, but clearly indicative of a collapse of the crystals with increasing dose. The gradual change versus dose is indicative of a cumulative process. To attempt to understand the behavior, we consider that over time and with increasing dose, some specific and non-specific damage occurs to individual proteins in the crystal. The radiation presumably causes some point defects randomly distributed in the crystal by destroying parts of the protein, at first changing slightly its quaternary structure. It is known that a crystal can accommodate a limited density of point defects while maintaining its long-range order and therefore a crystal can withstand some small changes in the structure of some of the proteins without any visible adverse effects. However, after a certain dose, the defects in the crystal reach a critical density and the structure of the crystal is no longer stable. Some parts of it therefore reorganize in a more stable structure. The reorganizations lead to the formation of small point defects in the crystal. In our case, the evidence indicates that these defects are contractions in the spacing of the crystal layer. Once these defects start to occur, their number keeps increasing until they are too numerous and the whole crystal is no longer stable and completely breaks up until a Bragg peak is no longer seen.

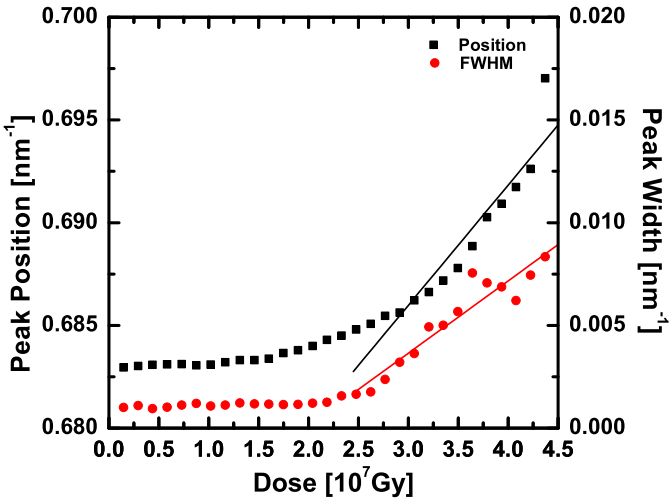


Figure 3.29: Position and width of the (200) Bragg peak shown in Figure 3.27 versus dose. There is a stable latent period followed by a roughly linear in time increase of both parameters.

In order to make this a little more quantitative, we performed an analysis of the data using a Hendricks-Teller type model [31]. Such a model is often used to study the diffraction from crystals with a varying number of random defects with a size mismatch to the host [68]. The exact model, in one dimension, applies to a linear array of objects spaced apart by distances L_j chosen at random with probability f_j. The diffracted intensity from a crystal with defects is thus modeled as a regular spacing L_0 with randomly inserted defects of spacings L_j of probability f_j. The intensity is given by the following expression

$$I(q) = \frac{1 - C^2}{1 - 2C\cos\phi + C^2} \tag{3.5}$$

where

$$C = \sum_j f_j \cos(qL_j + \phi) \tag{3.6}$$

where ϕ is an average crystallographic phase factor given by

$$\tan\phi = \frac{\sum_j f_j \sin(qL_j)}{\sum_j f_j \cos(qL_j)} \tag{3.7}$$

The f_j are the probabilities that a spacing of length L_j will occur.

In our case we used the simplest model possible with only one type of defects with a spacing different from the expected crystal spacing. The intensity profile was then completely determined by performing a fit of the data with only two free parameters : a spacing L_1 and a probability f_1 in a host of known spacing L_0 with probability $f_0 = 1 - f_1$. Each curve shown as a solid line on Figures 3.26 and 3.27 are the fits to the data using this Hendricks-Teller two state model. The two fitted parameters, a probability for the defects with a lattice spacing of the defect are shown on Figure 3.30 for the (200) Bragg peak.

The probability of occurrence of defects is fitted to essentially zero for the latent period indicating that even though there has to be damage to individual proteins, the damage is small enough to not cause any change in the crystal structure. At the end of this latent period, the Bragg peak position starts to change and a non-zero value of the probability of defects is fitted. A large number of individual defects on separate proteins combine to cause a structural change to the crystal. A layer of the crystal collapses to a distance a little closer to the previous layer than it should be. Over time as the absorbed dose increases, more of these defects occur and many more layer spacings collapse to values smaller than the fcc crystal spacing. Eventually, a probability of 0.55 is reached before there is no diffracted intensity measured, which means that the crystal diffracts until 55% of the layers have collapsed.

The fitted values of the spacing seen on FIG. 3.30b) are unreliable during the latent period due to the fact that the probability is essentially zero. A defect of any size which has a zero probability of occurrence will give the same diffracted intensity as any other defect with a zero probability. This latent period is seen to last until a dose of roughly 1.5×10^7 Gy is absorbed. After that point, the spacing fitted can be trusted and it is seen to fluctuate slightly around a value of 90Å. This value was allowed to vary freely because we have no reason to expect a particular new spacing to occur. The process by which this new spacing is created is a highly uncontrolled destructive process which we have no means of predicting. The perfect crystal spacing for the (200) Bragg peak of fcc ferritin is 91.5Å. This means that individual layers of the crystal collapse by 1.5Å or 1.6%. This is fairly large considering the fact that the fcc crystal of ferritin is close packed and the distance should not be allowed to contract. The contacts between proteins is along the face-diagonal of the crystal and mediated by 2 Cd^{2+} salt bridges. There is essentially no room between the proteins along the face diagonal with the insertion of these salt bridges. In order for the (200) lattice spacing to collapse, which corresponds to the edges of the cubic unit cell, one mechanism might be entire removal of the salt bridges, which each occupy roughly 3Å. The damage from the x-rays must cause these links to be destroyed, most likely by destroying the residues involved in the contact. However, that fact alone does not explain why the crystal collapses. The removal of the salt bridges could just as well cause the proteins to move apart rather than move closer together. Water could fill the empty space between the proteins and cause them to move apart. There must then also be a removal of the interstitial water near the protein contacts allowing the observed collapse. This could be due to a heating effect from the intense x-ray beam causing the interstitial water to boil off. This is however unlikely considering that the sample is mounted on a large heat sink which means that very slow heating of the sample cannot occur. It is conceivable that rapid heating could occur for a small volume on time scales shorter than the time is takes for the heat to be conducted away. However the whole process occurs over tens of minutes which is more than enough time for the heat sink to cool any hot spot in the sample. Also, it is a well-known fact that x-ray radiation damage is a dose related problem [58], which indicates the main factor is not boiling of the solvent.

Our data is low resolution data which does not allow us to make any specific claims as to the mechanism involved in the destruction of the crystal. Since we only measure the

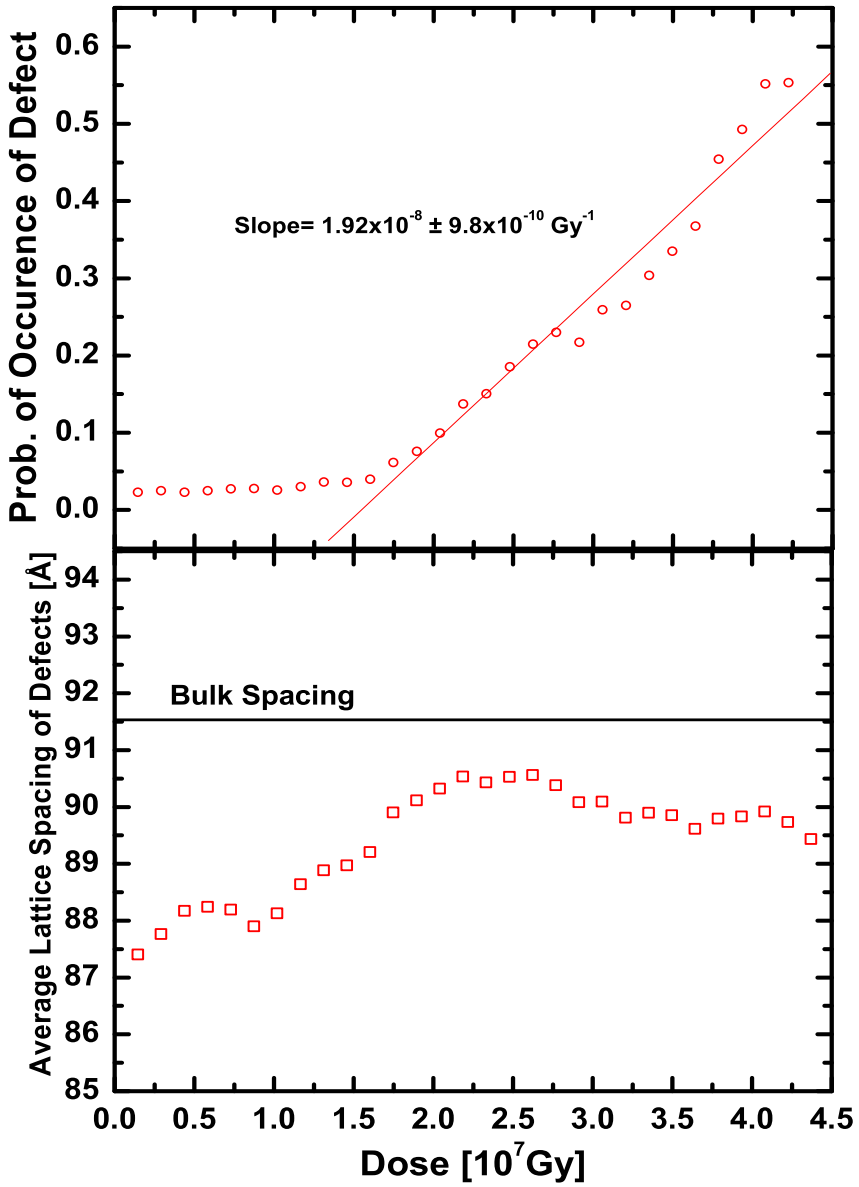


Figure 3.30: a) Probability of occurrence of a radiation induced defect in the crystal obtained by fitting a Hendricks-Teller model to the time evolution of a (200) Bragg peak. After a latent period, the probability increases linearly with time. b) Fitted value of the spacing of the radiation induced defects in the crystal. The values obtained for the latent period are unreliable. Once the crystal starts to disintegrate, the fitted value is fairly constant (90Å) and is about 2Å below the indicate bulk value for a defect free crystal (91.5Å).

first orders of Bragg peaks, we are fairly insensitive to small changes in individual proteins caused by the beam. It is well-known that the higher order reflections disappear first during data collection on a protein crystal. By the time we measure a significant change in the (200) peak, the crystal should have long stopped diffracting at higher q. We are examining the stability of the crystal lattice rather than the specific changes in individual molecules. Spot fading studies, sensitive to the latter effect, find changes at high q at lower doses than what we see. However, the measured onset of the linear changes in the lattice with dose matches well with the Henderson limit. This is surprising because it is expected that the limit is 70 times lower at room temperature, where our measurements were made. There is speculation that very small protein crystals near 1 micron in size would be less sensitive to radiation than larger crystals since the photelectrons created by the interaction of the x-rays with matter would be able to escape the small crystal before creating a lot of damage [59]. Our results could be an indication of that. However, we cannot tell what is happening to the higher diffraction orders and can't conclude about the spot fading of these peaks.

Although no specific crystal interactions can be inferred from our data, we can still learn some interesting facts about the late stage of radiation damage on protein crystals, especially small crystals only a few microns in size. The entire crystal is illuminated by the x-ray beam and the measurements is indicative of the entire crystal. It is interesting to note that the crystal can survive with as much as 55% defects in the lattice spacing. Also, the crystals seem to break up in many (between 2 and 8) separate blocks which all have slightly different lattice spacings and each of these blocks is stable for a significant amount of time, as much as 5 minutes.

3.7 KB Mirror Focusing

KB focusing mirrors were used for a few experimental runs to increase the signal level of the CXD data measured. A stronger signal should allow the measurements to be performed faster therefore minimizing the effect of the thermal motions in the sample. The KB mirrors focused a 25×25 μm^2 beam down to a 2×2 μm^2 beam with close to 100% reflectivity. This leads to an increase in the incident intensity per area by a factor of roughly 150. Therefore the accumulation time needed to obtain good statistics on the data could be accordingly reduced.

However, problems accompany this increased flux. The first one is obviously an increase in radiation damage. Damage leading to the collapse of the crystals previously observed to occur after 1500 seconds or so were now observed on a time scale of 10 to 30 seconds. A second problem with the focusing mirror was related to the small size of the beam. With a beam only 2 microns wide, it was much easier for crystals to diffuse in and out of the beam. Thermal motions were found to be large, especially rotations. Direct measurement of translations was impossible. However, the fact that often Bragg peaks suddenly appeared or disappeared is indicative of crystals diffusing in or out of the beam. The smaller beam gave a much small illuminated volume and crystals moved in and out of it rapidly. Therefore, the use of the KB mirrors did not allow us to obtain better statistics. The experiments could be done faster, but without any gain in statistics.

3.8 CXD Data

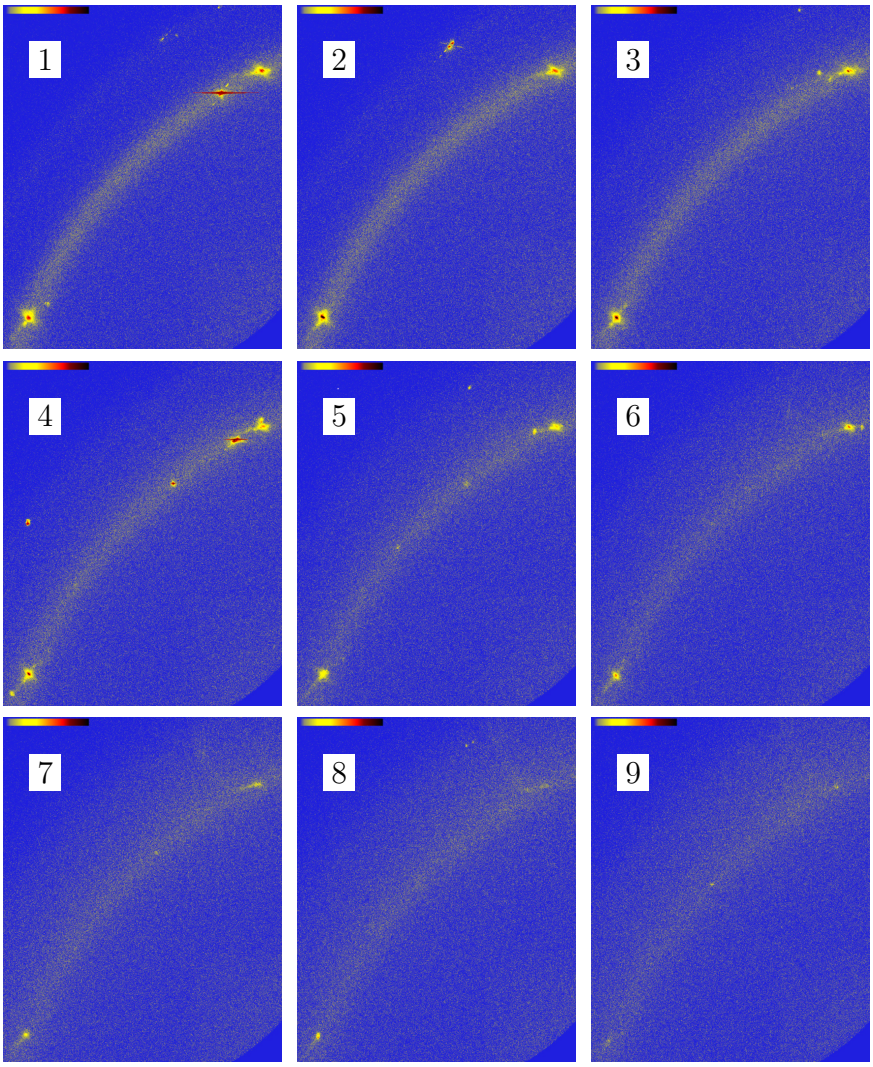


Figure 3.31: Intensity on a logarithmic scale for multiple exposures on a sample containing many small crystals of holoferritin. A continuous ring is seen due to the presence of a large number of small crystals. A few individual peaks can be identified coming from single crystals.

The previous few sections of this thesis have dealt with the technical difficulties of performing coherent diffraction imaging measurements on micron size crystals of proteins which are suspended in solution. The thermal motions in the sample as well as the damage to the sample caused by the x-ray beam limit the accumulation time for a single crystal. Long exposures are required to measure the weak fine features near the intense Bragg peak. However, we have shown that after a short period of time, which is generally shorter than the time that would be required to adequately perform the measurement, the sample has changed significantly and often to the point that a Bragg peak is no longer seen. There are large changes over time which make the measurement difficult but also small variations

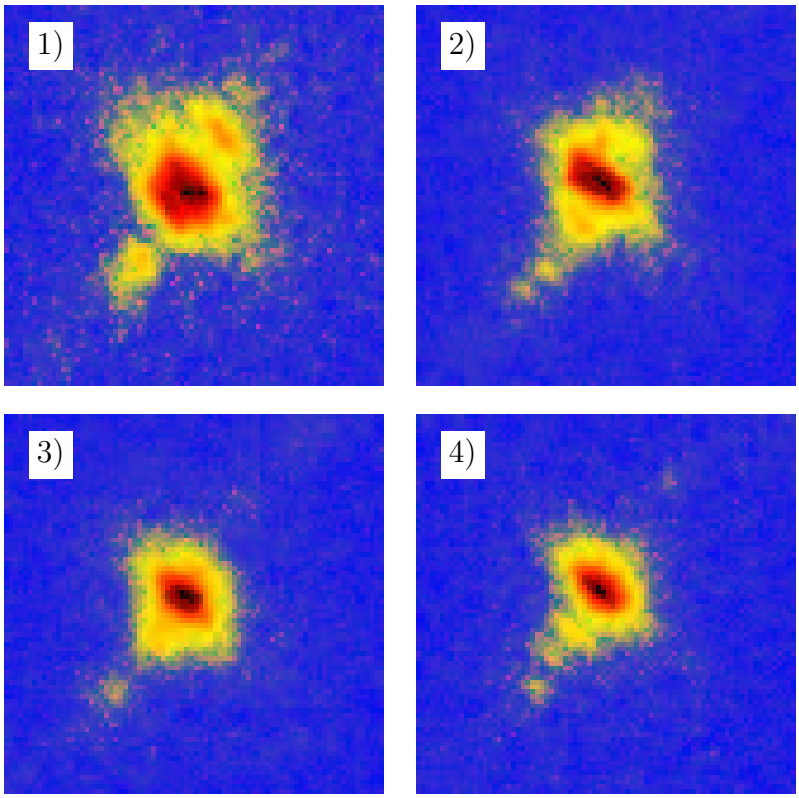


Figure 3.32: Intensity on a logarithmic scale of a single Bragg peak from the same crystal over time. Each image corresponds to an exposure of 60 seconds.

over time can end up smearing the data to the point where useful information is lost. Rapid oscillations of the crystal would lead to an averaging of many slices through the diffraction pattern which would smear the coherent diffraction features. Despite all the technical difficulties limiting the effectiveness of the technique, some CXD patterns displaying expected features such as intensity fringes were measured and they are presented here.

Figure 3.31 shows a time series on a sample that was prepared in a drop, starting with 4μl of holoferritin at 100mg/ml to which 4μl of 45mM $CdCl_2$ was added. Crystals were formed right away and data were collected immediately after the experimental hutch could be closed. Each image in the time series corresponds to an exposure time of 60 seconds. The continuous {111} ring can be seen indicating the presence of many smaller crystals. The effect of radiation damage on the powder ring can clearly be seen in this time series. The ring becomes broader and weaker over time as crystals become destroyed by the beam. Only a few individual Bragg peaks can be identified on the ring. On the first image, 3 peaks are observed. The second one from the top right displays a sharp streak. This is not a streak due to the sample. It is due to the fact that the intensity at the central pixel of this peak was too high and the pixel was saturated. In such a case the saturated pixel tends to leak into neighboring pixels in the direction of readout of the CCD chip.

We focus our attention on the Bragg peak at the bottom left of the first image. It is

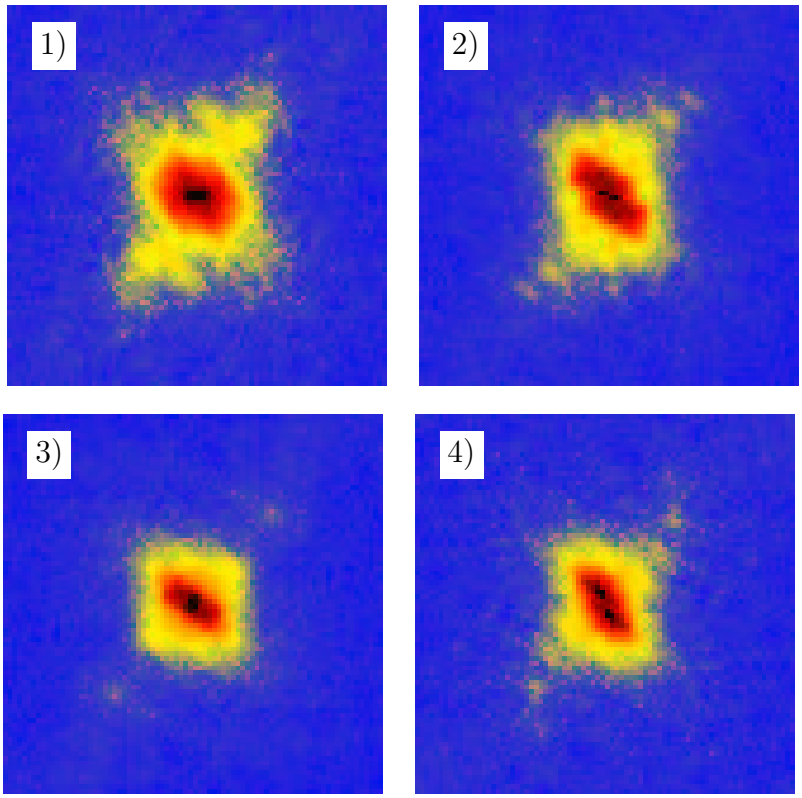


Figure 3.33: Symmetrized intensity on a logarithmic scale of CXD patterns from a single crystal. The patterns of Figure 3.32 were symmetrized about the center of the pattern.

seen for a few images, gets progressively weaker and eventually disappears completely. A zoom on this peak for the first four images is shown on Figure 3.32. Intensity flares with modulations, which are expected for faceted crystals can be seen. The patterns are seen to change in time which is expected from the thermal motions and the radiation damage discussed previously. These CXD images are not particular good when compared with the ones obtained on gold or lead nanocrystals [70, 89]. However, when considering the much weaker scattering power and the shorter time one has to perform a measurement, the CXD data shown here is not too bad.

The CXD pattern, which corresponds to the Fourier transform of the shape of the crystal should be centrosymmetric since the particle is supposed to have a real electron density and is assumed to be a perfect crystal with exact translational symmetry. The patterns measured are clearly not centrosymmetric. There are two possible reasons for this. The first one is, as discussed before, that the 3D pattern is centrosymmetric but any random 2D slice through the pattern does not necessarily have to be symmetric. Only a slice through the exact center of the pattern will be symmetric. There is no reason here to expect the slice to be through the center. Therefore, each of these slices is the Fourier transform of the projection of the electron density onto some unknown plane, the detector plane.

Another possible reason for the absence of centrosymmetry could be the presence of

strain in the crystal as discussed in Chapter 2. Strain would show up as an imaginary part to the electron density in a reconstruction. The quality of the data does not allow us to accurately reconstruct a strain field for the crystal. Therefore, in order to obtain a reconstruction of the projection of the shape of the crystal, the patterns were symmetrized about the center. This is shown on Figure 3.33 were the same four images were symmetrized.

The symmetrized patterns were then put through a series of algorithms to retrieve the phases of the scattered waves in order to reconstruct the projection of the shape of the crystal.

3.9 Fitting Procedure

A series of programs to treat the data obtained with the CCD detector have been developed primarily by John Pitney [61] and Garth Williams [88]. The suite of programs can be used directly to convert the data into a format the phase retrieval routines can handle. The first step is to convert the raw data from the spe format obtained using the WinView™program into a binary format developed by John Pitney [61] called sp4. This is done using a program called **2dspetosp4**. The pattern is then centered roughly by eye and the image is cropped using **2dcrop**. The inverse Fourier transform of the intensity can be calculated using **dofft**. The result is the autocorrelation function of the electron density and this quantity should be real, that is the phase of the autocorrelation function should be constant. If a phase gradient is present, it indicates a miscentering of the diffraction pattern. The pattern is then recentered until the autocorrelation function displays no phase gradient.

Once a centered pattern is obtained, the scattered amplitude is obtained by taking the square root of the intensity using **dosqrt**. A background is subtracted from the array using **2daddbkgd**. The background value is determined separately by finding the average value of the pixels far away from the peak. The data is then padded to increase the real space resolution of the reconstruction. This is done by putting the array at the center of a larger array of zeroes using **2dcroppad2**. This array containing the amplitude of the scattered waves is real but needs to be converted into a complex quantity using **realtocomplex**. Due to the way the FFT packs the array, the final step is to change to order of the array by putting the center at the origin using **2dinvert**. The array of data is now ready to be used in the phase retrieval program.

The phase retrieval program called **2dmultifit8** is extensively described in the thesis of Williams [88]. It is based on the iterative method described in Chapter 2. It has many modes of operation corresponding to different real and reciprocal space constraints applied during the loop. It allows for different constraints to be used on different loops in order to possibly improve the fit. Along with the reciprocal space data, the program requires a real space starting array. This array needs to be of the same size and contains the real space support. This support can be created using the **2dpoly** program which creates a polygon of desired shape and size. This size needs to be guessed beforehand by looking at the spacing of the fringes in the data. The support has a value of 1 where the electron density can be non-zero and 0 everywhere else. For these data, due to the orientation of the flares, the support used was a square of 60×60 pixels rotated by 45°. The program requires a

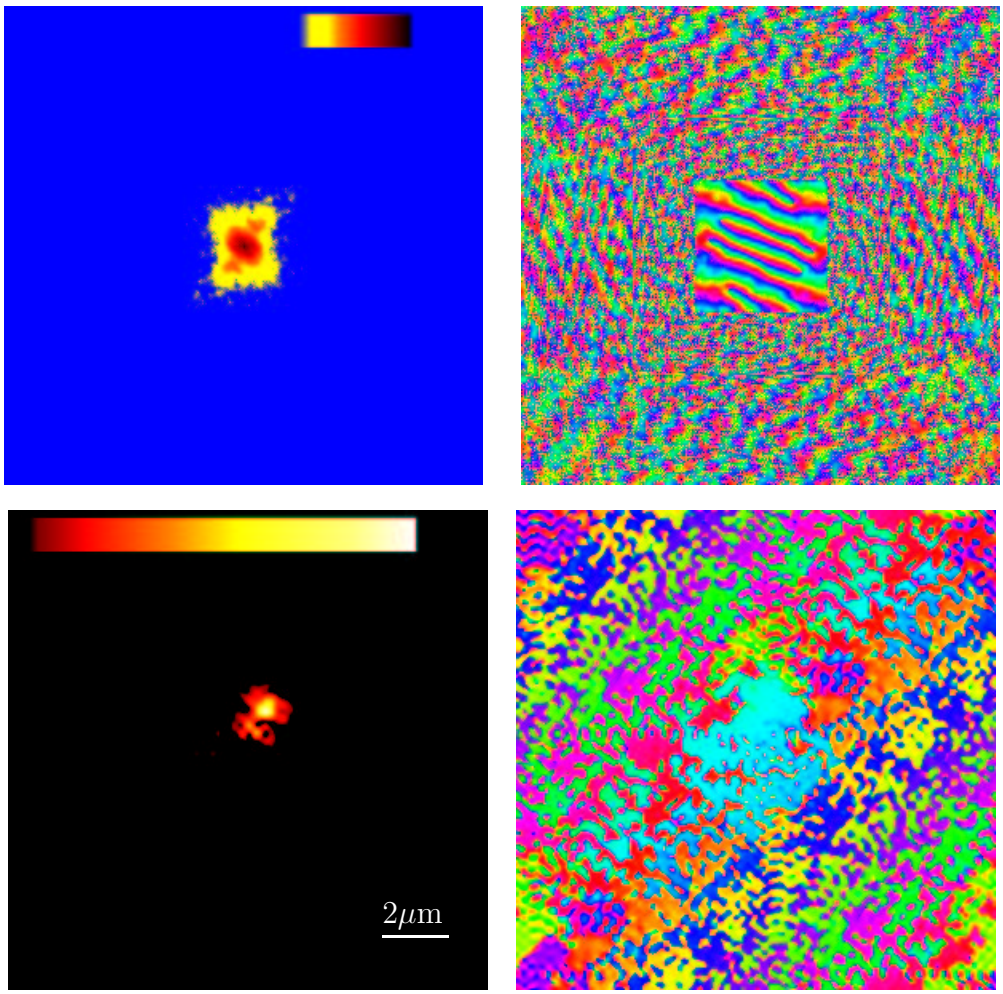


Figure 3.34: Fit of the symmetrized slice number 1 of Figure 3.33. Top left : Fitted scattered amplitude. Top right : Phases retrieved by the algorithm. Bottom left : Reconstructed amplitude of the electron density. Bottom right : Phase of the reconstructed electron density.

configuration file telling it which constraints to apply and for how many iterations, as well as how many fits to perform and how many of the best ones to keep. In the present case, we used a combination of Hybrid Input-Output (HIO) and Error Reduction (ER). Both of these are simply different constraints and are discussed thoroughly in the thesis of Williams [88]. A total of 50 different fits were performed with different random starting phases for each of the images and the 5 best fits were kept for each. The very best fit obtained is shown for two images on Figures 3.34, 3.35.

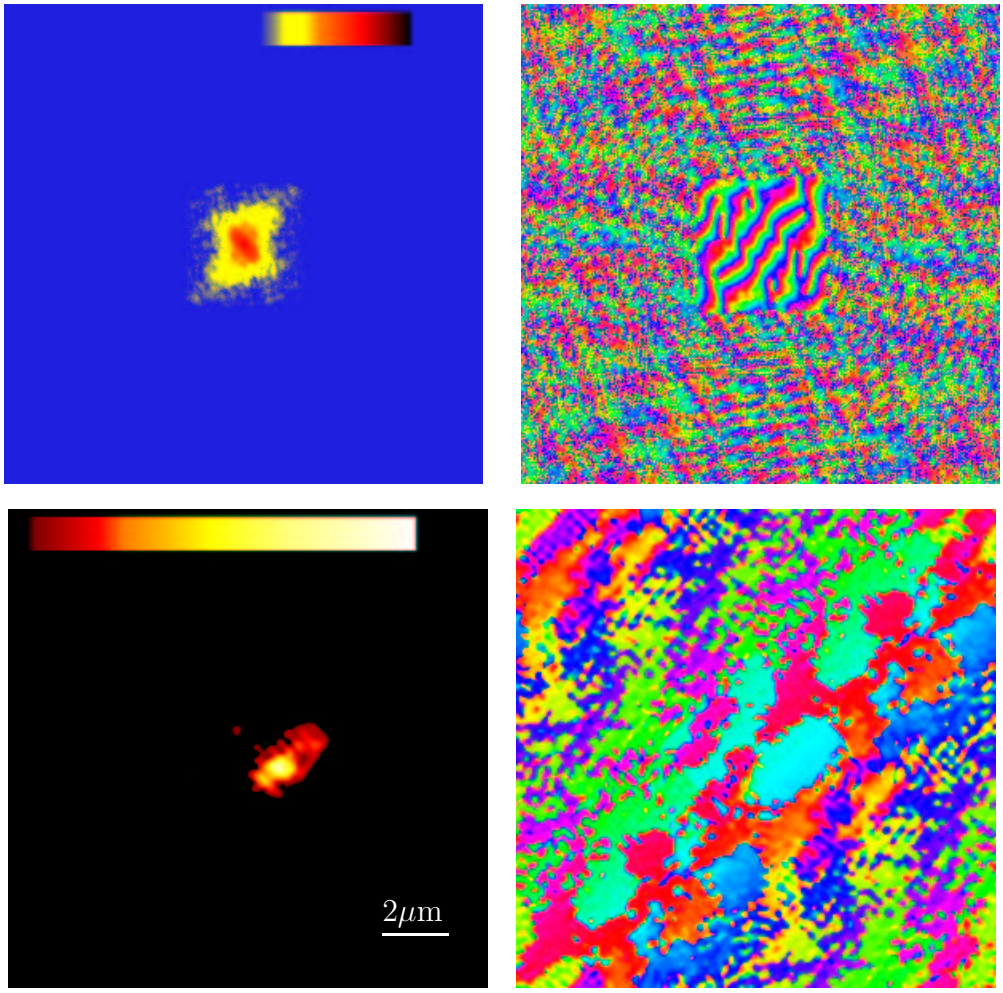


Figure 3.35: Fit of the symmetrized slice number 2 of Figure 3.33. Top left : Fitted scattered amplitude. Top right : Phases retrieved by the algorithm. Bottom left : Reconstructed amplitude of the electron density. Bottom right : Phase of the reconstructed electron density.

3.10 Interpretation of the Fits

The reconstructions obtained are more for a proof of principle than for a quantitative analysis at this point. Both of the reconstructions obtained are quite good considering all the difficulties on the experiment and the low quality of the data. Figure 3.34 shows the fit and reconstruction from slice number 1 from Figure 3.33. The top two images are the reciprocal space amplitude and phase from left to right. The bottom images are the reconstructed projection of the electron density of the crystal. The amplitude is on the left with the phase on the right.

The algorithm retrieved the reciprocal phase and used it along with the measured scattered amplitude to generate the real space images. The appearance of a square in the phase is due to the padding of the array. The measured amplitudes were only within this square box and they were included in the larger array of zeros. The phases outside this box are irrelevant since the measured amplitudes there are zero.

The reconstructed projection of the electron density displays a compact object with some well-defined facets on both reconstructions shown on Figures 3.34 and 3.35. Even though the same object is illuminated in both cases, the changes in time due to rotations and radiation damage explain why the two reconstructions look different. The rotation of the crystal means a projection of the electron density on two different planes is measured. In the second reconstruction, a slightly more elongated object is recovered with again fairly sharp interfaces. The size of the objects reconstructed is around 2μm.

The real space phase of the reconstructed object is interpreted as the projection of the strain onto the q vector. The phase is seen in both cases to be constant where the electron density is non-zero. This indicates there is no strain in the crystal and the recovered electron density is real. This is not surprising since we forced it to be so by centrosymmetrizing the measured scattered amplitude. The flat phase is simply a consequence of that.

The top left shows the fitted scattered amplitude obtained from the reconstructed electron density. Comparing the fitted amplitude with the data, we get value of χ^2 of 0.021096 for the first image and 0.0274489 for the second. This means the fitted amplitude is on average 21% and 27% different from the measured amplitude.

Using the program called **2dreprod**, we can estimate the reproducibility of the reconstructed electron density since multiple reconstructions were performed on a single diffraction pattern. The program calculates a value similar to χ^2, that is the sum of the squared differences between the two solutions, normalized by the sum of the squared density of one of the 2 solutions. A comparison of the two best fits yields an average 20% difference per pixel for the first slice shown on Figure 3.34 and a 10% difference for the first slice shown on Figure 3.35. Comparing the other fits for each reveals that the other solutions found with higher χ^2 rapidly become very different. Therefore, the reconstructed electron density is not very reproducible and strongly depends on the initial random phases.

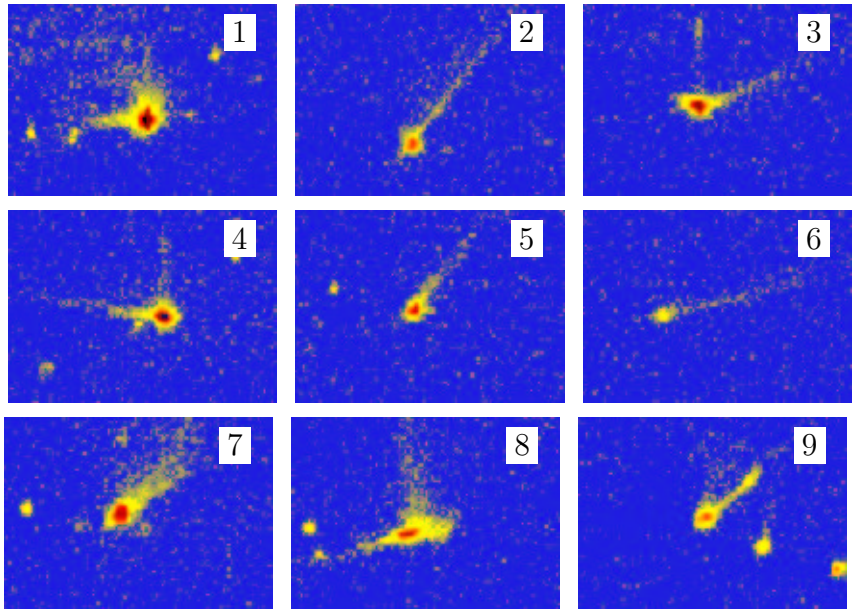


Figure 3.36: Multiple examples of single Bragg peaks with intensity streaks. The direction of these streaks is neither radial nor in the direction of the CCD readout and is always consistent with either a (111) or (200) direction.

3.11 Simulations

The lack of quality of the CXD data makes it impossible to obtain more than a two dimensional projection of the electron density in some rare cases. However, some extra information can be deduced from the overall characteristics of the many time series than were measured. A typical example of such data is shown in Appendix A obtained on a sample prepared in a gel. Each image corresponds to an exposure of 50 seconds. Every so often, some of the Bragg peaks seen display sharp streaks of intensity in a direction which is not radial. Some examples are shown zoomed in on Figure 3.36.

These streaks appear and disappear with a corresponding Bragg peak since the crystals are rotating. The streaks therefore come from a single crystal. Furthermore, in some cases, the streak and the central part of the peak can be seen to change over a few images. Sometimes, the streaks are seen first as the crystal rotates into the Bragg condition and the Bragg peak then appears with a corresponding change in the intensity of the streak. This is the signature of a rotating crystal with facets. Intensity streaks which are not radial can only come from a facet of the crystal in a direction different from the $\mathbf{q}$ vector [6]. The presence of these streaks is not surprising since faceting of the crystal is expected. What is unexpected is that most of the streaking is toward a higher q value than that of the Bragg peak. The streaks are almost always on the high q side of the Bragg ring. Two of the 8 $\{111\}$ directions in the crystal are radial on the CCD detector. Any rotation around this vector is allowed. There is no preferred orientation and therefore no reason to expect facets to be only pointing in directions leading to streaking to higher q. For a real object, the diffraction pattern is supposed to be centrosymmetric and streaks should be seen on both

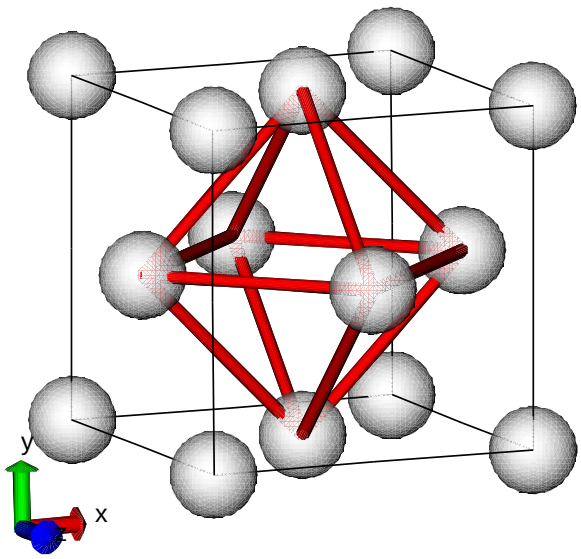


Figure 3.37: The close-packing directions in an FCC crystal are in the {111} directions. The intersection of these planes define an octahedral volume. The close packed faces are more stable and therefore we may expect the crystals to have an octahedral habit.

sides with equal probability but this is clearly not the case.

The expected shape of a face centered cubic crystal is an octahedron with the 8 facets corresponding to the {111} directions. This is expected because the {111} directions are the close-packed directions in the crystal as shown on Figure 3.37. The close-packing means that these are the directions where the fewest bonds are left dangling in order to expose the surface. Even though the {111} faces are the most stable, they are not expected to be the only surface present. However, any large facet on the crystal giving rise to a strong streak in intensity is expected to be a low order surface with fairly tight packing such as the {111} and {200} directions.

The angles between all 8 {111} directions and all 6 {200} directions are fixed by the lattice. The angle between two (111) faces has to be either 71°, 109° or 180 °, while the angle between two (200) directions is always a multiple of 90°. The angle between a (111) face and a (200) face can take on a value of 35°, 55°, 126° or 144°. Therefore, by measuring the angle between the q vector of the Bragg peak, which is always the (111) direction and the direction of the streak in intensity, we can determine if the facets producing the streaks are {111} or {200}. These angles are shown on Figure 3.36 and they are found to always correspond, within the accuracy of the measurement to one of the allowed angles between the (111) and either another {111} or a {200} direction. We can therefore conclude that the small crystals seen possess primarily {111} and {200} facets.

Turning now to the question of why the streak almost always point toward higher q. To answer this question, we performed some simulations. We first assume that the crystal has an octahedral shape with {111} facets as shown on Figure 3.37 and on Figure 3.38 in green. The actual shape of the crystal does not matter in what follows so we picked a convenient and likely shape of the crystal. The Fourier transform of this shape is shown on Figure

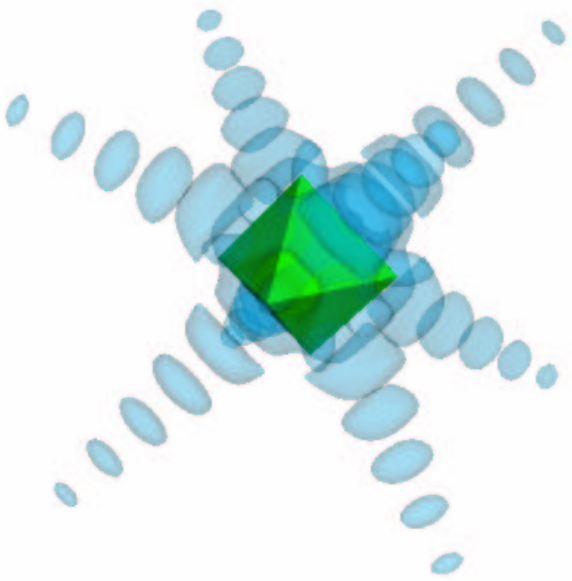


Figure 3.38: Simulation of an octahedral FCC crystal shown in green with its Fourier transform shown in transparent blue. Each {111} crystal face gives rise to a streak of intensity perpendicular to it. The streaks from two parallel faces interfere which gives rise to modulations in the intensity, *i.e* fringes.

3.38 in transparent blue. The real and reciprocal space are shown together on obviously different scales in order to emphasize the relationship between faceting in real space and intensity flares in reciprocal space. There are 8 streaks and choosing one of them as the radial direction determines that 4 of them point toward a lower q value while the remaining 4 point toward higher q.

Now suppose there is strain present in the crystal. This will introduce asymmetry in the intensity. The strain is modeled as an imaginary part of the electron density as discussed in Chapter 2 and the Fourier transform of a complex object is generally not centrosymmetric. The intensity flares from the presence of facets are still present but the intensity of the flare on on side of the peak is not longer the same as the corresponding one on the other side. This is shown on Figure 3.39.

The question is then what kind of strain field will lead to enhancement of the intensity on the high q side with strong cancellation on the low side. One possible answer is a surface contraction. If the top layer of the crystal is shifted down from the bulk value, the asymmetry to the Bragg peak will be toward higher q. This is what is shown on Figure 3.39 where a 2D slice through reciprocal space of a small octahedral fcc crystal is shown in the same way as it was shown on Figure 3.8 in the case of an unstrained crystal. The presence of the surface strain has the effect of distorting the diffraction pattern radially from the origin. The zeroth order peak at the origin is unchanged by the presence of strain and the entire reciprocal space is still centrosymmetric about the origin but no longer locally symmetric around each Bragg peak.

The simulation shown on Figure 3.39 was done for a fairly large amount of strain (16%)

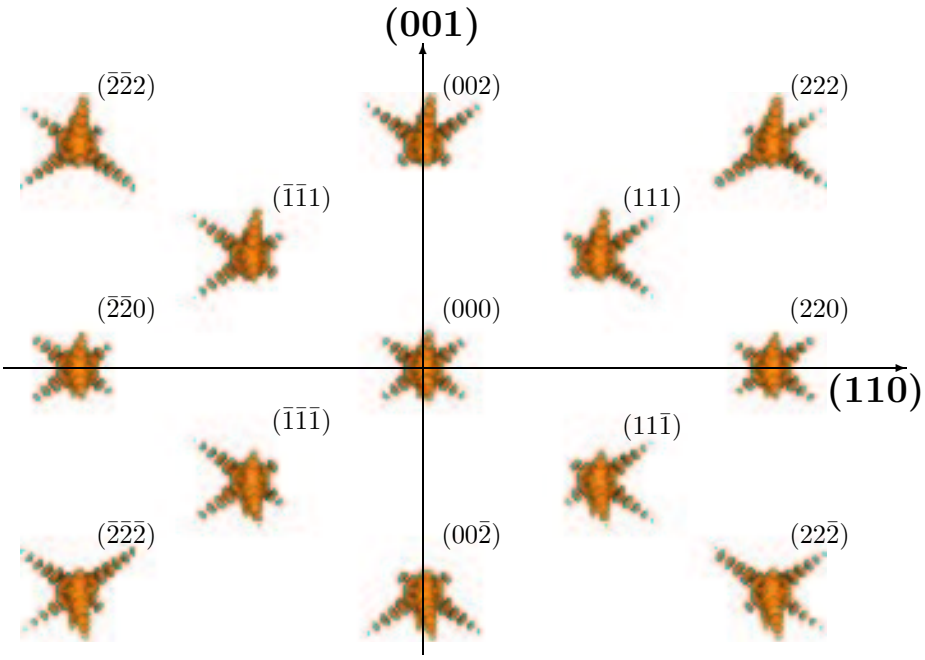


Figure 3.39: Cut through reciprocal space of an octahedral fcc crystal with 16% of contraction of the spacing of the top layer in the crystal. This surface strain causes the Bragg peaks to be asymmetric while the unstrained crystal leads to centrosymmetric Bragg peaks as shown in Figure 3.8.

in order to emphasize the effect. This means the top layer at every surface is shifted down by 16% of the bulk spacing. The bulk spacing of the (111) direction in ferritin is 106Å and the strained spacing is 89Å. The actual amount of surface strain cannot be determined accurately with the data obtained.

The presence of surface strain giving rise to the observed asymmetry toward higher q is a reasonable explanation. Such a surface contraction was previously measured on large millimeter size ferritin crystals. The crystal truncation rod [67] was measured and also showed a strong asymmetry toward the high q side. This was fitted to a 4% contraction of the top layer of the crystal [6].

The surface contraction of the small crystals does not completely cancel the streaking at lower q, it just reduces its intensity relative to that of the higher q side. Therefore we should still be able to observe some streaking toward lower q and this was indeed the case. It was simply observed very rarely compared with the high q side.

Therefore, we conclude that the small micron sized crystals possess mostly {111} and {200} facets and there is significant surface contraction or strain present in these crystals.

Chapter 4

Nucleation at Low Temperature

It often proves necessary in the study of biological samples with x-rays at synchrotron sources to cool the sample down to cryogenic temperatures in order to extend the life of the sample. Radiation damage is found to be mitigated at such low temperatures [58]. Furthermore, as discussed in Chapter 3, the motions of micron sized crystals in solution make it difficult to observe very small crystals. It was then thought that freezing the sample to cryogenic temperature would not only reduce the observed radiation damage but also dampen out the motion of the small crystals, thus allowing for a longer stable period for measurement. Experiments were then performed on frozen samples of ferritin as described in Chapter 3 yielding a surprising new feature to the scattering pattern. This current chapter is dedicated to the study of this new feature, a broad peak in the SAXS pattern.

4.1 Facility and Equipment

The experiments were performed at Sector 34-ID-C at the Advanced Photon Source at Argonne National Laboratory. The beamline used was the same as described in Chapter 3. Only slight modifications to the setup described above were required in order to perform small-angle scattering experiments.

4.1.1 Experimental Setup

The SAXS scattering setup is schematically similar to the CXD setup shown on Figure 3.2b) with a slight difference with the detector arm. The camera in the CXD experiments was far away from the sample and also set slightly off-axis to measure the first Bragg peak. In the SAXS setup the camera is moved closer to the sample to a distance between 300 and 1000mm. This is needed because the size of the CCD chip is small and the camera needs to be close to subtend the desired angle. The camera is also placed on-axis with the beam stop protecting the center from damage from the beam.

The key difference between the CXD and SAXS setup is that the latter consisting of measurements much closer to the direct beam, much greater care must be taken to produce a well collimated beam devoid of unwanted spurious features such as scattering from slits. The beam defining slits cut the edges of the beam and this leads to some flares coming from them. These flares can be quite strong, often more intense than the signal from the sample. They arise mostly from specular reflection of the beam touching the edges of the beam. They are obviously not desired and can be dealt with by using multiple slits. By

placing the beam defining slits first, followed by cleanup slits further downstream, the flares can be eliminated. The second slits are set to a gap slightly larger than the first slits so that they do not cut into the direct beam and cause flares themselves. They therefore block the scattering from the first slit while letting the beam through and not causing more unwanted scattering. The cleanup slit is placed as close to the sample as possible.

The monochromatic beam is used and slit down to a size of 50 to 500μm depending on the sample. The beam does not need to be coherent and can be much larger than for CXD measurements. A wider beam illuminates more of the sample and increases the signal level. Ideally, the size of the beam is matched to the size of the sample. To make the beam larger than 200μm, the roller-blade slits were pulled all the way out of the beam and the set of rough slits 1m upstream were used as the beam defining aperture. The roller-blade slits proved to not be ideal for SAXS measurements because they produce a lot of intensity flares at small angles. The polished round surface of the molybdenum cylinders reflect the x-rays very well at all angles. The roller-blades could still be used but more effort was needed to prepare a quality beam. In most cases, the size of the beam used was between 50 and 200μm. The small beam was chosen because it provided enough signal and it was easier to use on the samples mounted on the Peltier cooler as shown on Figure 3.9. A small beam could be scanned along the vertical while a larger beam would have hit the edges of the sample faster and more importantly hit the Peltier casting a large shadow on the detector.

4.2 Experiment

4.2.1 Sample Preparation

As was the case for the CXD measurements, the samples were prepared in both drop placed in air or in capillaries as shown on Figure 3.9. The samples consisted of protein solutions at varying concentrations and varying solvent conditions. High concentrations were obtained by centrifugation as described in Chapter 3.

4.2.2 Data Collection

The temperature dependence of many samples were investigated in the 20°C to -40°C range. For each sample, the SAXS pattern was measured at a temperature above freezing, generally at 10°C. At this temperature, the sample was stable in solution and the form factor of the protein could be measured since the interference between different protein molecules was minimal. The structure factor was close to 1 in the region of reciprocal space measured although at the highest concentrations measured, some features to the structure factor could be seen as will be discussed in Section 4.3.

Each SAXS pattern was measured using a varying exposure time generally around 3 minutes including the readout time of the camera. The exposure time per accumulation was set so that no point on the detector would be saturated with the highest intensity at roughly 80% of saturation. Generally 10 such accumulations were added together to get good statistics. The total time of 3 minutes to accumulate all the data was chosen because it was within the time where signs of radiation damage started to occur. The effect of radiation

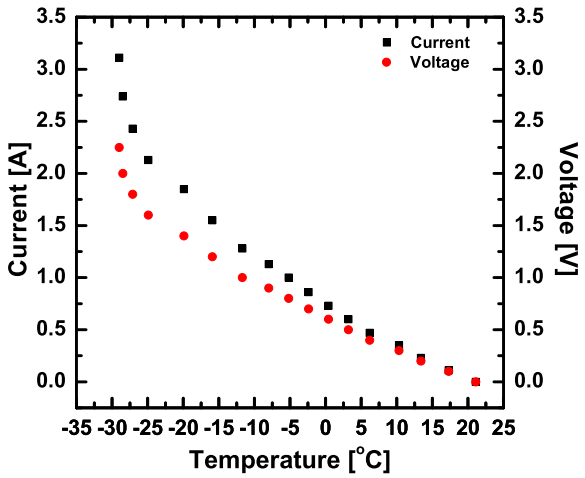


Figure 4.1: Typical Peltier cooler calibration of the voltage and current needed to reach a certain temperature.

damage on the SAXS pattern will be discussed in Section 4.4.2. The main signature of radiation damage is a reduction of the peaks of intensity with an increase of intensity at the minima, in other words, the features get washed out over time.

After the data waw collected at 10°C, the temperature was changed by increasing the DC current through the Peltier cooler. A new SAXS pattern at the new temperature was collected in the same way. The sample was translated by an amount slightly larger than the width of the beam between each temperature to expose a fresh part of the sample to avoid radiation damage. The process was repeated, incrementally reducing the temperature until the maximum power of the Peltier was reached and the lowest temperature possible was achieved. A typical calibration curve of the current and voltage needed to reach a certain temperature is shown on Figure 4.1. The highest temperature reached depends on the temperature of the water circulating through the copper heat sink, in this case 20°C. The temperature decreases linearly with current and voltage applied until the maximum cooling power is reached and increasing the electrical power does not reduce the temperature and the curves start to diverge.

Background images were measured with no sample in place *i.e* with only air scatter present, with an empty capillary and with a capillary containing only solvent. These could be used later for background subtraction on the data. It was however not straightforward to do because different capillaries have slightly different sizes which means they scatter differently. Furthermore, the translation of the sample for each temperature means that different parts of the capillary are illuminated and non uniformities in the capillary lead to different backgrounds and also different signal levels.

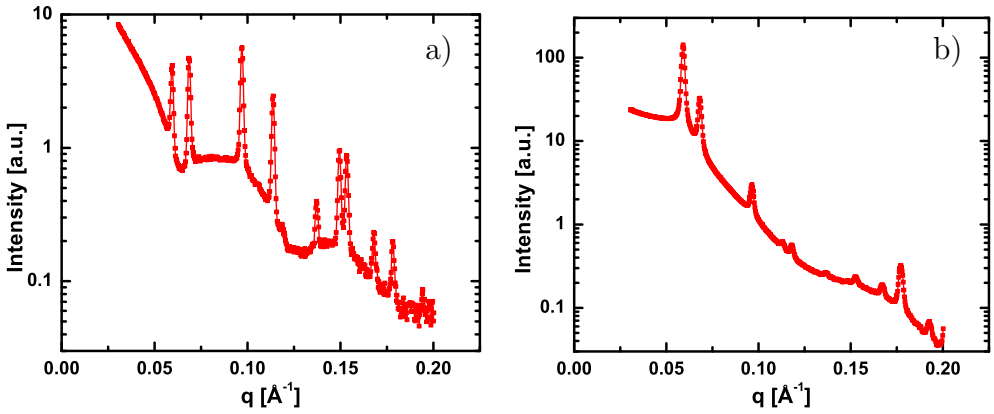


Figure 4.2: a) Integrated powder diffraction of apoferritin with added cadmium salt. b) Powder diffraction of holoferritin with the same amount of cadmium salt.

4.3 Data and Data Treatment

4.3.1 Calibration and Integration

The raw data obtained from the CCD detector was a 2D array of scattered intensities on an arbitrary scale. Typical examples of the raw data obtained at room temperature for both apoferritin and holoferritin were shown previously on Figure 3.14a) and b) respectively. Prior to measuring the samples of protein solution, a calibration measurement was performed. The calibrant was simply a powder sample of apoferritin as shown on Figure 3.15a). The lattice spacing of the small crystals is well-known and the many rings measured allowed for an excellent determination of the center of the pattern, as well as a calibration of the sample to detector distance. The only two other quantities needed to determine the value of $\mathbf{q}$ of each pixel was the pixel size and the wavelength of the radiation, which were known. The program Fit2D used to determine the center and the distance has an optimization routine to determine the center and the tilt of the detector with respect to the incident beam direction [27]. The distance was adjusted manually by looking at the q values of the Bragg rings and comparing with the expected values, until an accurate match was found.

This calibration needed to be redone every time any part of the experimental setup affecting the position of the detector with respect to the beam or the energy of the beam was changed. Generally, it was possible to keep everything fixed for long periods of time and the calibration was performed only at the beginning of the experimental run. With the calibration done, the samples could be measured and each 2D image could be integrated immediately while the next image was being measured.

The integration was performed using the Fit2D program [27]. The intensity of the SAXS pattern was spherically summed from the origin earlier determined during calibration. Figure 4.2 shows this integrated intensity for each of the two patterns of Figure 3.15. The dark noise background on the detector was generally weak enough compared with the signal that it could be considered as uniform and there was no need for a pixel by pixel background subtraction. The subtraction could be performed after the integration.

In some cases, entire sections of the data were bad for various reasons. For example, the beam may have been too low on the sample and parts of the scattered rays were blocked by the Peltier cooler, leading to a shadow on the bottom part of the chip. In other cases, the beam was hitting the top of the capillary and a sharp specular reflection ridge could be seen on the top part of the detector. It was possible in these case to ignore the features by integrating only the good data located within a desired arc. The software also allowed to mask small areas of bad pixels.

4.3.2 Normalization

Each integrated image shown on Figure 4.3a) was first scaled to an exposure time of 1 second to get rid of the differences arising from different exposure times as shown on Figure 4.3b). The intensity was then plotted on a log scale versus q^2. The slope of this curve close to the origin is related to the radius of gyration of the particles in solution according to Guinier's law of Equation 2.41. The results obtained on the radius of gyration in this way proved to be mostly meaningless since the SAXS patterns were generally not measured at low enough q for such an analysis. This is shown on Figure 4.3c). The next step was to plot Iq^4 versus q^4 as shown on Figure 4.3d). According to Porod's law of Equation 2.44, such a plot should be linear at high q with a slope corresponding to the background factor B for particle with sharp interfaces. A straight line was fitted and the value of the slope was then subtracted from the data as shown on Figure 4.3e). The fitted value of the intercept is a scale factor of the intensity which should correspond to the concentration of the solution. It is expected to be constant for multiple images of the same sample. It was however found to vary a little because each image was taken at a different location on the sample and therefore different volumes were illuminated by the beam. The data, after subtracting B was rescaled by dividing by the intercept A, which means every data set was normalized to the same number of protein molecules illuminated, getting rid of density fluctuations within the sample as shown on Figure 4.3f).

4.3.3 Form Factors and Concentration Dependence

These appropriately scaled curves were then divided by an accordingly scaled form factor. The form factor was measured on the unfrozen protein solution. The structure factor of the solution was obtained in this way. In the case of the unfrozen sample used as the form factor, the structure factor was assumed to be exactly equal to 1. Slight variations from unity were observed on samples at high concentrations, which is expected due to the close proximity of the molecules. The different intensity curves for different concentrations are shown on Figure 4.4. As the concentration increases, a peak starts to appear at low q indicating the presence of long wavelength density fluctuations for the liquid-like arrangement of the protein molecules. If the arrangement of the molecules is liquid-like, the limit of the peak position as the concentration increases will be $\frac{2\pi}{D}$ where D is the diameter of protein. Other types of ordering such as crystalline ordering are also possible which changes the peak position. The arrangements of the protein molecules in highly concentrated solutions giving rise to such peak was fitted to a paracrystalline model by Haußler *et al.* [40]. It should

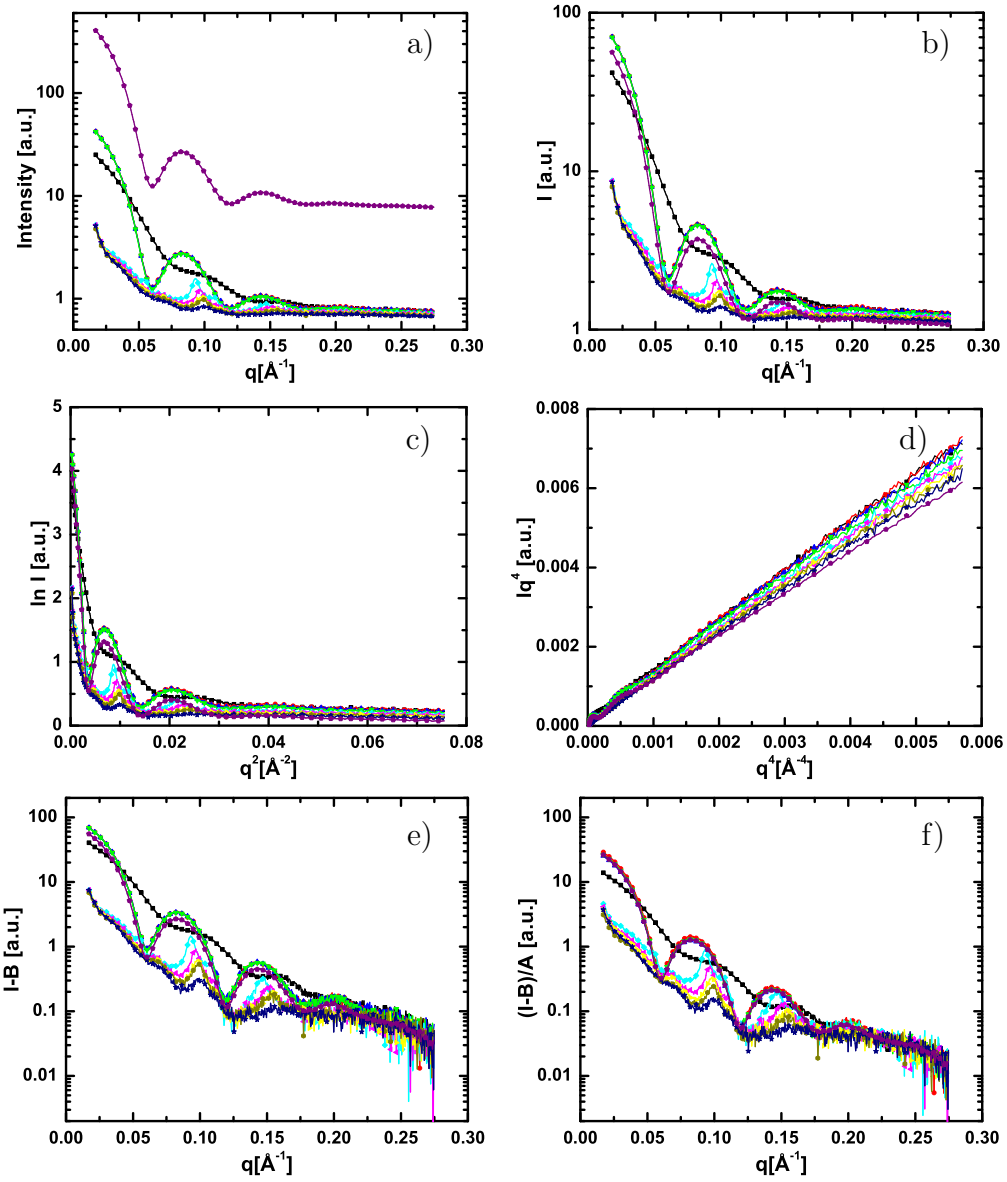


Figure 4.3: a) Raw integrated scattered counts versus q. b) Scattered intensity obtained by dividing each curve by the exposure time. c) Guinier plot : the low q part of the $\ln(I)$ versus q^2 should be linear with the slope related to the radius of gyration of the protein. d) Asymptotic curve of Iq^4 versus q^4. The slope gives a background value with the intercept yielding a scaling factor. e) Intensity curves with the fitted background subtracted. f) Intensity curves scaled with the fitted scale factor.

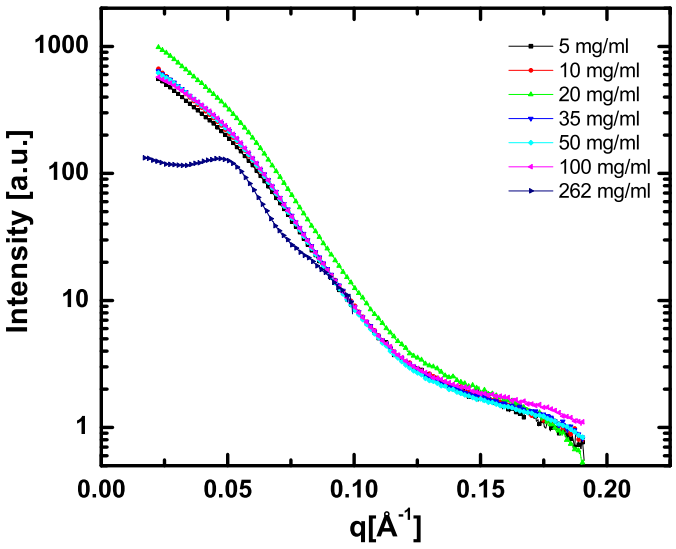


Figure 4.4: Scattered intensity from holoferritin and different concentrations. The higher concentration brings protein molecules into close proximity, which is seen in small angle x-ray scattering as a peak in intensity.

be pointed out that the division by the rotationally averaged form factor explicitly assumes that the molecules are identical and isotropic. This is very nearly the case in apoferritin but not true in holoferritin. In the latter case, the structure factor obtained is only a first order approximation.

4.3.4 Temperature Dependence

Figure 4.5 shows a typical SAXS pattern when the solution of holoferritin is cooled below the freezing point of the solution. This particular image was obtained on a solution at 100mg/ml with 150mM NaCl at a temperature of -20°C. The monotonic decrease in intensity with q is replaced by a broad peak at a q value near the {111} Bragg ring. The new peak, first seen at cryogenic temperatures as shown on Figure 3.24 is much broader than the powder diffraction peaks on Figure 3.15. In general, a peak in intensity indicates the presence of ordering. Cooling the sample therefore introduces new ordering in the sample.

The position of the peak is seen to be very temperature dependent. This is shown on Figure 4.6. The peak clearly shifts to higher q for lower temperatures. In a more subtle manner, the width of the peak also changes with temperature. It is reasonable to assume the peak is caused by some aggregation of the protein. A typical spacing can be obtained from the position of the peak and a typical size of the clusters formed can be deduced from the width of the peak using the Scherrer formula. This was done for all the cases and a typical example is shown on Figure 4.7. The spacing of the molecules in the clusters is seen to vary by almost 30% over a small range of 30°C. This is discussed further in Section 4.4.1.

The clusters were never present at temperature above the freezing point of the solution. A peak was only observed when the solution visibly went through a phase transition from

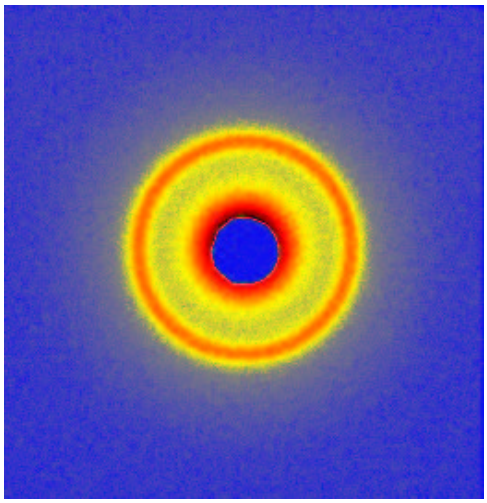


Figure 4.5: Raw data obtained with a CCD detector showing the small angle x-ray scattering of a solution of 100mg/ml holoferritin at -20°C. A broad ring appears when the solution is frozen.

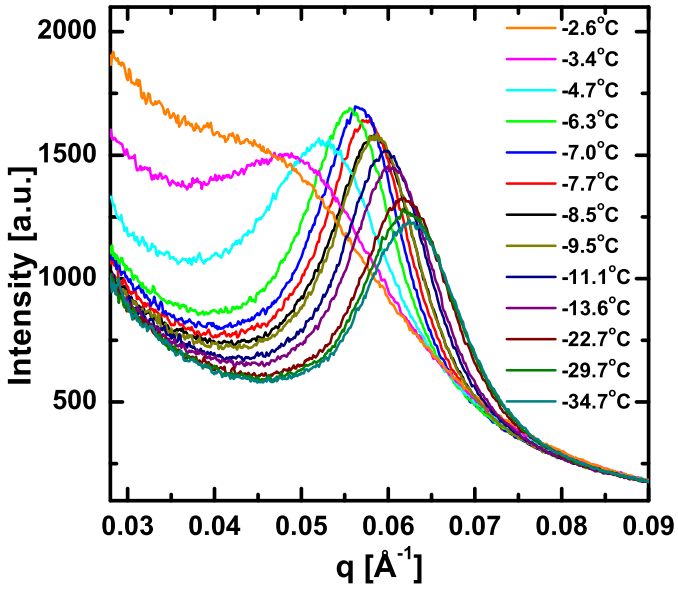


Figure 4.6: Temperature dependence of the SAXS intensity of a solution of 100mg/ml holoferritin. The 2D SAXS patterns were circularly integrated and plotted versus q.

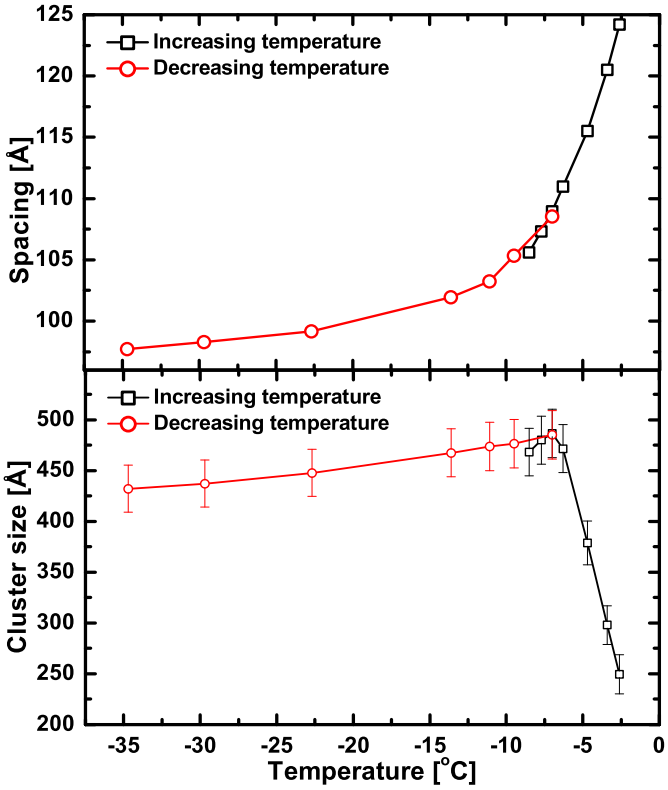


Figure 4.7: The lattice spacing and size of the clusters were obtained from position and width of the peaks in the structure factor shown on Figure 4.8. A 30% change in spacing is seen over a 30°C range.

a liquid state to a solidified state as judged by the appearance of opaqueness. The freezing point was dependent on the concentration of protein and salt in the solution. For 100mg/ml holoferritin in 150mM NaCl, the freezing point was found to be roughly -7.5°C. Therefore, while cooling the sample, no peak was observed until this temperature was reached. The sample was further cooled and changes in the peak position were seen. When the sample was heated, the melting point was found to be higher than the freezing point by a few degrees. The solution generally melted at -2°C, at which point no peak was observed. There was significant hysteresis observed and the direction of the change in temperature was very much relevant. The values of the cluster size and molecular spacing shown on Figure 4.7 were plotted using two different colors to emphasize whether the sample was being cooled or heated.

We explain the observed behavior by a phase separation of the sample into domains which are rich in protein and domains almost devoid of proteins. The sample can be cooled below the freezing point of water due to the high protein and salt content. Once it freezes, a network of ice crystals is formed. The protein molecules do no fit very well into the structure of ice and are pushed out to allow the formation of hydrogen bonds in the ice network. Voids are created in the ice and the protein molecules are trapped in them. As the temperature is increased, the protein content in the ice is actually very low due to the phase separation.

The melting point is then much closer to the actual melting point of pure water.

In the frozen solution, the protein molecules are effectively at a much higher concentration, trapped into small volumes. At such high concentration, the molecules are very close to each other giving rise to a peak. From the position of the peak however, we can conclude that it is not simply due to liquid-like ordering. Such a situation would yield a peak in intensity at a q value corresponding to a spacing equal or slightly larger than the size of the protein, *i.e.* 130Å. The presence of the peak at a higher q value indicates the presence of some level of crystalline ordering. In the fcc crystal of ferritin, the nearest neighbor distance is 130Å in the $\langle 110 \rangle$ directions but the $\{110\}$ reflections do not exist due to the exact cancellation in the unit cell. It is therefore clear from the data that there is some level of crystalline order present in the frozen samples. The position of the peak is generally near the expected $\{111\}$ peak of the fcc crystal suggesting a structure similar to fcc is present.

4.4 Structure Factors

To better understand quantitatively the observed behavior, it is usefully to examine the structure factors of the solutions measured rather than the intensity. As discussed in Chapter 2, the contribution to the scattered intensity from the electron density of the individual molecules, the form factor, can be factored out if the molecules are isotropic. What is left is the structure factor of the sample depending in principle solely on the arrangement of the molecules with no dependence on the actual shape of the molecules. The form factor was measured using a dilute solution of the protein studied, making the explicit assumption that at low concentration the structure factor was exactly equal to 1 and the scattered intensity was therefore the square of the structure factor. The structure factor obtained for both apoferritin and holoferritin powder crystalline samples shown on Figure 4.2 are shown on Figure 4.8.

Figure 4.8 clearly indicates that one must be careful in interpreting the structure factor obtained in this way. The structure factor for both proteins, which crystallize in the same structure should be identical. They however show some obvious differences. The ratio of the Bragg peaks are different for both proteins. Furthermore, some peaks seem to be completely absent. The reason for this is that the explicit assumption that the protein molecules in the crystals are isotropic and identical breaks down in the case of holoferritin due to the presence of the iron core. Direct calculation of the 3D form factor of apoferritin from the known atomic structure shown on Figure 3.5a) has shown it to be nearly isotropic and well described by a spherical shell. However, the iron cores from different protein molecules are all different and therefore the intensity is no longer factorisable. Also, it is likely that the cores are not isotropic. They are however not expected to have preferred orientation within the molecule even though the residues inside the shell are believed to play a role in the nucleation of the iron [19].

A strong cancellation in the structure factor for holoferritin is seen near q=0.12Å^{-1}. This value of q corresponds to a path length difference of the x-rays in the sample of ∼105Å. This value corresponds to a 25Å thick protein shell with an 80Å inner core. Since the ratio of the shell thickness to the inner core is roughly 0.25, it means that a quarter wavelength

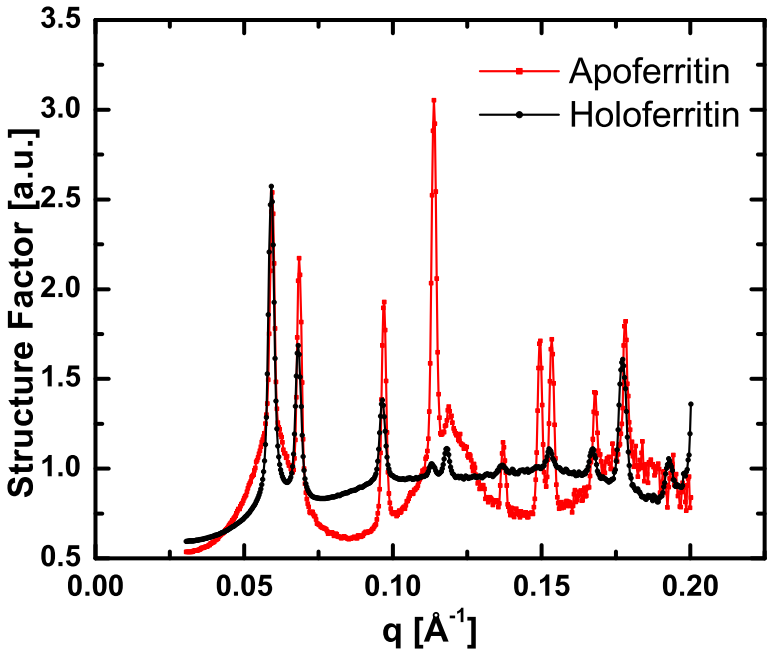


Figure 4.8: The structure factors calculated by dividing the intensities shown on Figure 4.2 by the measured form factors of apoferritin and holoferritin. Different peak heights are attributed to the fact that the form factor for holoferritin is ill-defined.

fits in the protein shell with 3 quarters in the iron core. The 2 parts therefore scatter out of phase. The same occurs for apoferritin. However, in holoferritin, the polydispersity of the iron core strongly influences the height of the peaks. So in summary, the form factor of the iron-loaded proteins is not known and cannot be properly measured, leading to wrong measured structure factors. However, as can be seen, the first 2 peaks of both proteins are fairly consistent. The analysis of the structure factors of holoferritin is therefore restricted to the low q part.

4.4.1 Holoferritin

The structure factors of the temperature series shown on Figure 4.6 are shown on Figure 4.9a). A second smaller peak not noticeable on the intensity curve is seen. The temperatures are shown in the order at which they were measured, with the first one at the top. The structure factor is seen to be unity at temperatures above the freezing point of the solution, even after freezing and thawing the solution. A large peak is present when the sample is frozen and this peak changes in height position and width with temperature. Figure 4.9b) shows the fitted lattice spacing in black and the fitted cluster size in red. The lines are guides to the eye.

The form factor in this case is a rapidly decreasing function of q. The product of a rapidly decreasing function and a sharply peaked function yields a curve with a sharp peak at a slightly shifted position. Therefore, in order to extract the position of the peaks to estimate the lattice spacing, one must use the structure factor and not the measured intensity. The

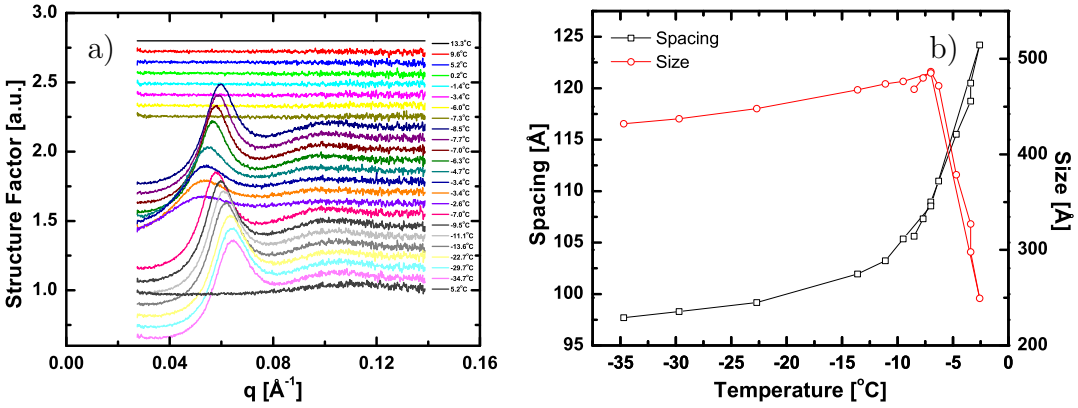


Figure 4.9: a) Structure factors of a solution of 100mg/ml holoferritin at different temperatures. They are shown in the order at which they were measured with the first one on top and the last at the bottom. b) Fitted lattice spacing (black) and fitted cluster size (red) obtained from the position and width of the first peak respectively.

first peak in the structure factor was fitted to a gaussian. The lattice spacing is estimated as $\frac{2\pi}{q_{\text{peak}}}$ and the Scherrer formula Size $= \frac{2\pi}{\text{FWHM}}$ is used to estimate the average size of the clusters in the sample. These results were shown on Figure 4.7 for this particular case.

4.4.2 Protein Concentration Dependence

The effect of protein concentration on the observed formation of an ordered state at low temperature was studied. The holoferritin solution purchased was concentrated using centrifugal filtering to 200mg/ml. All the samples discussed in this section were in a 150mM NaCl solution. The structure factors measured at 200mg/ml are shown on Figure 4.10a). A similar behavior to the 100mg/ml sample is observed with a slight difference in the value of the freezing point due to the higher solute concentration. A peak in the structure factor appears when the solution freezes and this peak changes with temperature.

Figure 4.10b) shows the fitted spacing and cluster size. The behavior of the two parameters is again similar. The spacing is slightly smaller for the 200mg/ml sample. This can be explained by the phase separation discussed above. There is less volume available to the protein molecules at higher concentration and they are forced to pack more tightly. The size of the clusters formed is comparable to the 100mg/ml case discussed above.

Performing the experiment again on a solution at 100mg/ml reveals a slightly different behavior as shown on Figure 4.11. In this case, a single broad peak is formed upon cooling. However, as the temperature is increased, this peak is seen to split into two narrower peaks. This indicates the formation of larger clusters as the temperature is increased. As simulations will show in Section 4.5, the single broad peak is in fact the combination of the first two peaks of the face-centered cubic lattice and as the size of the cluster is increased, they get narrower at become distinct peaks. Therefore, the fitted lattice spacing is fixed at a value near the expected lattice spacing of 106Å as the temperature is increased.

The size of the clusters increases with temperature up to a value of roughly 1000Å

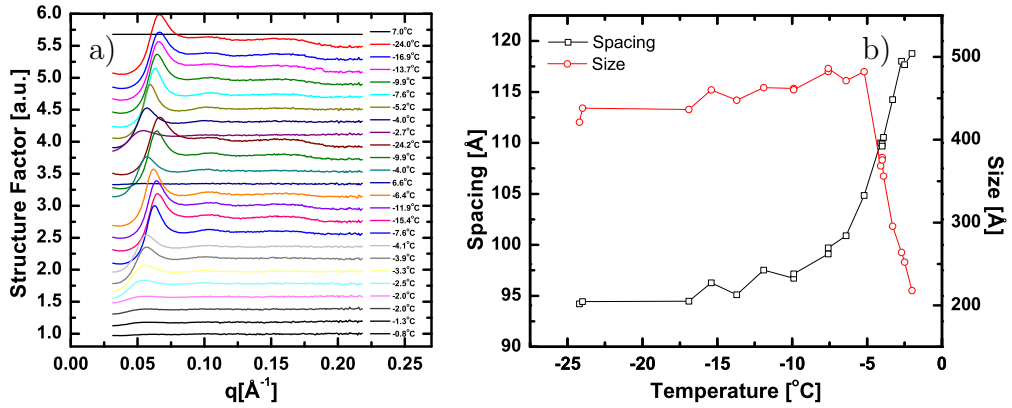


Figure 4.10: a) Structure factors of a solution of 200mg/ml holoferritin at different temperatures. They are shown in the order at which they were measured with the first one on top and the last at the bottom. b) Fitted lattice spacing (black) and fitted cluster size (red) obtained from the position and width of the first peak respectively.

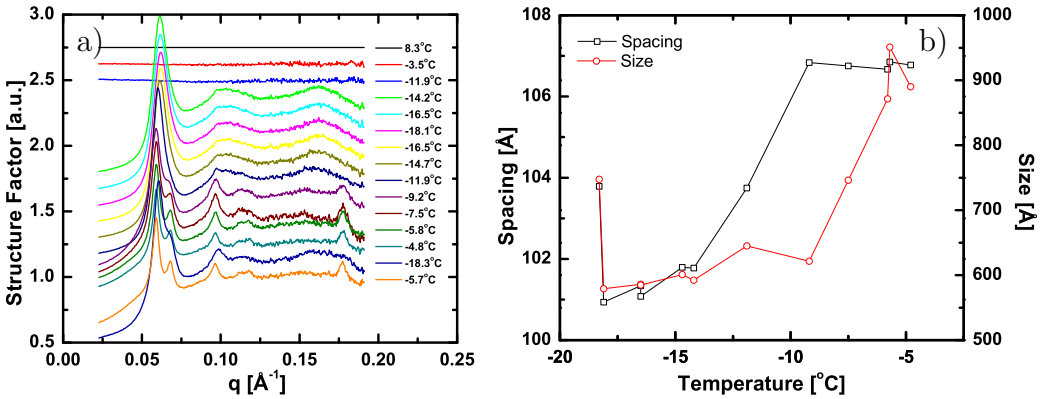


Figure 4.11: a) Structure factors of a solution of 100mg/ml holoferritin at different temperatures. They are shown in the order at which they were measured with the first one on top and the last at the bottom. b) Fitted lattice spacing (black) and fitted cluster size (red) obtained from the position and width of the first peak respectively.

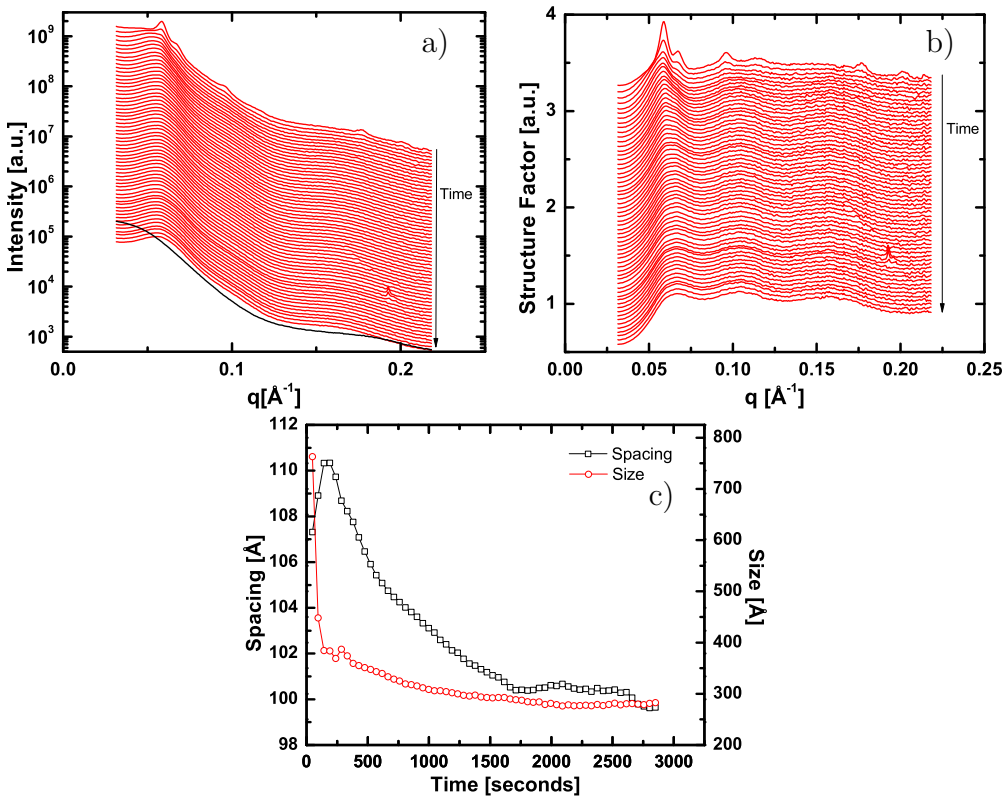


Figure 4.12: Effects of radiation on small clusters of holoferritin. a) Intensity versus time for a 100mg/ml holoferritin sample at -5°C. Each curve is taken 47.5 seconds apart. The black curve is the unfrozen form factor. b) Structure factor versus time. c) Fitted spacing (black) and size (red) versus time.

compared with 500Å at 200mg/ml. These are very small clusters considering the size of the protein. A cluster 500Å across contains only about 4 proteins in 1 direction for a total of roughly 64 molecules. The difference between the two cases at 100mg/ml shown could be real since the concentration of one of the two samples could have been slightly wrong. However it is most likely due to radiation damage to the sample. Most likely, not enough care was taken in preventing damage to the sample by translating it regularly for the case shown of Figure 4.9. As can be seen on Figure 4.12, the effect of radiation is to transform over a short period of time a structure factor showing two peaks and even a sharp third peak at higher q into a single broad peak. In other words, radiation transforms the data shown on Figure 4.11 into that of Figure 4.9.

The fitted spacing and size of the clusters changes significantly over time, *i.e* with dose. Clusters shrink from 800Å to 300Å. The fitted lattice spacing shrinks with radiation after a short period of increase. This is not what is observed when comparing Figures 4.11 and 4.9. Therefore radiation damage cannot explain the large thermal expansion observed. It can however explain the presence of two peaks in one case and a single peak in another case. It needs to be pointed out that great care needs to be taken in interpreting the results for the fitted lattice spacing and cluster size due to the fact that often two peaks are so broad

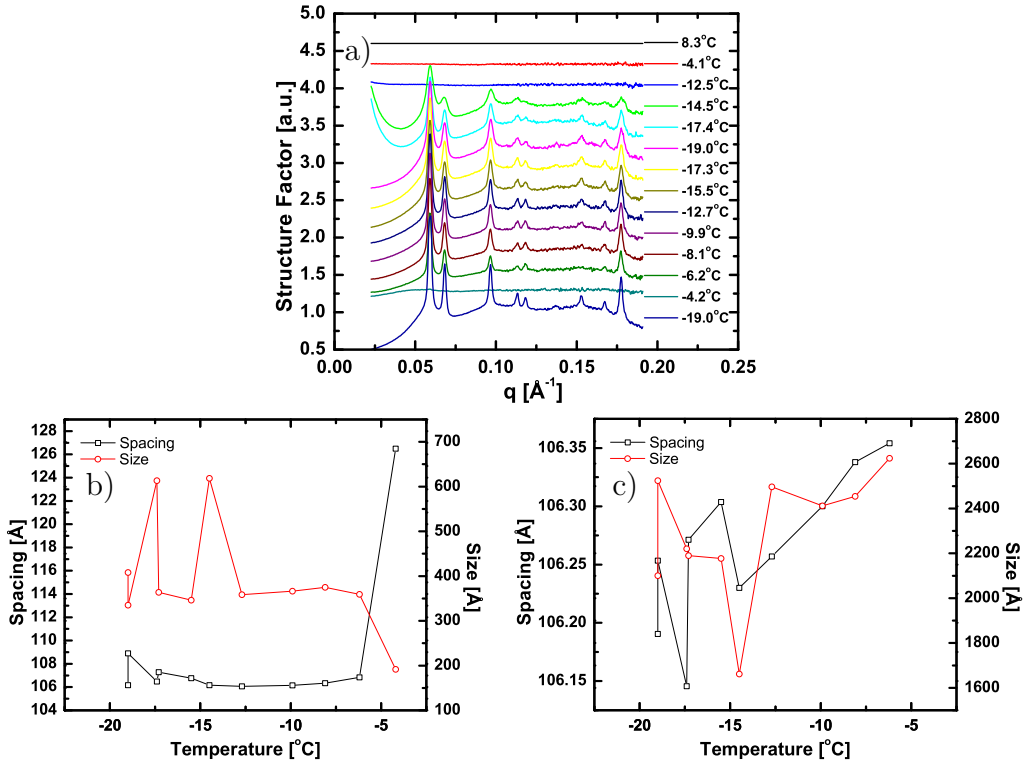


Figure 4.13: a) Structure factors of a solution of 50mg/ml holoferritin at different temperatures. They are shown in the order at which they were measured with the first one on top and the last at the bottom. b) Fitted lattice spacing (black) and fitted cluster size (red) obtained from the position and width of the first broad peak respectively. c) Fitted lattice spacing (black) and fitted cluster size (red) obtained from the position and width of the first sharp peak respectively.

that they look like a single peak. In such a case, the relation between the position and the spacing as well as that between the width and the size are no longer valid. This will be made clearer in Section 4.5.

The experiment was repeated on a sample of 50mg/ml holoferritin with again 150mM NaCl. The results are shown on Figure 4.13. As seen on the sample at 100mg/ml, there is a formation of small clusters giving rise to broad peaks when the solution is frozen. Figure 4.13a) shows that as the temperature is increased, these peaks get narrower indicating the clusters are growing in size. There is however now two clear distinguishable features at q values near 0.06Å^{-1}. There are 2 sharp peaks and a third much broader peak similar to that seen at 200mg/ml. Figure 4.13b) shows the fitted lattice spacing and cluster size obtained from the broad peak. This peak corresponds to loose small clusters of size roughly 400Å displaying a similar thermal expansion as observed before. The fitted spacing and size from the first sharp peak is shown on Figure 4.13c). The spacing is very constant and corresponds to the {111} spacing of fcc ferritin crystals. The crystal size is now much larger, between 1600 and 2600Å. The crystals grow larger with temperature, but there is still a significant amount of much smaller clusters with a spacing different from the crystalline spacing.

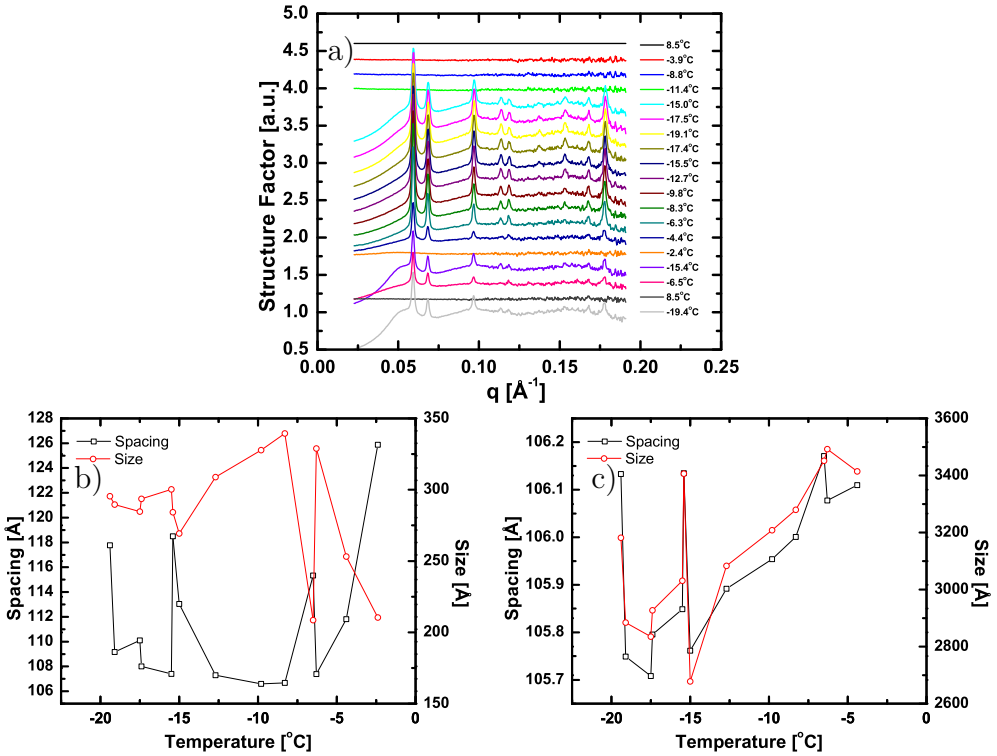


Figure 4.14: a) Structure factors of a solution of 35mg/ml holoferritin at different temperatures. They are shown in the order at which they were measured with the first one on top and the last at the bottom. b) Fitted lattice spacing (black) and fitted cluster size (red) obtained from the position and width of the first broad peak respectively. c) Fitted lattice spacing (black) and fitted cluster size (red) obtained from the position and width of the first sharp peak respectively.

The data obtained on samples containing 35mg/ml, 20mg/ml, 10mg/ml and 5mg/ml of holoferritin in 150mM NaCl is shown on Figures 4.14, 4.15,4.16 and 4.17 respectively. A consistent trend can be observed. First of all, the spacing corresponding to the sharp crystalline peaks is always near the expected value of the 105.6Å obtained from the PDB structure. The slight variations can be due to real thermal expansion or also possibly to radiation damage which is known to change the lattice spacing.

4.4.3 Interpretation

The lower protein concentration appears to be more favorable to the formation of large crystals when the temperature is lowered below the freezing point. At the lower concentrations, there is a coexistence of two phases with large micron-size crystals as well as much smaller protein clusters.

As the concentration is reduced, the number of protein molecules caged in a certain volume after phase separation gets smaller. It is then not surprising that the spacing corresponding to the broad peak gets larger since there is less crowding. The fitted size of the clusters remains fairly constant for all concentrations at roughly 300Å. For the lower con-

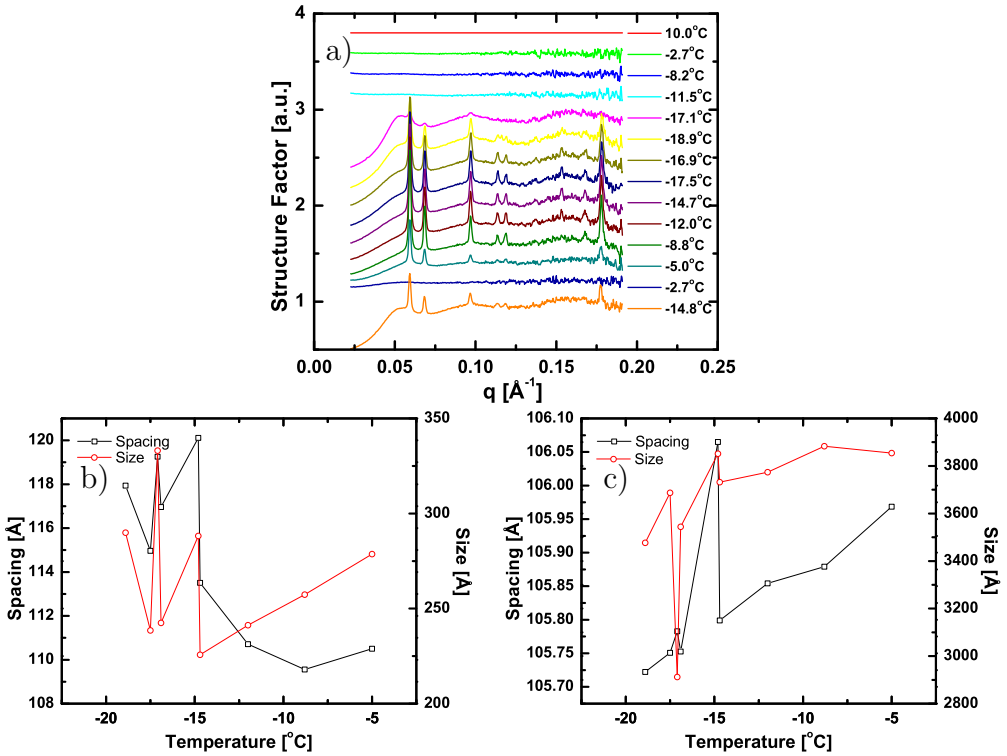


Figure 4.15: a) Structure factors of a solution of 20mg/ml holoferritin at different temperatures. They are shown in the order at which they were measured with the first one on top and the last at the bottom. b) Fitted lattice spacing (black) and fitted cluster size (red) obtained from the position and width of the first broad peak respectively. c) Fitted lattice spacing (black) and fitted cluster size (red) obtained from the position and width of the first sharp peak respectively.

centrations, the spacing starts to approach the size of the molecules at 130Å. The packing therefore resembles more a liquid-like structure than small crystalline clusters. This will be discussed further in Section 4.5.

The size of the crystals formed for all concentrations is seen to increase with temperature but also increase with decreasing concentration. This can be explained again by the trapping of proteins into voids in the ice after phase separation has occurred. After rapid cooling, the molecules are originally trapped into a certain conformation. At high concentration, there is a lot of crowding and the molecules cannot reorganize at any temperature before melting occurs. However, at lower concentrations, an increase in temperature provides enough energy to the system and increases the available space to allow the protein molecules to reorganize in such a way as to produce larger crystals. This is similar to an annealing process. The molecules have some room to move around and are not trapped in there original configuration. Therefore, the lower the concentration, the easier it is for large crystals to form, to the point that they are even seen immediately upon freezing. At lower concentration, the number of crystals produced is much smaller and therefore the height of the peaks is smaller. The smoother oscillations in the structure factor from the small

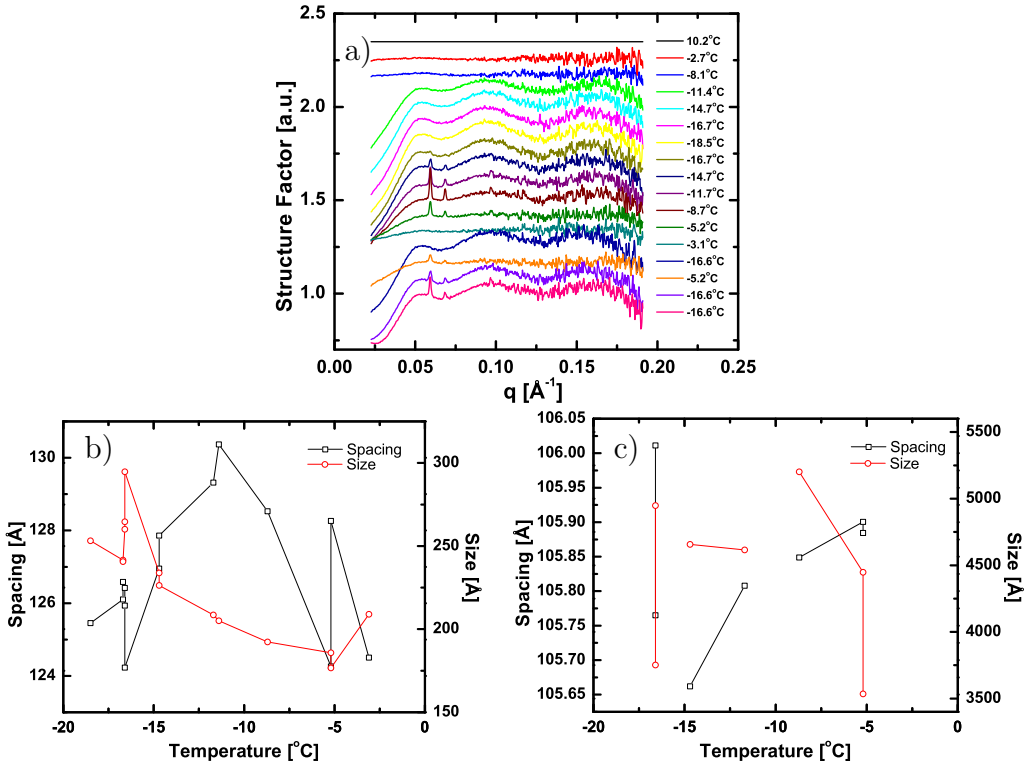


Figure 4.16: a) Structure factors of a solution of 10mg/ml holoferritin at different temperatures. They are shown in the order at which they were measured with the first one on top and the last at the bottom. b) Fitted lattice spacing (black) and fitted cluster size (red) obtained from the position and width of the first broad peak respectively. c) Fitted lattice spacing (black) and fitted cluster size (red) obtained from the position and width of the first sharp peak respectively.

clusters or possibly liquid-like ordering become more significant compared with the crystal peaks.

To explain the observed lattice spacing changes in the small clusters at the higher concentrations measured, it is first necessary to understand the actual structures seen. We therefore discussed this topic below in Section 4.7.

4.4.4 Ionic Strength Dependence

Proteins is solution are generally charged and therefore tend to interact strongly. The presence of counterions, in this case Na^+ and Cl^- screen these interactions to a certain extent. It was shown that for ferritin molecules is solution at room temperature, the concentration of NaCl greatly affects the SAXS intensity curve [40]. A peak in the structure factor appears as the ionic strength is decreased and this peak gets stronger and sharper with decreasing NaCl concentration. The authors attribute this behavior to the fact that interactions are stronger at lower ionic strength, giving rise to a more structured solution with a stronger, more well-defined structure factor.

In order to gain better insight into the mechanisms involved in the formation of small

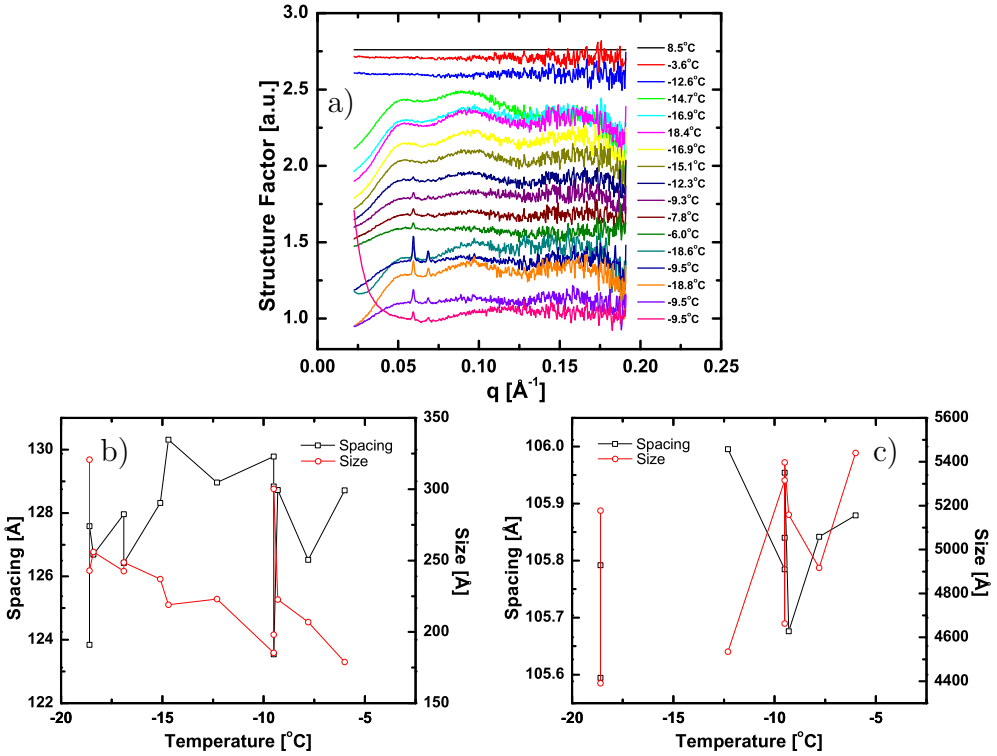


Figure 4.17: a) Structure factors of a solution of 5mg/ml holoferritin at different temperatures. They are shown in the order at which they were measured with the first one on top and the last at the bottom. b) Fitted lattice spacing (black) and fitted cluster size (red) obtained from the position and width of the first broad peak respectively. c) Fitted lattice spacing (black) and fitted cluster size (red) obtained from the position and width of the first sharp peak respectively.

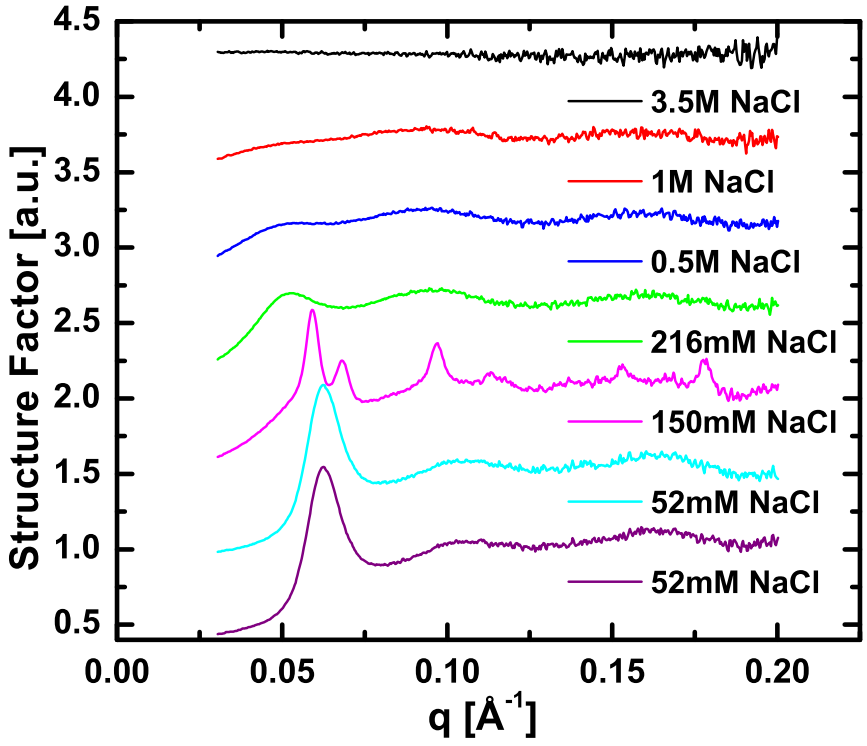


Figure 4.18: Structure factor for holoferritin at 35mg/ml with varying concentration of NaCl at -20°C.

clusters of ferritin at low temperature, solutions of holoferritin at 35mg/ml with different NaCl concentrations were prepared. Samples containing 3.5M, 1M, 0.5M, 216mM, 150mM and 52mM NaCl were prepared by diluting a 100mg/ml stock protein solution with the appropriate salt solution. Each solution was cooled to -20°C and the structure factor was calculated in the same manner as described before. The results are shown on Figure 4.18.

The solution with 3.5M of NaCl did not freeze in the temperature range accessible to the experiment. This is because of the high solute concentration which significantly depresses the freezing point. Therefore, the structure factor measured is equal to unity, with some noise at the high q end. Every other solution did eventually freeze before -20°C was reached, although they all froze at a different temperature. The behavior at 150mM NaCl is the same as observed before since the protein concentration data were collected with samples containing 150mM of NaCl. At high ionic strengths, the structure factor displays characteristics consistent with a liquid-like ordering of the protein molecules with a first peak near the value of $2\pi/D$ where D is the diameter of the protein. This could be due to more screening of the charges on the protein molecules and therefore weaker interactions not sufficient to produce ordered structures such as crystals.

At lower ionic strength, the two peaks characteristic of a larger crystalline cluster merge into a single broad peak at a position roughly half way between the two sharper peaks. This is an indication of crystalline ordering with smaller overall size. The stronger interactions between the molecules at lower ionic strength lead to more rapid crystallization yielding

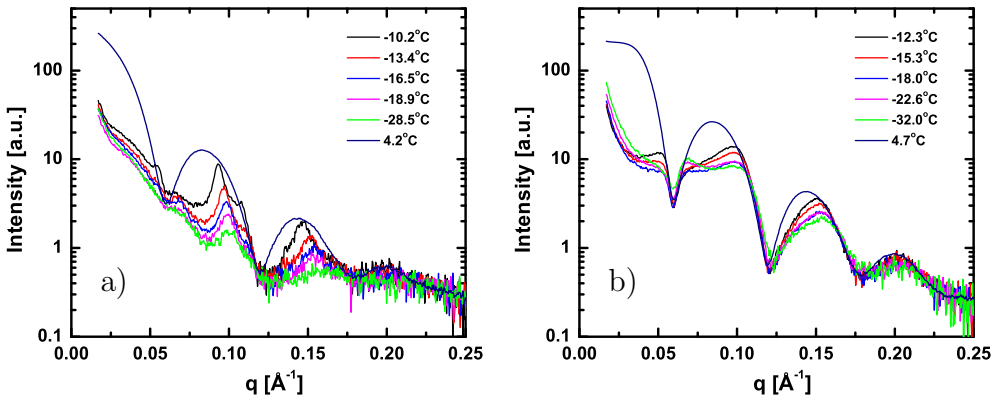


Figure 4.19: a) SAXS intensity from a 50mg/ml solution of apoferritin containing 100mM NaCl at various temperature. b) SAXS intensity for multiple temperature for a solution of 150mg/ml apoferritin.

smaller crystals. It is well-known in protein crystallography that if the interactions are stronger, the crystals produced get smaller [71].

4.4.5 Apoferritin

Similar experiments were performed using solutions of the empty protein shell apoferritin. The signal from these samples was significantly weaker and made measurements of low concentrations difficult. Figure 4.19 shows the intensity measured from samples of apoferritin containing 50mg/ml and 150mg/ml with 100mM NaCl. The most intense curve in each case corresponds to the unfrozen sample. Both have Bessel function like oscillations due to the spherical shape of the protein molecules. The form factor of Equation 2.22 with a low level of polydispersity describes the intensity fairly well except at low q for the more concentrated solution. Comparing the two concentrations, one can see the appearance of a slight peak at roughly q=0.035Å at 150mg/ml. This peak could be attributed to liquid-like ordering but careful studies indicate the presence of paracrystalline ordering in solution [40].

As in the case of holoferritin solutions, as the temperature is decreased, the solution eventually freezes and a distinct change to the SAXS pattern can be seen. At both 50 and 150mg/ml, the intensity near the origin drops and the oscillations from the form factor of the protein are distorted and even canceled off. The decrease in intensity at low q is indicative of aggregation. The total scattered intensity from the sample remains the same since the total number of electrons illuminated remains the same. However, cluster formation redistributes the intensity to different values of q. The maximum at the origin gets narrower as the cluster size increases and becomes a delta function in the limit of an infinite crystal. New peaks in the intensity appear where peaks in the structure factor are present.

When compared with the case of holoferritin, the apoferritin intensity curves seem like a very complicated function with no clearly identifiable peak which one could interpret as a typical spacing as was done in Section 4.4.1. There is however a great advantage in using the empty protein shells. They are very monodisperse and very closely isotropic. Therefore,

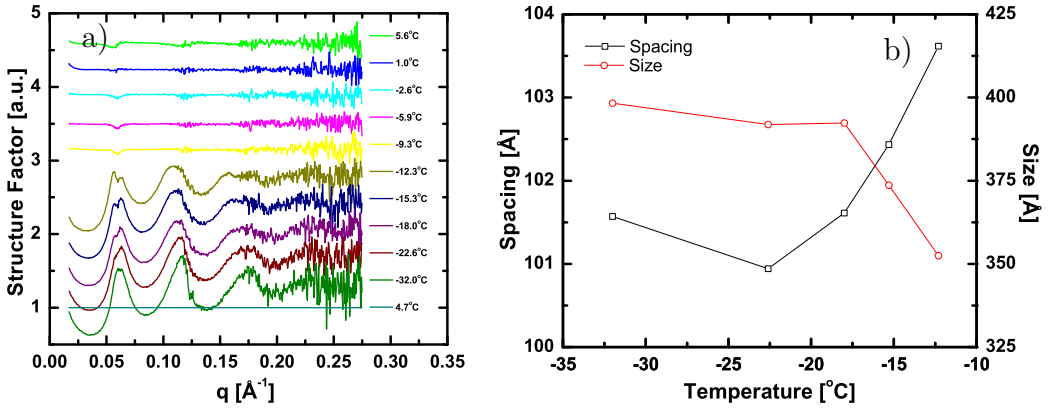


Figure 4.20: a) Structure factors of a solution of 150mg/ml apoferritin at different temperatures. They are shown in the order at which they were measured with the first one on top and the last at the bottom. b) Fitted lattice spacing (black) and fitted cluster size (red) obtained from the position and width of the first peak respectively.

the structure factor obtained is much more meaningful than in the iron-loaded protein case. In the latter, the structure factor measured was only a first order approximation valid over a small q range. In the apoferritin case, the entire structure factor curve obtained should be valid.

The structure factors for both cases shown on Figure 4.19 are shown on Figures 4.20 and 4.21 for 150mg/ml and 50mg/ml respectively. The complicated intensity curves turn into smooth curves when divided by the proper form factor. Peaks in the structure factor are again indicative of some form of ordering. The presence of multiple peaks indicates a good level of long-range order.

At 150mg/ml, the solution did not freeze until a temperature lower than -9°C was reached. The first 3 peaks in the structure factor are clearly visible. One thing to notice is the slight splitting of the first peak. The interesting aspect of it is the fact that the lower q part of the peak has a lower intensity than the higher q part. The powder crystalline data measured shown on Figure 4.8 clearly shows that the second peak, the {200} peak of fcc ferritin crystals is weaker than the first peak, the {111} peak.

The position and width of the first peak in the structure factor was used, as before in the case of holoferritin to estimate a molecular spacing as well as a cluster size. The results on Figure 4.20b) show a spacing of roughly 101Å increasing with temperature. The fitted size of the ordered region is near 400Å and decreases sharply when the temperature approaches the melting point. Both of these results are consistent with the results above for holoferritin.

The structure factors for the solution at 50mg/ml display more features. Most importantly, the second broad peak at q ~0.12Å^{-1} is split, with a sharp peak appearing at q ~0.1Å^{-1}. The position, height and width of this peak changes significantly with temperature, moving to lower q with increasing temperature. Another small peak at q ~0.15Å^{-1} is seen for a few of the measured temperatures.

The fitted spacing and cluster size are shown on Figure 4.21b). Results from another sample with the same protein and salt concentrations is shown on Figure 4.22 to show

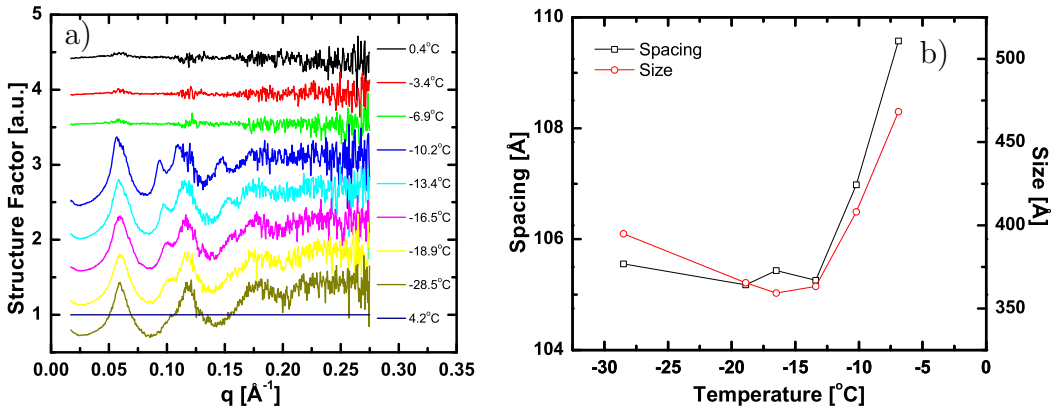


Figure 4.21: a) Structure factors of a solution of 50mg/ml apoferritin at different temperatures. They are shown in the order at which they were measured with the first one on top and the last at the bottom. b) Fitted lattice spacing (black) and fitted cluster size (red) obtained from the position and width of the first peak respectively.

that the results are somewhat reproducible, but not perfectly. These two samples came from different batches purchased at different times. Biological samples are known to vary from batch to batch and results often cannot be duplicated exactly. The lattice spacing at 50mg/ml shown on Figure 4.22b) shows the same large change with temperature as holoferritin. The fitted size of the clusters is again in the 400-500Å range corresponding to clusters of less than 100 molecules.

The thermal expansion measured is compared on Figure 4.23 to that of holoferritin previously shown on Figure 4.7. The holoferritin sample was at 100mg/ml with 150mM NaCl while the apoferritin sample was at 50mg/ml with 100mM NaCl. The difference in mass concentration is due to the heavier protein with iron. If the molar concentration of the protein is used instead of the mass per volume, there is a less the 30% difference between the two samples. Therefore, the results are expected to be very similar unless the presence of the iron core plays an important role in the aggregation mechanism. This is not the case as the two proteins show similar behavior. The increase in the molecular spacing follows the same curve for both cases, with the absolute values differing only because of the slight difference in the molar concentration. The holoferritin sample is slightly more concentrated giving rise to a smaller spacing. As a reference, a dashed line was drawn at a value corresponding to the spacing of the {111} planes in the fcc crystal of ferritin. The spacing measured is smaller than the crystal spacing, but as mentioned before, this is due partly to the merging of the first two Bragg peaks into one broader peak.

Results for a solution of apoferritin at 25mg/ml are shown on Figure 4.24. The first thing that is noticeable is that the data is much more noisy. This is due to the smaller fraction of the sample actually containing proteins. A larger proportion of scattering comes from the solvent leading to noisier data. Nevertheless, the familiar features are present, with a few broad peaks which can be identified as coming from small clusters and a few sharper peaks, generally attributed to the presence of larger well-ordered structures. Figure 4.24b) again shows the changes in the fitted spacing and cluster size with temperature. A very

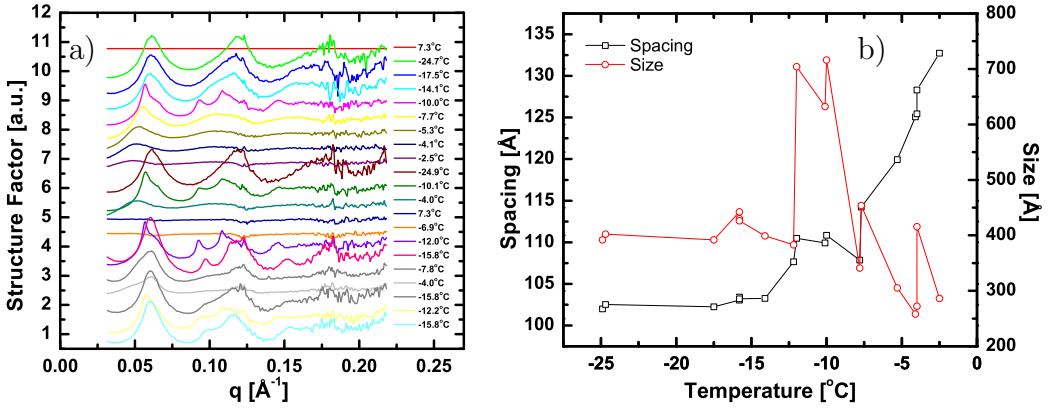


Figure 4.22: a) Structure factors of a solution of 50mg/ml apoferritin at different temperatures. They are shown in the order at which they were measured with the first one on top and the last at the bottom. b) Fitted lattice spacing (black) and fitted cluster size (red) obtained from the position and width of the first peak respectively.

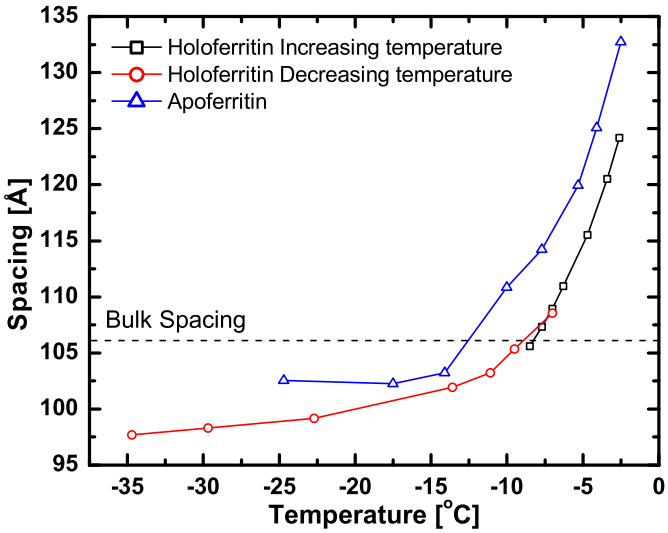


Figure 4.23: Thermal expansion of the fitted spacing of apoferritin small clusters formed at low temperature from a solution of 50mg/ml compared with the thermal expansion of from similar clusters of holoferritin formed from a solution of 100mg/ml. The two proteins follow similar curves. The dashed line indicates the ideal bulk spacing of a large fcc ferritin crystal.

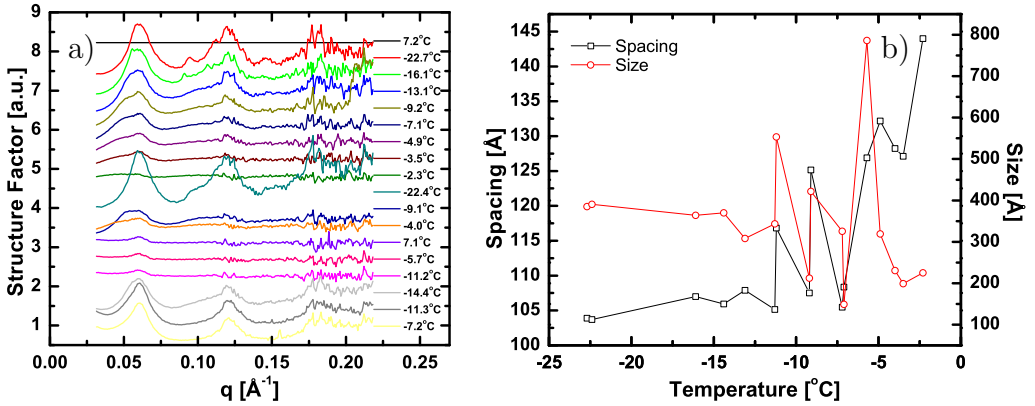


Figure 4.24: a) Structure factors of a solution of 25mg/ml apoferritin at different temperatures. They are shown in the order at which they were measured with the first one on top and the last at the bottom. b) Fitted lattice spacing (black) and fitted cluster size (red) obtained from the position and width of the first peak respectively.

large change in the spacing is observed with the cluster size remaining mostly constant. The large spacing measured is most likely due to a change in the structure of the clusters as the temperature is increased. The frozen clusters clearly show crystalline ordering while as the temperature approaches the melting point, the clusters appear to break up a form a more liquid-like structure at high concentration.

The experiment was performed at lower protein concentrations and the results are shown for a 10mg/ml solution on Figure 4.25. At such low concentration, the signal level was very low and the fine features of the structure factor, if any were present, may not have been measured accurately. Nevertheless, oscillations are still present in the structure factor. The pattern is almost independent of temperature as shown on Figure 4.25b). The spacing and the size of the ordered domains remain mostly constant. Such a structure factor, with only periodic oscillations and no characteristic features such as the sharp peaks of Figure 4.21 or a splitting of the first maximum can be fitted using multiple models. As will be discussed in Section 4.5, simple liquid-like adhesive hard sphere models, paracrystalline models and small ordered clusters give similar structure factors. It is therefore difficult to draw any conclusion from the weakly scattering low concentration apoferritin samples.

4.5 Simulations

The position of the peaks in the structure factors of ferritin solutions at low temperature suggests that there is formation of small aggregates with crystalline order. Furthermore, the width of the peaks indicate that the overall size of the aggregates corresponds to a total number of protein molecules smaller than 100 in many cases. This small number makes it possible to directly calculate the structure factor arising from clusters of a known structure.

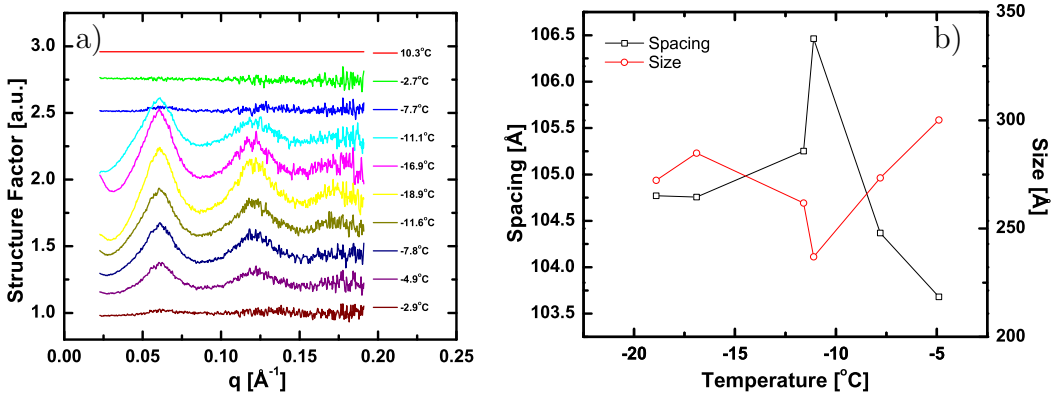


Figure 4.25: a) Structure factors of a solution of 10mg/ml apoferritin at different temperatures. They are shown in the order at which they were measured with the first one on top and the last at the bottom. b) Fitted lattice spacing (black) and fitted cluster size (red) obtained from the position and width of the first peak respectively.

4.5.1 Isotropic Clusters

It is logical to start by assuming the clusters formed possess a face-centered cubic structure since this is the most easily formed crystal structure of the protein. As a first guess of the structure of the clusters, one might assume a roughly spherical aggregation, with a central molecule to which new molecules attach forming an isotropic structure. This is shown on Figure 4.26. Supermolecular clusters were built up starting with a central molecule located at the origin. Molecules were added one by one on fcc lattice sites. The structure of the dimer is unique. However, the trimer structure and most other oligomeric structure is degenerate. Even if one limits the number of possible locations where to add a new molecule to the nearest neighbor shell so as to keep the overall supermolecule roughly spherical, there are still multiple possibilities. The first neighbors from the central molecule in the fcc structure are all the $\langle 011 \rangle$ lattice sites. There are 12 of these. It does not matter which one of the 12 is chosen first since a simple rotation transforms any cluster formed into any other one. However, different structures are formed depending on which of the remaining $\langle 011 \rangle$ sites is chosen next. In general, all supermolecules are degenerate. However, clusters containing only filled shells of nearest neighbors are unique. If all the 12 $\langle 011 \rangle$ sites are occupied, the supermolecules containing 13 ferritin molecules is unique. The same goes for all other filled shells. It is those filled shells that are shown on Figure 4.26.

The next closest lattice points in the fcc structure are the 6 $\langle 002 \rangle$ sites. Continuing this process to build up supermolecules, one gets that the number of protein molecules in filled-shell structures shown on Figure 4.26 are 13, 19, 43, 55, 79, 87, 135, 141, 153, 177 and 201 corresponding to lattice sites filled from the origin up to the $\langle 110 \rangle$, $\langle 002 \rangle$, $\langle 112 \rangle$, $\langle 022 \rangle$, $\langle 013 \rangle$, $\langle 222 \rangle$, $\langle 123 \rangle$, $\langle 004 \rangle$, $\langle 033 \rangle$, $\langle 114 \rangle$ and $\langle 024 \rangle$. Some lattice points, although different such as the $\langle 033 \rangle$ and $\langle 114 \rangle$ actually are at the same distance from the origin. They however have different coordination and different binding energies.

The structure factor of such supermolecules was calculated using the Debye formula of Equation 2.30. The structure factors calculated for all the filled shells up to 767 molecules

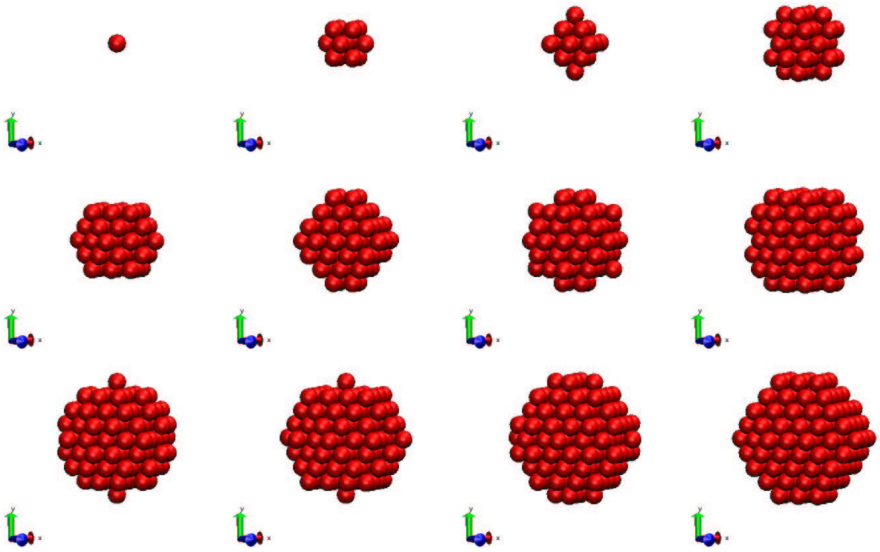


Figure 4.26: Supermolecules of ferritin. A central molecule is located at the origin. Shells of nearest neighbors are added keeping the structure isotropic and roughly spherical. The number of molecules in each structure shown is 1, 13, 19, 43, 55, 79, 87, 135, 141, 153, 177 and 201.

are shown on Figure 4.27. The structure factor was also calculated for each number of molecule between the filled shell numbers. Only one structure per number of molecule was calculated. As the number of proteins increases, the difference in the structure factor from degenerate structures becomes smaller and smaller.

The structure factor from the monomer is exactly unity. The structure factor from the dimer is the sinc function. As the cluster size grows, broad peaks get more well-defined. These peaks are at q values near the fcc Bragg peaks. Focusing our attention on the first peak at q $\sim$0.06Å^{-1}, one notices that as the number of proteins is increased, this peak gets narrower, then splits into two distinct peaks that also get progressively narrower. This is because there are two Bragg peaks near q=0.06Å^{-1}. The {111} Bragg peak is at q=0.0591Å^{-1} and the {002} peak is at q=0.0686Å^{-1}. For a small cluster of 13 molecules, the diameter of the cluster is 390Å corresponding to a width of the diffraction pattern of 0.016Å^{-1}. The two peaks are then indistinguishable. The same thing occurs at other q values. As the crystal grows, Bragg peaks become better defined, until they essentially become delta functions in the infinite limit.

The simulations shown on Figure 4.27 show a good similarity with the structure factor data obtained for holoferritin at low q and also for apoferritin over the whole q range measured. The first peak was often measured to split with increasing temperature as shown on Figure 4.11. The very rapid oscillations in the very low q part of the simulation arises from the exact shape of the clusters. This part is very sensitive to the exact structure.Iit is not expected that every cluster in the sample will be identical. Instead, a weighted sum over many of the simulated clusters is a better description of the system. Such a summation

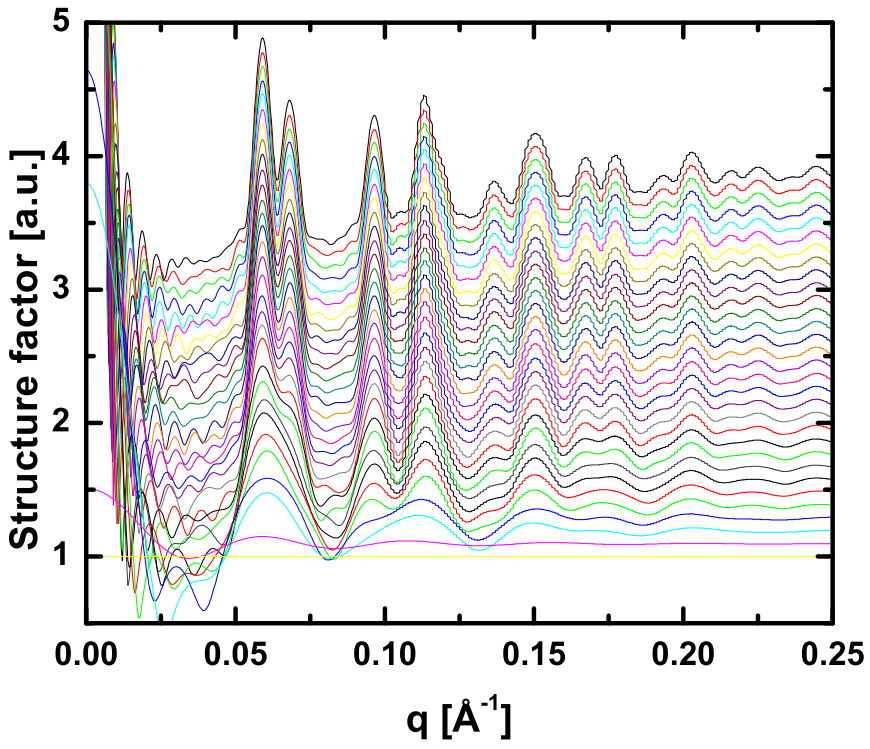


Figure 4.27: Orientationally averaged structure factors simulated from filled shell fcc supermolecules. Each curve was offset for clarity. The number of protein molecules in each supermolecule is from bottom to top : 1, 2, 13, 19, 43, 55, 79, 87, 135, 141, 153, 177, 201, 225, 249, 273, 321, 369, 381, 405, 429, 435, 459, 483, 531, 555, 603, 627, 675, 683, 707, 719, 767. Each curve was multiplied by a Debye-Waller factor to damp the high q oscillations to make them look more like the real data.

over a small range of sizes will affect very little the overall structure factor at values of q above the first minimum. The very low q part, which is often not even measured in our experiments would then average out to a smooth function.

It is worth asking the question of whether the small clusters can be composed of other plausible structures. One such possibility would be the formation of a hexagonal close-packed (hcp) structure. The hcp structure is similar to the fcc structure with identical layers in the {111} plane but with a different stacking of these layers. Similar simulations of the structure factor were performed using an hcp structure. The result clearly showed the data is inconsistent with the hcp model. Most of the Bragg peaks of fcc and hcp are the same due to the similarity of the two structures. However, the hcp lattice gives rise to some peaks which are clearly not present in the data.

4.5.2 Planar clusters

Work published by Yau and Vekilov presents evidence that the nucleation of apoferritin proceeds along a planar rather than an isotropic pathway [92]. They measured directly the structure of small clusters of proteins using atomic force microscopy (AFM). Their data

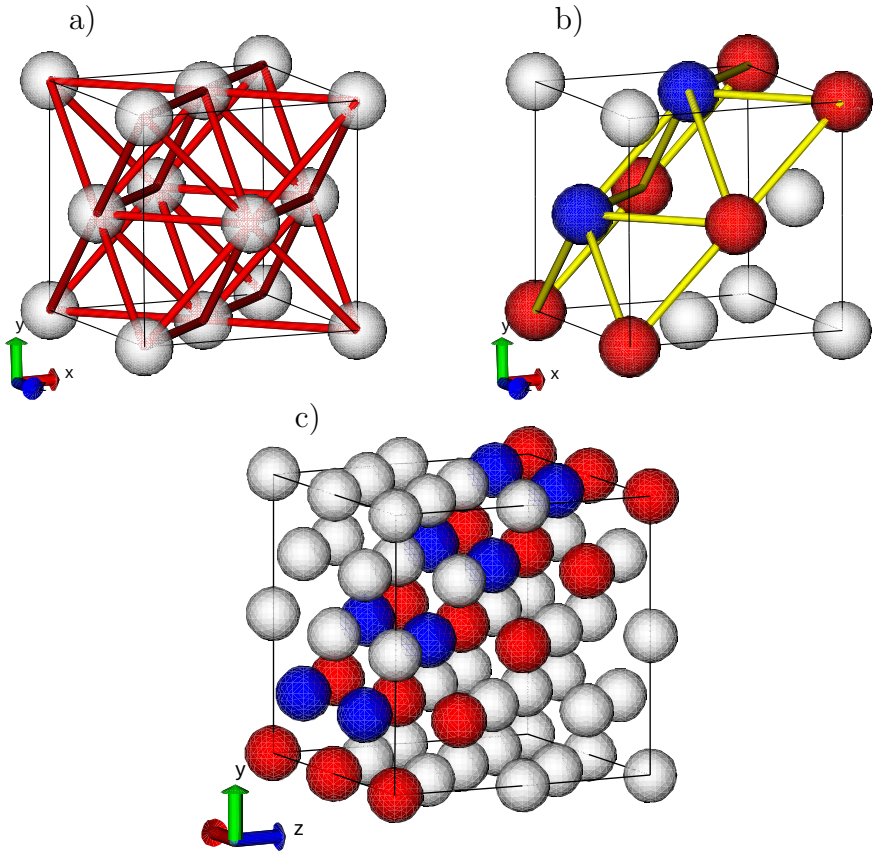


Figure 4.28: a) Bonds (red line) between ferritin proteins in an fcc crystal. b) Contacts, shown in yellow between two parallel [110] rods (red balls) and a third out of plane [110] rod (blue balls). c) Quasi-planar structures of apoferritin measured by Yau and Vekilov [92]. The red balls form one plane with the blue balls forming a second plane. This quasi-planar structure is a subset of the fcc structure also shown.

indicates the formation of rod structures aggregating into a close to planar structure. Supermolecules of this type were simulated in order to determine if they are a good representation of our data.

In the fcc ferritin crystal structure, all the contacts between the molecules are along the $\langle 110 \rangle$ directions as shown on Figure 4.28a). Each of the contacts is represented as a red line. Any structure with underlying fcc symmetry will be formed by conserving some of the bonds and getting rid of others. The structure measured by Yau and Vekilov consisted of [110] rods as shown on Figure 4.28b) where two such rods are shown using red balls and one is shown using blue balls. The two red rods, along the face diagonal of the unit cell form a plane. They however do not have any contacts with each other. Therefore, a third rod is included in the structure to bring them in contact. This blue rod is in a different plane than the first two. The bonds involved in such a scaffold structure are shown in yellow.

As more rods are included, two distinct planes are formed as shown on Figure 4.28c). The final structure formed is therefore close to planar, but is in fact a triangular wave. It is an fcc structure with a different habit. The AFM data of Yau and Vekilov indicates

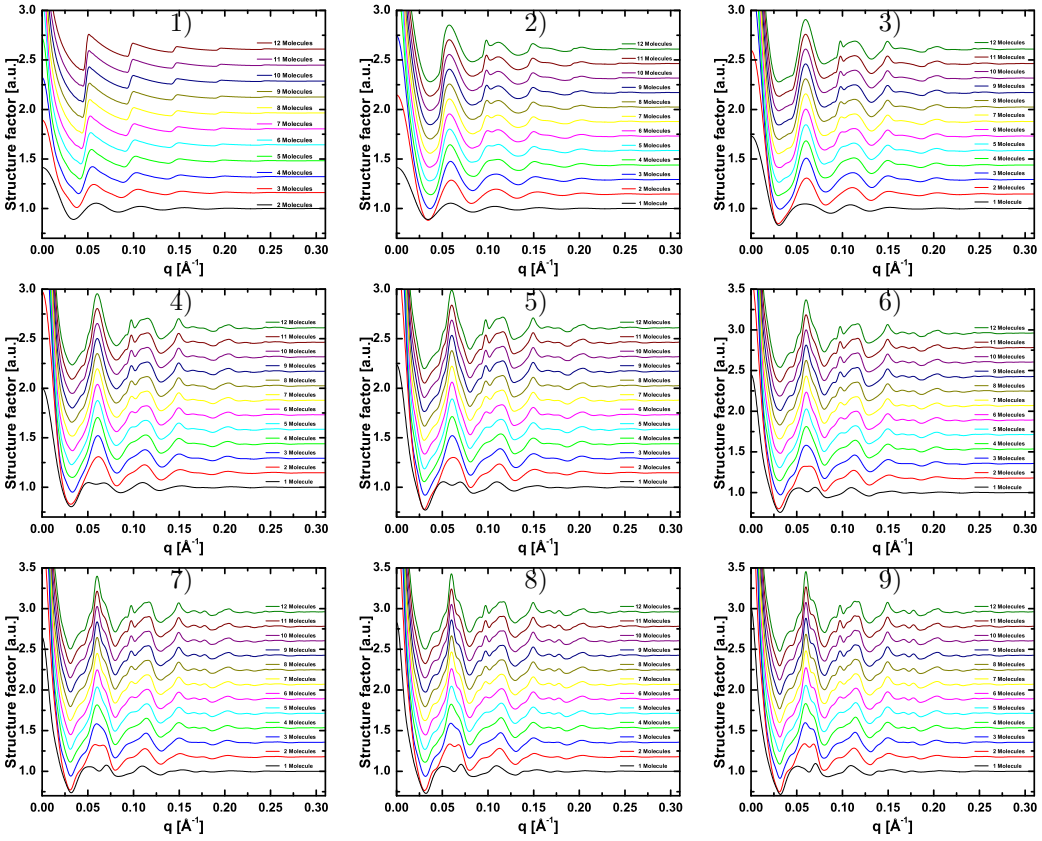


Figure 4.29: Simulated structure factors for ⟨110⟩ rods described on Figure 4.28. The number of the graph indicates the number of rods for this particular simulation. For each graph, the number of molecule per rod was varied from 1 to 12.

that clusters possessing this kind of structure containing between 5 to 10 rods with 4 to 8 molecules in each formed in apoferritin solution.

The simulated structure factors from these quasi-planar structures are shown on Figure 4.29 for many different number of rods with a varying number of molecules in each rod. The number associated with each graph represents the number of rods for this graph. These simulations were done for 1 to 9 rods containing between 1 to 12 molecules per rod.

There are significant differences between the isotropic cluster structure factors of Figure 4.27 and the structure factors from the planar structures. This indicates that SAXS is a good technique to be used to distinguish between the two even if they are randomly oriented. The peaks for both cases are and should be at the same q values and get narrower with increasing cluster size since the position of the peaks depends on the lattice, which is fcc in both cases. However, the planar structure has many planes of the fcc structure missing and therefore some characteristic distances either occur less often or not at all. This leads to different widths of the peaks. Similarly, the size of the cluster is different in all directions. The directions where it is larger will give rise to narrower peaks while the peaks corresponding to the directions where the cluster is small will be much broader. This leads to the first peak getting much narrower before it starts to split into two peaks when comparing to the

isotropic cluster case.

4.6 Fitting

In order to identify which of the simulated structure, if any, best represents our data, a fitting routine was written based on the Levenberg-Marquardt method [44]. This method is a combination of the steepest descent method and the Taylor method. It is a standard minimization routine used for curve fitting. It requires complete knowledge of the fit function and its derivatives with respect to all the fit parameters. In our present problem, there is no simple function of a set of variables to be fitted. Instead, what needs to be fitted to the data is a set of numerical functions which depend in some way on a set of variables. Each of the simulated structure factor for a certain structure and number of molecules was stored in a library. The goal of the algorithm is to identify which of theses functions better represents the data.

The simulated structure factors are not expected to directly match the data. First of all, not all the protein molecules are necessarily included into a cluster. Therefore the simulated structure factor is multiplied by a probability or a fraction of the total number of proteins crystallized in this particular structure. In the case of a single type of structure formed denoted by n, the rest of the molecules are isolated, with a structure factor equal to unity. The total structure factor, denoting the crystallized fraction as p_n is

$$S(q) = p_n S_n(q, a) + (1 - p_n) \tag{4.1}$$

where the probability of protein molecules not to be included in a cluster is $1 - p_n$ and a is the lattice spacing of cluster n. In order to account for possible error in the overall scale of the measured structure factor, the $1 - p_n$ term is replaced by a background value (A) to be fitted by the algorithm. The structure factor to be fitted is then

$$S(q) = p_n S_n(q) + A \tag{4.2}$$

When the structure factor is measured, the square root of the intensity is divided by the form factor, which was itself measured as the square root of the intensity of a dilute sample. If the background subtraction for either or both of the data sets used was not perfect, this introduces a quadratic drift to the structure factor. The measured structure factor is, assuming the backgrounds are small

$$\begin{aligned} S_{\text{meas}}(q) =& \frac{S(q)F(q) + c_1}{F(q) + c_2} \\ \approx& \frac{S(q)F(q)}{F(q)(1 + c_2/F(q))} \\ \approx& S(q)(1 - c_2/F(q)) \\ =& S(q)(1 - c_2 q^2) \\ =& S(q) - c_2 q^2 \end{aligned} \tag{4.3}$$

where the first order approximation of the form factor as q^{-2} was used and $S(q)$ was approximated as unity in the second term. The value of c_2 is a second fit parameter to be determined.

As mentioned before, there is disorder as well as thermal vibrations. This leads to a damping of the structure factor at higher q, the Debye-Waller factor discussed in Chapter 2. The Debye-Waller factor damps the intensity to zero in the high q limit. However, because the structure factor is normalized to unity in our measurements, the effect of the Debye-Waller factor on the structure factor is to damp the oscillations around unity. Therefore, the structure factor is constructed to be

$$S_{\text{fit}}(q) = A + c_2 q^2 + p_n \left[(S_n(q,a) - 1)(e^{-\frac{1}{2}q^2\langle u^2\rangle} + 1 \right] \tag{4.4}$$

where the r.m.s. displacement $\langle u^2 \rangle$ is the third fit parameter. This equation in analytic and therefore its partial derivatives with respect to the three fit parameters can be calculated as needed by the fitting routine.

In Equation 4.4 above, the simulated structure factors, which were shown on Figure 4.27 and 4.29 are explicitly denoted as functions of a, the lattice spacing of the structure. Since our data strongly suggests this spacing varies with temperature, it must be another fit parameter that is free to vary. However, we only have a list of numbers for $S_n(q,a)$ and no function directly relating how it depends on a. We need not only this, but also the a dependence of the derivative of $S_n(q,a)$. We can understand the a dependence by going back to the original Debye formula of Equation 2.30. The structure factor is a function of a and q, but only in the multiplicative combination qa. This allows us to realize that a change in a is equivalent to a change in q such that qa is the same. Therefore, the effect of changing the lattice spacing is to stretch or shrink the entire simulated structure factor by rescaling the values of q.

As for the derivative with respect to a, one can calculate it as

$$\frac{\partial S(q,a)}{\partial a} = \frac{\partial S(q,a)}{\partial q}\frac{\partial q}{\partial qa}\frac{\partial qa}{\partial a} = \frac{q}{a}\frac{\partial S(q,a)}{\partial q} \tag{4.5}$$

Therefore, by calculating numerically the derivative of the simulated $S_n(q)$ with respect to q, we can deduce the derivative with respect to the lattice spacing. We then have all the partial derivatives necessary to proceed. The four fit parameters are a background, a slope, a fraction for clusters n and a lattice spacing.

It is logical to assume that there is a cluster size which is more likely than others. The distribution could be gaussian around a central supermolecule size. In order to determine this central value, a fit was performed for each structure factor measured assuming only 1 type of supermolecule was present. A fit was performed for each model in the library, including spherical clusters up to 177 molecules and planar clusters containing anywhere between 1 to 11 rods of 1 to 12 molecules. The program written called **lsqfit**, kept track of the parameters obtained for each of these fits and at the end, the supermolecular model yielding the lowest value of $\chi^2 = \sum[S(\text{measured}) - S(\text{fit})]^2$ was output. Thus 309 fits were performed on each data set and the best fit was kept as the most likely supermolecule, among all the simulations performed.

Good fits were not obtained in every case using this automated method. For some of the data, the very best fit obtained was clearly not a valid representation of the data. In other cases, the fit was good for parts of the data but failed at other parts. The lack of a good fit could be due to a few reasons. First, there is noise in the data which looks like real features that the program tries to fit but simply cannot accommodate since they are not real. Second, one of the models used was a good model, but somewhere during the fitting routine, a local minimum was found and yielded a poor fit. This is unlikely to have occurred since the initial guess of every fit was set to closely match the observed quantities such as the lattice spacing. Third and most simply, none of the simulated supermolecules correspond to the sample. In this case, one obviously does not converge to a proper solution.

4.6.1 Holoferritin

Some results of the fits obtained are shown for holoferritin on Figures 4.30, 4.31 and 4.32 for the 100mg/ml sample previously shown on Figure 4.9, the 200mg/ml sample shown on Figure 4.10 and the second 100mg/ml sample shown on Figure 4.11 respectively. Of all the holoferritin samples shown above, these were the only ones where no large fcc crystal formation was observed. The samples with lower concentrations produced large crystals and it is very easy to determine the structure of these. The Scherrer formula used to estimated the size of the crystallites breaks down when the size becomes too small, but is a very good estimate for larger crystals. Therefore, the results for the crystal size obtained previously for holoferritin samples at concentrations between 5 and 50mg/ml are correct and there is no need to fit the habit of the crystals since the information is washed out. The Bragg peaks are also well-defined and the lattice spacing can be determined accurately.

The fits shown on Figure 4.30a) are the ones obtained for this sample that were deemed good. The fits for the data obtained at other temperatures were not adequate for the various reasons discussed above. The fits shown as from data ranging from -7.7°C to -2.6°C. The curves are shown in order at which they were obtained with the first at the bottom and the last at the top. There is a good agreement between similar temperatures taken at different times and with different thermal history. The best fit obtained for some of the curves was using spherical supermolecules while for some others it was using the planar clusters. This result in this case is not fully conclusive because the curves at low q are fairly similar for both types of clusters. There is a different ratio of the height of the first peak to its width and this leads to a better fit for planar clusters. However, to truly see the difference between the two simulations, one must go to higher q. In the holoferritin data, the higher q part is not fully reliable due to the polydispersity of the molecules as previously discussed. Nevertheless, there seems to be an indication of planar clustering in the system.

The fitted value to the lattice spacing is shown on Figure 4.30b) for all the temperatures measured. Even in the case where the fits were not very good over the whole data range, the position of the first peak was always accurately matched by the fit. Therefore, this fitted spacing is a reliable value. Also plotted on Figure 4.30b) is the estimated guess to the spacing using the position of the peak shown previously on Figure 4.9b). A good agreement between the two is found despite the first peak actually being a double peak. There is a

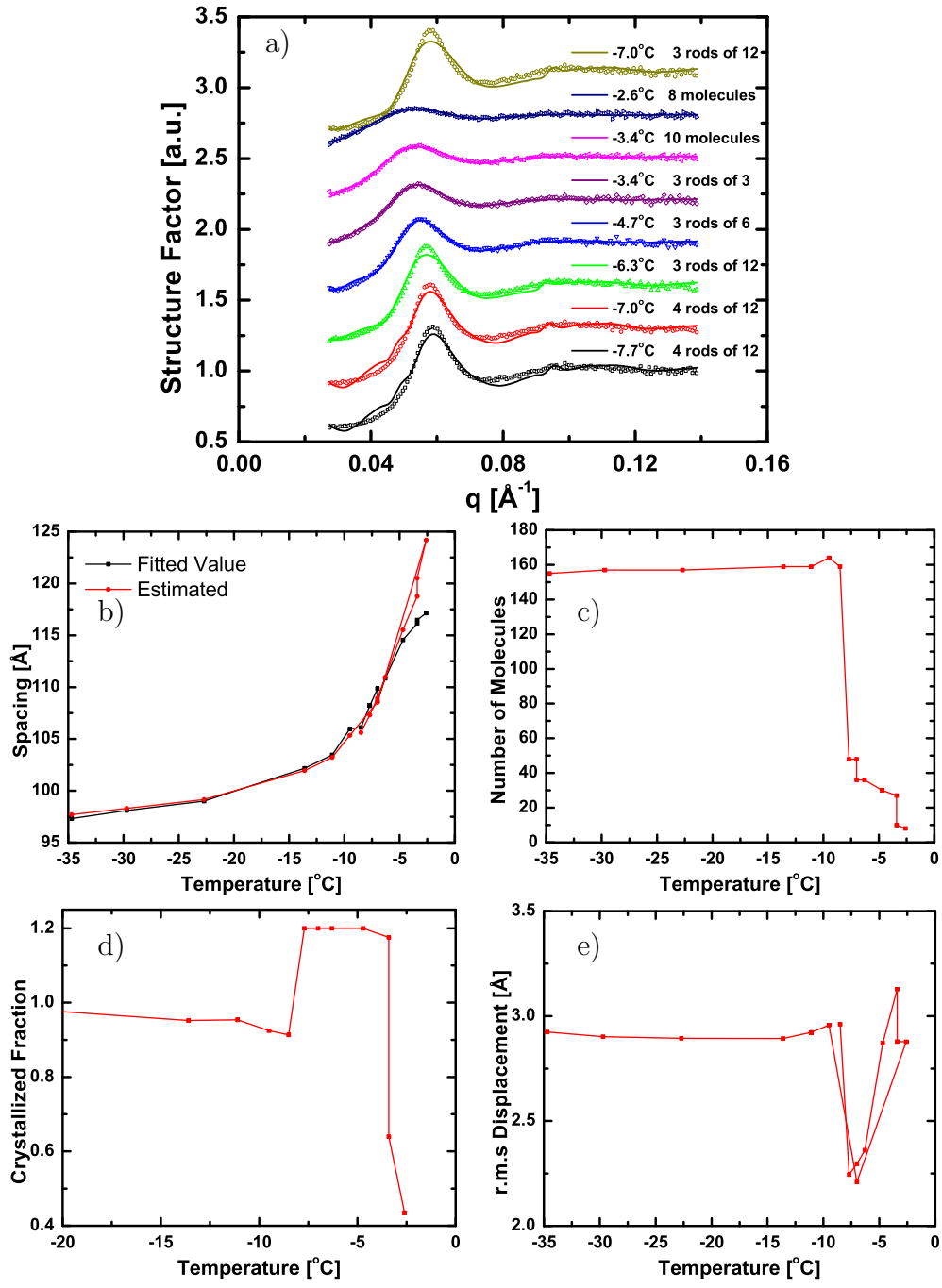


Figure 4.30: Results of fits to the structure factors for 100mg/ml holoferritin shown on Figure 4.9. a) Symbols are the data points and the solid lines are the fits. The model to which the fit corresponds is written on the right along with the temperature of the sample. b) Fitted lattice spacing (black) compared with the estimated spacing previously shown. c) Total number of molecules in the supermolecular model yielding the best fit. d) Crystallized fraction, *i.e.* the fraction of proteins illuminated by the x-ray beam that are included in one of the clusters. e) Root Mean Square displacement from the ideal lattice positions obtained during the fit corresponding to both thermal and configurational disorder.

disagreement between the two at high temperature when the clusters get smaller and the shift in peak position from the presence of the double peak is the greatest.

The total number of molecules in the simulated supermolecule which the fitting algorithm identified as representing the data best is shown on Figure 4.30c). This is shown for all temperatures including the poor fits. The values at temperatures below -7.7°C are therefore not reliable. The results above -7.7°C indicate the clusters break apart as the temperature increases in that region. This is the range of temperature where the sample can be superheated before melting. The sample freezes at roughly -8°C but only melts back at roughly -2°C.

The crystallized fraction fitted is shown on Figure 4.30d). This value should not be allowed to be greater then unity. However, some of the fitted values are greater than one. This parameter only scales the simulated structure factor in order to make it match the data. The fact that it comes out greater than 1 can indicate that either the simulation is wrong, another model leading to a larger peak with the same width would be better or more likely the polydispersity of the molecules due to the presence of iron leads to a wrong measured structure factor. The simulations at temperatures above -7.7°C match the data very well in terms of peak position and width. They only fail to match the peak height which is not a free parameter, other than the crystallized fraction showed. Nevertheless, there is clear indication the clusters are breaking apart as the sample is heated with a decreasing crystallized fraction. The protein molecules leaving clusters revert back to a more dilute disordered state.

The root mean square displacement of the molecules from their ideal lattice positions leading to a Debye-Waller factor decrease in the intensity was fitted as described above and plotted on Figure 4.30e). Again the data for the poor fits is also shown along with the data from the good fits above -7.7°C. For this range of temperature, an increase in disorder is measured, indicating again a destruction of the order structures. The damping of the oscillations can be fairly accurately fitted even if the overall fit does not match the peak positions very well as long as the relative peak heights are well matched. Therefore, the data at all temperatures is valid.

Figure 4.31a) shows the good fits obtained on the holoferritin sample at 200mg/ml. Similarly to the 100mg/ml sample, good fits were only obtained at the highest temperature measured, above -5°C. The other fits not shown were able to match the first peak with good agreement but the other peaks were not matched. However, as mentioned, the polydispersity of the holoferritin molecules means we don't expect to be able to match these other peaks. Therefore, the fits parameters obtained and shown can be expected to be near the real values even for the poor fits. It is noteworthy to mention that the sample at 6.6°C was fitted to the dimer structure. This would seem to indicate that the freezing and melting process irreversibly forms clusters which do not break apart. A solution with little dimers forms a large amount of them after it is frozen and then heated back above the melting point.

On Figure 4.31b), the fitted lattice spacing agrees again well with the estimated value from Figure 4.10b) until the clusters get small at high temperature. The position of the peak is not an accurate measure of the lattice spacing when the supermolecules become too small, as is the case here with fitted number of molecules per cluster below 20 as shown on

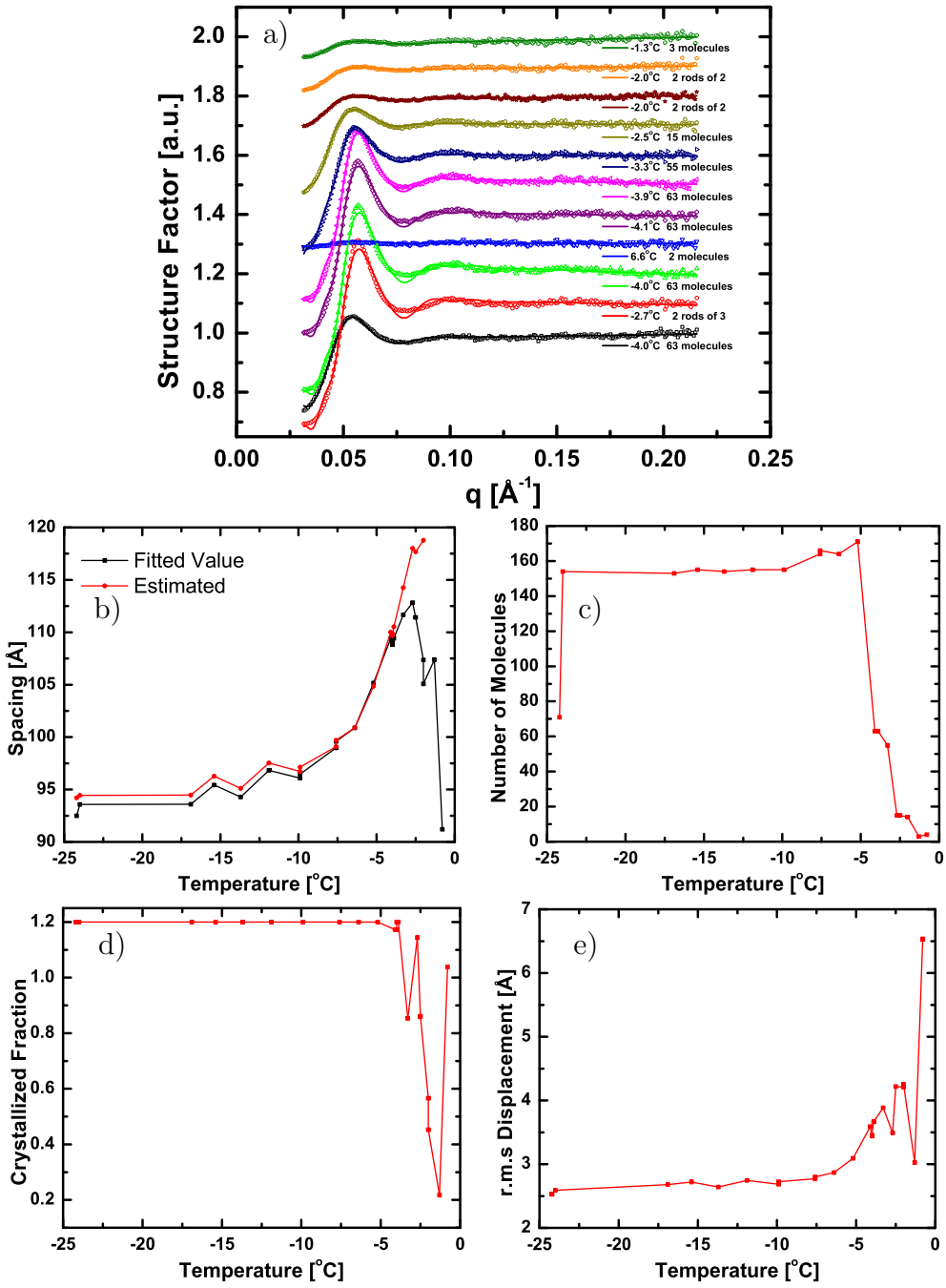


Figure 4.31: Results of fits to the structure factors for 200mg/ml holoferritin shown on Figure 4.10. a) Symbols are the data points and the solid lines are the fits. The model to which the fit corresponds is written on the right along with the temperature of the sample. b) Fitted lattice spacing (black) compared with the estimated spacing previously shown. c) Total number of molecules in the supermolecular model yielding the best fit. d) Crystallized fraction, *i.e.* the fraction of proteins illuminated by the x-ray beam that are included in one of the clusters. e) Root Mean Square displacement from the ideal lattice positions obtained during the fit corresponding to both thermal and configurational disorder.

Figure 4.31c). The values below -5°C for the fitted number are unreliable.

The crystallized fraction on Figure 4.31d) shows a trend consistent with the number of clusters decreasing with temperature. The r.m.s. displacement on Figure 4.31e) shows an increasing level of disorder with increasing temperature, consistent with clusters breaking up.

The fits for the second sample at 100mg/ml previously shown on Figure 4.11 are shown on Figure 4.32. As we recall, this particular sample produced slightly larger clusters than the of sample at 100mg/ml as indicated by the split first peak. Furthermore, annealing at high temperature produced even larger crystals with more well-defined peaks. The fits shown on Figure 4.32a) represent that fact with larger supermolecular models representing the data well. The fitted lattice spacing on Figure 4.32b) agrees well with the estimated spacing from the peak position alone. The number of molecules in the best fit shown on Figure 4.32c) show a value near 160 molecules which goes up to 309 at the lowest temperature. The value of 309 is meaningless in the sense that this was the largest cluster simulated. The sharp peaks in the data were best represented by this model since it produced the sharpest peaks. This number simply indicates that an fcc structure larger than what was simulated was present.

The crystallized fraction on Figure 4.32d) decreasing at high temperatures despite the formation of larger crystals is somewhat surprising. It suggests that larger crystals form from the existing smaller ones, but that not all of the small clusters get included in the larger crystals. Some clusters simply redissolve. The r.m.s. displacement shown on Figure 4.32e) displays a different behavior from before. The disorder in the clusters decreases with increasing temperature. This is due to the formation of larger crystals as opposed to clusters falling apart. Larger crystals must be well-ordered to exist while smaller aggregates can be more disordered. A high level of disorder would prevent the very formation of the large crystals.

4.6.2 Apoferritin

The structure factors for apoferritin are more reliable than for holoferritin due to the higher monodispersity of the molecules. We therefore expect the fits of these structure factors to reveal more information about the crystallization process. The apoferritin structure factors show a lot more features, which makes it easier for the fitting algorithm to get stuck in local a minimum. Therefore, good fits of the whole SAXS patterns were difficult to obtain. The fits for apoferritin at 50mg/ml are shown Figures 4.33 and 4.34 for the two samples, previously shown on Figure 4.21 and 4.22.

The fits at four temperatures are shown on Figure 4.33a). The agreement between the fit and the data gets a little worse as the temperature decreases. The second, third and even fourth maxima are fairly well captured before the signal level drops below what could be measured. The lattice spacing fitted does not match well the estimated value at the lowest temperatures on Figure 4.33b). This can clearly be seen from the fit at -18.9°C. The position of the first and second peaks are inconsistent with each other in the sense that their positions are not an acceptable ratio of each other according to the simulated supermolecules. The

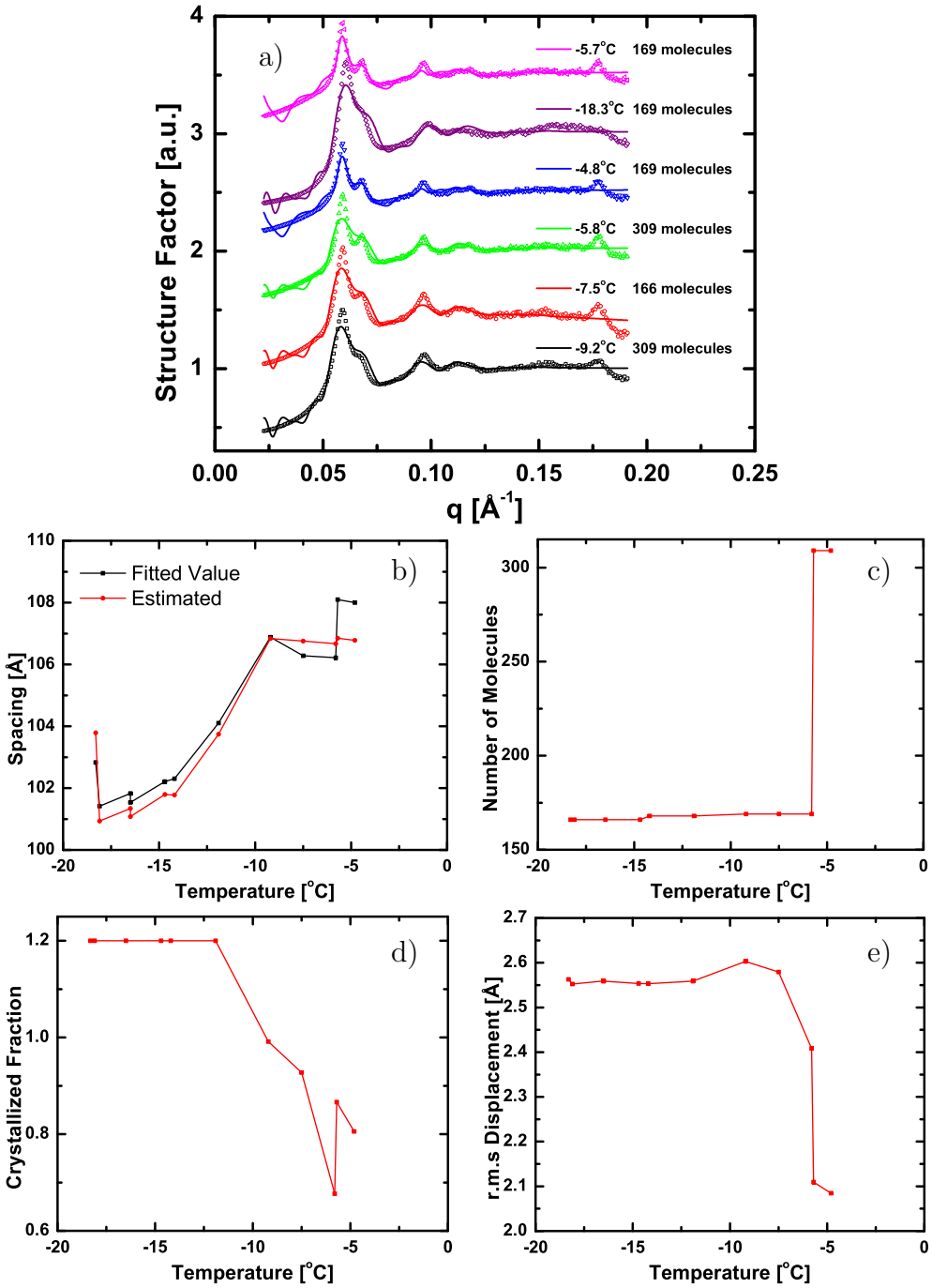


Figure 4.32: Results of fits to the structure factors for 100mg/ml holoferritin shown on Figure 4.11. a) Symbols are the data points and the solid lines are the fits. The model to which the fit corresponds is written on the right along with the temperature of the sample. b) Fitted lattice spacing (black) compared with the estimated spacing previously shown. c) Total number of molecules in the supermolecular model yielding the best fit. d) Crystallized fraction, *i.e.* the fraction of proteins illuminated by the x-ray beam that are included in one of the clusters. e) Root Mean Square displacement from the ideal lattice positions obtained during the fit corresponding to both thermal and configurational disorder.

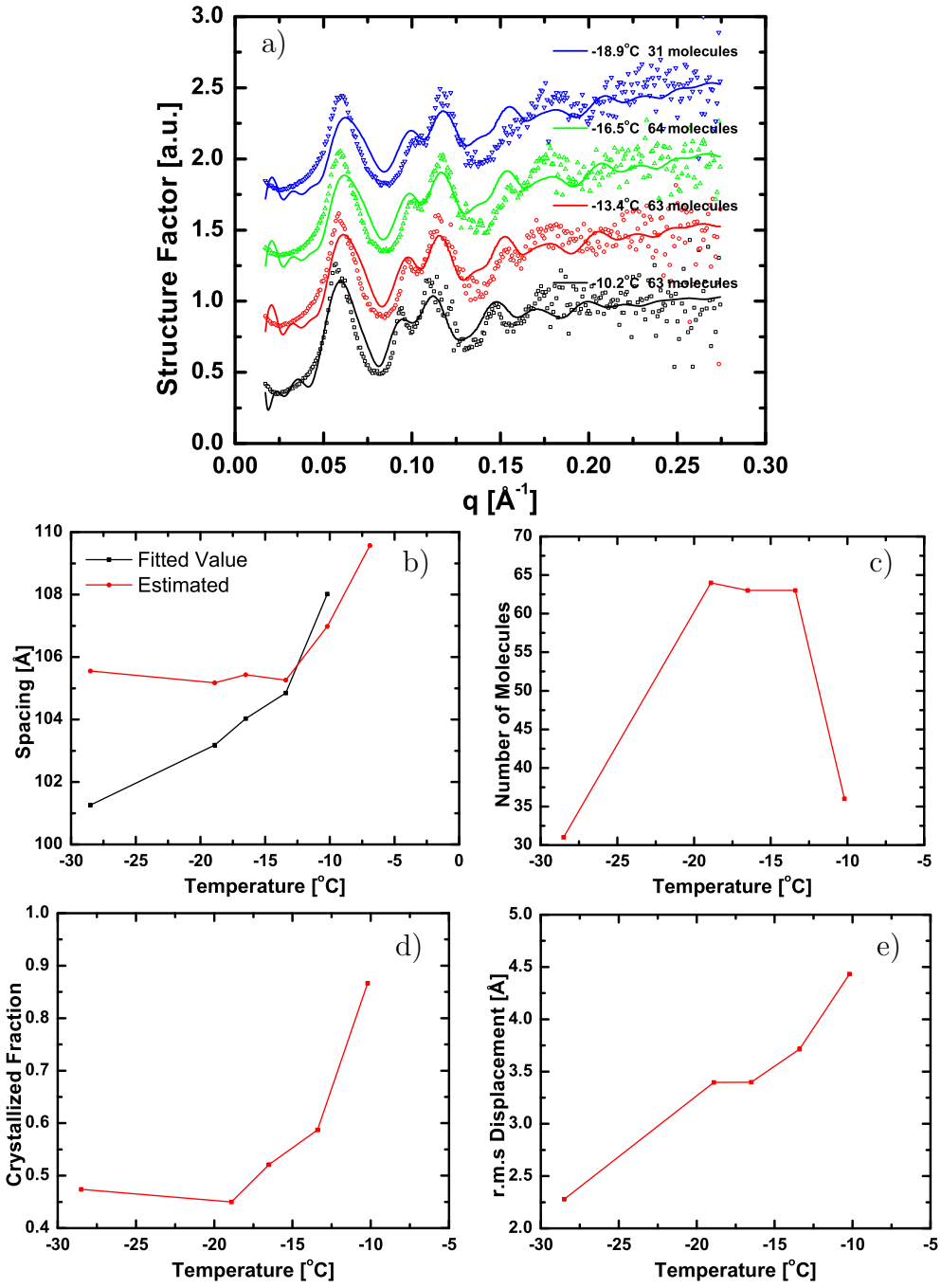


Figure 4.33: Results of fits to the structure factors for 50mg/ml apoferritin shown on Figure 4.21. a) Symbols are the data points and the solid lines are the fits. The model to which the fit corresponds is written on the right along with the temperature of the sample. b) Fitted lattice spacing (black) compared with the estimated spacing previously shown. c) Total number of molecules in the supermolecular model yielding the best fit. d) Crystallized fraction, *i.e.* the fraction of proteins illuminated by the x-ray beam that are included in one of the clusters. e) Root Mean Square displacement from the ideal lattice positions obtained during the fit corresponding to both thermal and configurational disorder.

fitting algorithm found a solution which matches the second and third peaks well but misses slightly the first one. Therefore, when using the first peak only to estimate the lattice spacing, the number obtained is wrong compared with the number obtained when using the entire data. The lattice is seen to expand with increasing temperature.

The number of molecules in the fitted cluster decreases with increasing temperature as shown on Figure 4.33c). The value at -29°C should be disregarded since this fit was a poor one. Contrary to what was observed before, the proportion of molecules that get included in a cluster increases with temperature as shown on Figure 4.33d). This is surprising and possibly indicative that the values obtained for this fitting parameter cannot be directly related to a real quantity but are rather a fit fudge factor allowing the overall scale of the data to be matched. The r.m.s. displacement fitted shown on Figure 4.33e) increases with temperature as before. Comparing the absolute value with the holoferritin data, the value is lower. This is likely due to the polydispersity of holoferritin which results in an artificially large r.m.s. displacement, a measure of disorder.

The fits for the second apoferritin sample at 50mg/ml are shown on Figure 4.34a). It is again clearly evident that fitting both the first peak and the secondary maxima at the same time with a single supermolecule is not a perfect model. The lattice spacing shown on Figure 4.34b) shows a large increase with temperature and matches the estimated value from Figure 4.22 well. The lack of perfect agreement between the fits and the data implies that the actual number of protein molecules in the model supermolecule is not a very reliable quantity. Looking at the trend on Figure 4.34c) however, one can see that the cluster size increases until roughly -10°C and then fall off. This again indicates the formation of larger crystals as the temperature is increased, until a certain temperature is reached and the clusters break apart.

The behavior of the crystallized fraction shown on Figure 4.34d) is a little messy. It does however seem to show a downward trend with increasing temperature consistent with clusters disintegrating. Finally, the r.m.s. displacement fitted shown on Figure 4.34e) shows the typical increase with temperature.

4.7 Interpretation of Results

The formation of small clusters of ferritin as the solution is frozen appears to be a phase separation process as mentioned before. Ice forms around the proteins leading to a low protein concentration in the ice and protein rich regions which a trapped into voids in the ice network. The protein rich regions may or may not be frozen. They may still be in a liquid aqueous environment due to their high solute percentage. The phase separation explains the hysteresis of the system causing it to have different freezing and melting temperatures

There is evidence that similar structures possessing face centered cubic symmetry are present at all temperatures. At the highest temperatures, they disintegrate into smaller fcc structures and eventually revert back to a solution. The evidence indicates the clusters beak up into smaller ones before they revert to liquid-like highly concentrated structures though there is some evidence of liquid-like packing leading to peaks in the structure factors. The large lattice spacings observed at high temperatures are better explained by a concentrated

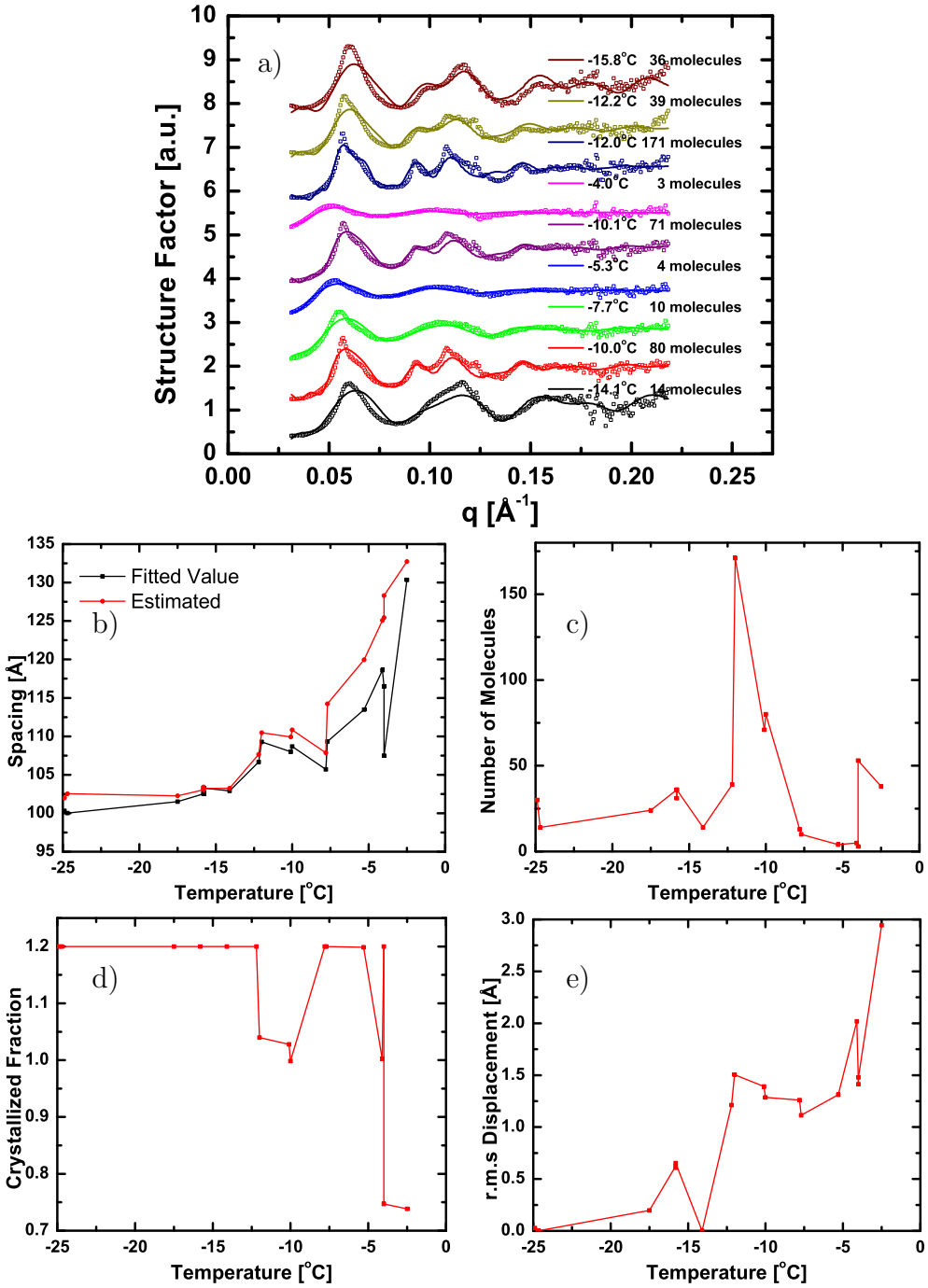


Figure 4.34: Results of fits to the structure factors for 50mg/ml apoferritin shown on Figure 4.22. a) Symbols are the data points and the solid lines are the fits. The model to which the fit corresponds is written on the right along with the temperature of the sample. b) Fitted lattice spacing (black) compared with the estimated spacing previously shown. c) Total number of molecules in the supermolecular model yielding the best fit. d) Crystallized fraction, *i.e.* the fraction of proteins illuminated by the x-ray beam that are included in one of the clusters. e) Root Mean Square displacement from the ideal lattice positions obtained during the fit corresponding to both thermal and configurational disorder.

liquid than an expanded lattice. However, there is clear evidence of a large thermal expansion of the fcc lattice with increasing temperature. The lattice spacing of 35mg/ml holoferritin producing large crystals on Figure 4.14c) is directly measured and shows a slight thermal expansion. It is however the smaller structure which show the larger thermal expansion for 100mg/ml and 200mg/ml holoferritin. Fits to the structure factors show that the sample consists of many small fcc supermolecules and the average lattice spacing greatly increases. This is quite interesting because it indicates that the fcc structure is a stable one even when the protein molecules are far apart and no longer directly in contact. This is reminiscent of colloidal systems in which crystal structures can be formed even at very low volume fractions, below the fraction one would get if the colloids where in direct contact. In these colloidal systems, the spacing between the particles in the crystal can be many time the diameter of the particles. It is well-known that the salt concentration in colloidal suspensions changes the organization in the sample from a fluid to various crystalline structures by screening the Coulomb interactions. The addition of salt to the ferritin solution changes the size of the fcc clusters formed and whether they occur at all. The ferritin proteins trapped in ice appears to behave like colloidal particles since they have crystalline order with a varying lattice spacing and without direct contact between nearest neighbors in the lattice.

The thermal expansion probably occurs because of changes in the ice network with changing temperature. As the temperature decreases, the void in the ice where the proteins are trapped changes slightly. This small changes has a large effect because this exerts pressure on the small clusters of proteins. It is known that protein crystals are very soft compared with inorganic crystals and also ice [21, 56, 57]. A small pressure applied to protein crystals from the ice will greatly affect their structure. As the temperature is increased, the pressure is relaxed, the protein molecules start to pull apart, yet they conserve their fcc structure even after they no longer are in contact.

The measured decrease in size and number of the clusters when approaching the melting point is consistent with previous measurements of Kilcoyne *et al.* using small angle neutron scattering [36]. Their analysis was based on the characteristics of the very low q range where the presence of large structures has a large effect. They did measure a few weak peaks in the intensity which they tentatively assigned to Bragg peaks from the hexagonal close-packed structure. Their data did not allow to fully verify this. No measurement of the annealing process discussed above was made in their case.

The fits to the structure factors discussed were of only a single cluster model at a time. There is no reason to expect the sample to contain a single cluster structure. There should be a range of different structures with different probabilities. Only the most likely structure was fitted for the following reasons. First, the structure factors simulated for a single type of clusters such as the spherical clusters is a slow varying function of the number of particles. There is not much difference between the structure factor calculated from a supermolecule containing 45 molecules and a supermolecule containing 50 molecules. Therefore, using multiple models and summing them together only slightly improved the fits. Furthermore, including too many models leads to many fit parameters which created too many degrees of freedom leading to multiple solutions. The real system probably includes a structure which is more likely to occur than others with the other structures within a range of sizes present

with some probability, including isomorphic structures which where not all simulated. It proved impossible with the available data to determine this range of size with reliability.

It is also possible that different habits of the clusters are present. Clusters could be forming in the sample with both planar and isotropic structures. It again becomes tricky to allow both to exist in the fitting routine, leading to ambivalent multiple solutions. Therefore, only the most likely structure of the supermolecules could be determined.

Some of the fits obtained indicate the most likely structure present is consistent with the planar nucleation structures measured by Yau and Vekilov [92]. Most of the data is however better explained by the isotropic simulations. It is then not fully conclusive that the planar structures exist since these fits were obtained only on the first peak in the structure factor which is similar for both types of clusters. The clear signature of rod-like aggregation was not seen, but there is evidence of such structures occurring.

The data clearly indicated the presence of very small clusters down to less than 10 molecules at times when the sample is cooled. As the temperature is raised, these small clusters grow into large crystals. The freezing process causes nucleation of the protein molecules. Theses structures are stable at a certain temperature. Raising this temperature causes a change in the level of supersaturation of the sample and these clusters can now grow. Therefore, the structure of the small supermolecules fitted and measured are identified as the nuclei, near the critical size of ferritin at these concentrations and temperatures. It would appear that isotropic nucleation is the main pathway under those conditions but the planar pathway seems to also be present and at the very least cannot be fully ruled out.

Chapter 5

Conclusion and Outlook

5.1 Summary

In Chapter 1 of this thesis, we introduced the problem at hand. That is we were trying to use the technique of coherent x-ray diffraction to measure the three dimensional structure of small protein crystals near the size of critical nuclei. A short review of protein crystallization was presented.

Then in Chapter 2 we introduced the theoretical concepts needed to understand the measurements performed using a coherent beam of x-rays to illuminate a small crystal. We also introduced the Gershberg-Saxton iterative phase retrieval method used later on with real data. There was also a discussion of small angle scattering in general as well as the SAXS patterns expected from a collection of randomly oriented supermolecular clusters of proteins.

With the basis for understanding the measurements in place, we moved on to describing in Chapter 3 the experimental details of a CXD experiment. The preparation of a clean coherent beam of x-rays as well as the preparation of samples fit for such measurements were discussed. Multiple time series of samples containing a large number of small crystals were measured, looking for the signature features of facetted crystals coherently illuminated. These features, which are intensity flares with fringes were rarely seen. It was demonstrated that this was mostly due to motions within the samples which not only smeared the diffraction patterns, but also limited to exposure time in a way that the fine features could not be measured appropriately.

We demonstrated that freezing the samples to reduce the motions did not help and created a new diffraction feature which was explored further in Chapter 4. It was also shown that trapping the small crystals in a gel significantly reduced the random rotation of crystals.

Not surprisingly, radiation damage proved to be an important issue. Surprisingly however, it was found that at room temperature, the exposure to x-rays caused the crystals to collapse linearly with dose. The collapse was fitted to a Hendricks-Teller two state model revealing a 2 to 3Å collapse of the center to center distance of nearest neighbor molecules in the crystals of the protein ferritin.

The reconstructed 2D projection of the electron density was shown for a few examples of CXD data. Even though the data were not of very high quality, a compact object with distinguishable facets was recovered.

Finally, the presence of intensity streaks, identified as truncation rods, which were seen almost only on the high q side of the Bragg peaks, was explained by the presence of inward strain at the surfaces of the crystals and the facets were identified as {111} and {200} facets.

In Chapter 4, the formation of small clusters of proteins at low temperature was investigated using small angle scattering. These clusters were found to be crystalline with a face centered cubic structure. There was evidence of a liquid-like dense packing of the protein molecules at temperatures close to the melting point. The crystalline packing was seen to be very temperature dependent with a large measured thermal expansion. Cooling the sample was shown to create small clusters which would grow to micron sized crystals when annealed. The small clusters could be identified as possibly being related to critical nuclei.

Similations of structure factors were made for spherical clusters as well as planar clusters that were measured by Yau *et al.* [92]. Comparison of the simulations with the data revealed that some of the features of both simulations matched with the data. A fitting routine was written to identify which of the simulated clusters best represented the data. Some evidence was found this way that the clusters may at times be planar in nature, but were mostly well described by the spherical cluster model.

The main result of Chapter 4 was that nanoclusters of ferritin can be produced by simply controlling the temperature of the sample and the size of these can be controlled, as well as their lattice spacing. These nanoclusters were shown to be precursor states to the crystalline state.

5.2 Future Directions

The current need for a better understanding of protein crystal nucleation is clear. Protein crystallography is the only effective way to determine the atomic stucture of proteins and the lack of good crystals is the main factor preventing new structures to be solved. Coherent x-ray diffractive imaging would be an ideally suited method to study the time evolution of small crystals as they grow if it were not for the damage radiation causes to the sample. Bearing in mind that the detection system used in the experiments described above was close to 100% efficient and the experiments were performed at the brightest source of coherent x-rays in the world, it would seem that there is not much hope the technique can be made into a useful tool to study nucleation. The level of irradiation on the sample required to get good statistics is very close to the tolerance of the sample. Increasing the intensity of the source would not solve the problem as it would simply yield the same results faster.

There is however already in the works a possible way to get around that problem. There are plans to build X-Ray Free Electron Lasers (XFEL) at Stanford and at DESY in Hamburg, Germany. The XFEL will be a long linear accelerator with a kilometer or so long array of magnets similar to the current undulators at synchrotron sources, except much longer. The electrons, packed in small bunches, will interact with the emitted x-rays from the previous oscillation, which amplifies the x-ray emission leading to self amplified spontaneous emission. The resulting beam will be up to 10 orders of magnitude more brilliant than the one used at the APS. The emitted radiation will be in short femtosecond pulses.

It is believed that the pulses can be made shorter than the time it takes for damage to the sample to occur. With such a beam, all the x-rays would be scattered by the sample and then the sample would completely disintegrate due to the intense beam. The measurement would then be possible, provided there exists a detector capable of capturing all these x-rays in such a short period of time.

With an XFEL, it should be possible to measure the projection of the electron density of a small crystal onto some plane. It would be impossible to get a 3D image since a single orientation could be measured before the sample gets obliterated. Since it is not possible to create a large amount of identical small crystals allowing for a 3D reconstruction, it appears that these kind of experiments would not be well-suited for the XFEL.

What seems to be a more promising use of the XFEL is the possibility of performing single molecule diffraction [48]. The intense beam would make it possible to measure the scattering from a single protein. This protein could be oriented in the x-ray beam using elliptically polarized laser fields [39]. This would allow many images of individual identical proteins in the same orientation to be measured to improve statistics. The orientation of the molecules could then be changed and images from all orientations would be added to get a 3D diffraction pattern. If this pattern is measured to high enough q and is oversampled, the phase retrieval methods discussed in Chapters 2 and 3 could be used to retrieved the atomic structure of the protein, doing away with the need to crystallize proteins.

This would of course require first great increases in computing power to handle the large data sets and the many FFT cycles needed. It would require more than anything however important detector developments. No detector exists today capable of capturing this kind of scattered beam with the desired speed, detector size and pixel size.

As for the imaging of crystals, it would appear that CXD will be a very useful tool to image the strain fields in 3D of nanocrystals. This strain field cannot be measured directly by any other experimental technique. It would appear that this will however be limited to inorganic nanocrystals where radiation damage is not an issue.

Appendix A

CXD Times Series

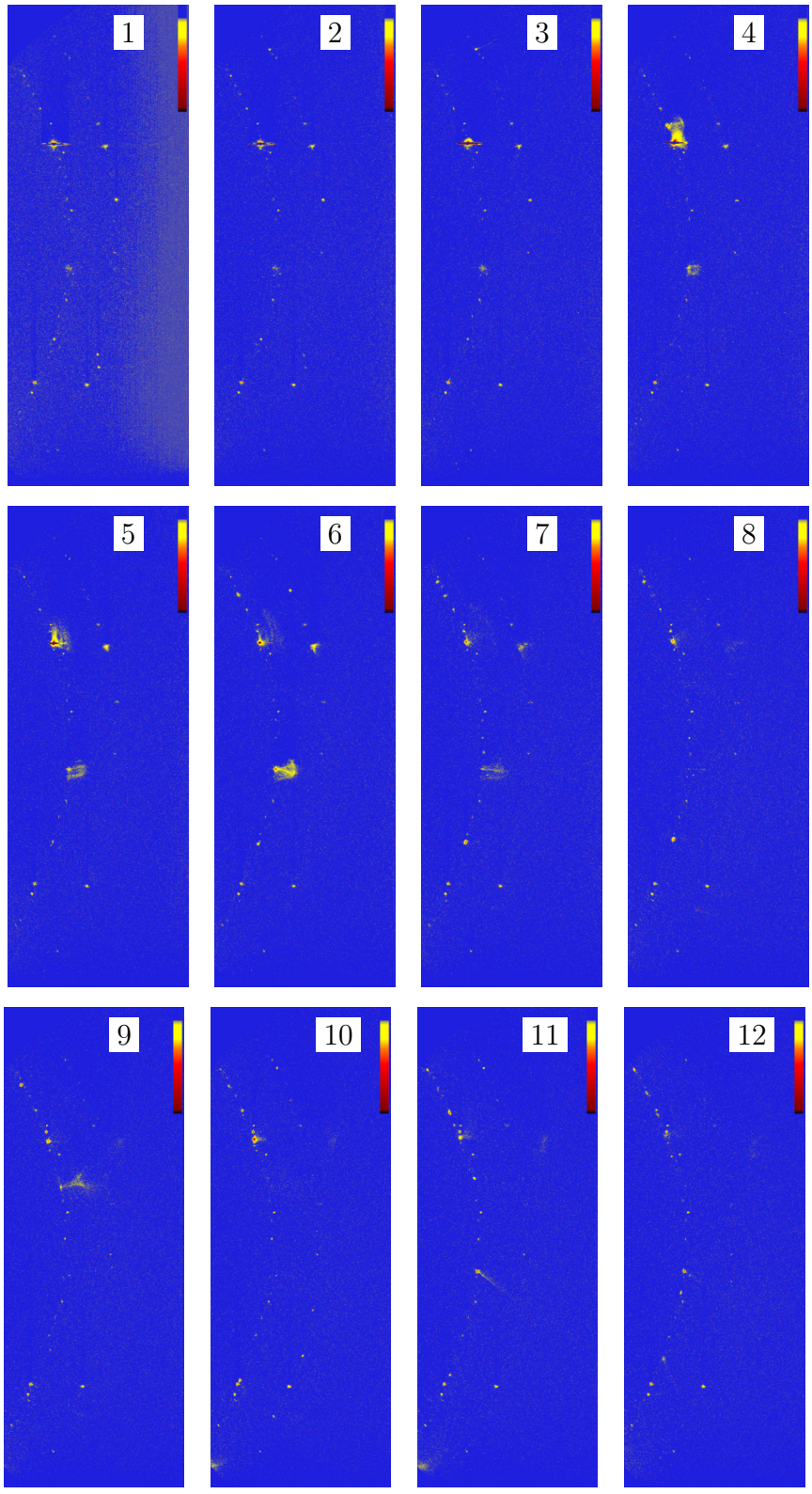


Figure A.1:

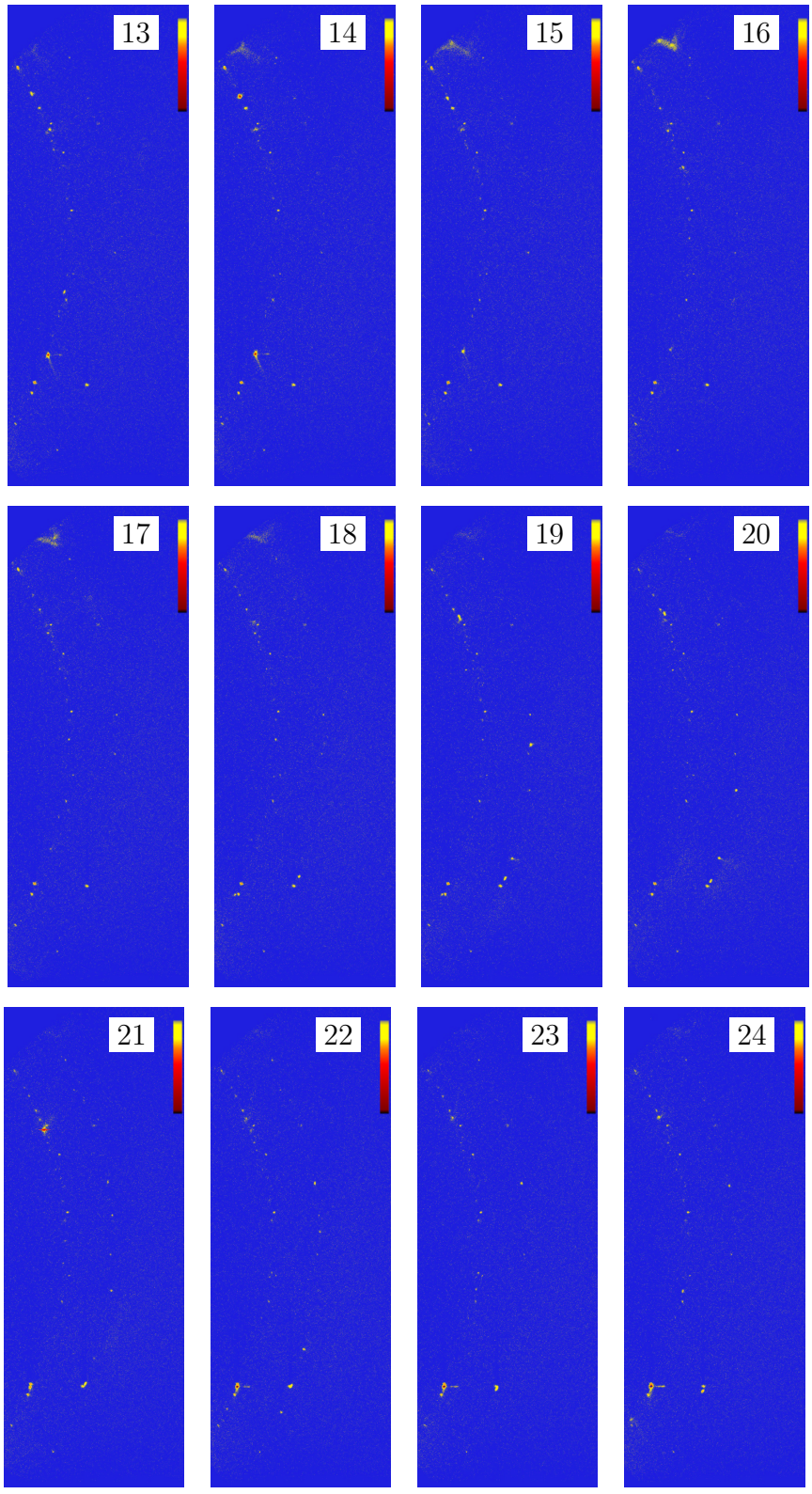


Figure A.2:

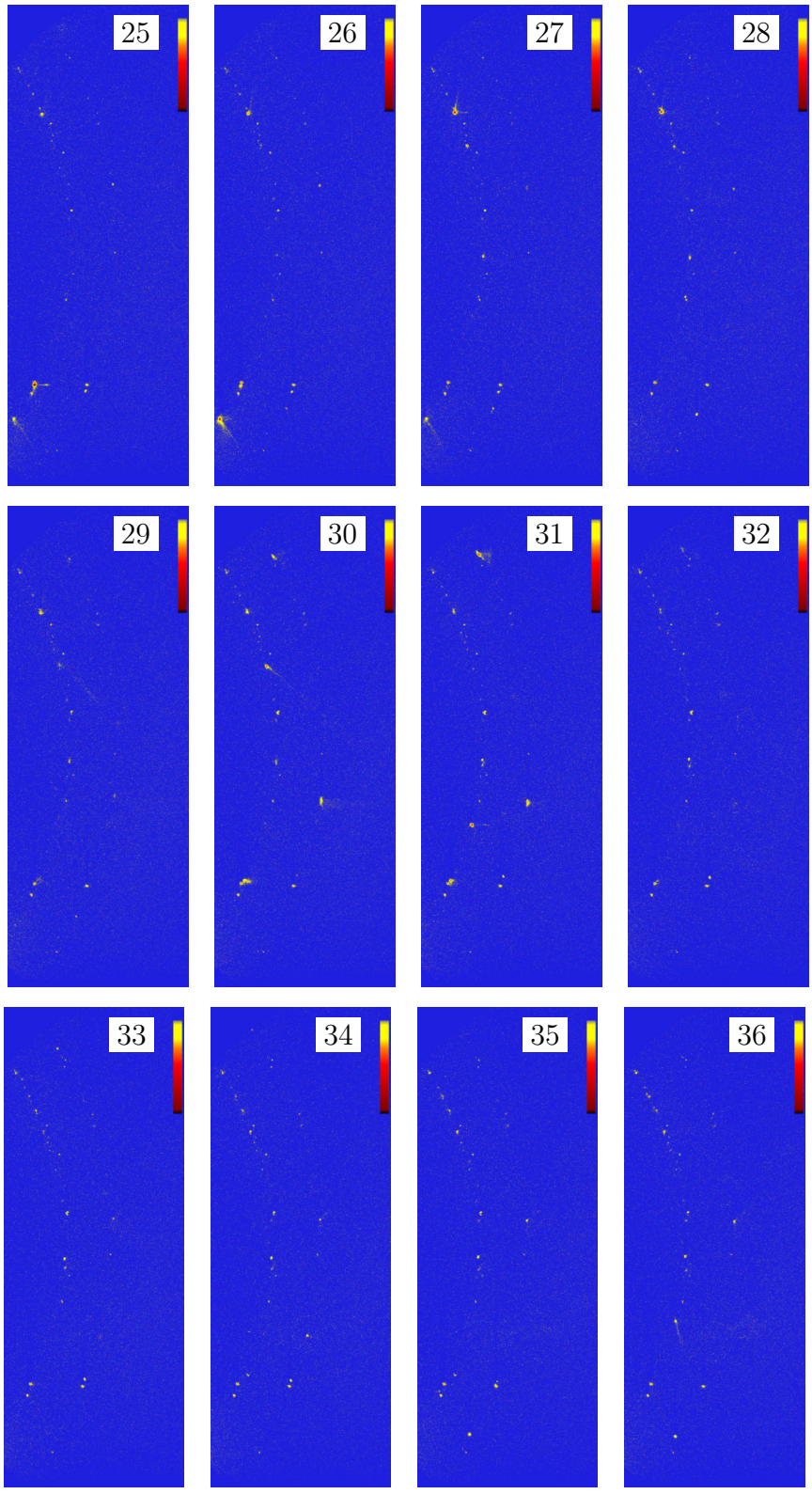


Figure A.3:

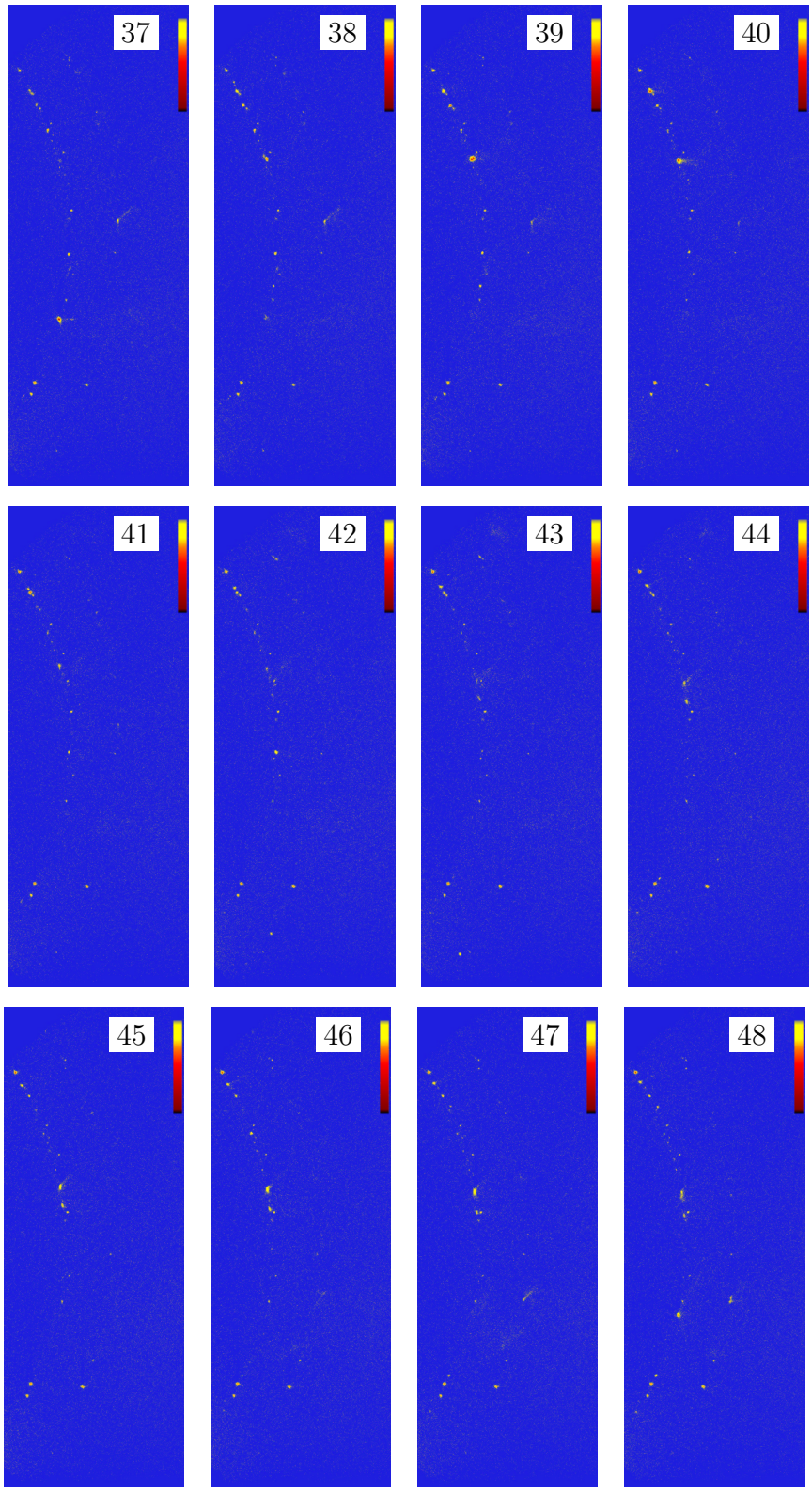


Figure A.4:

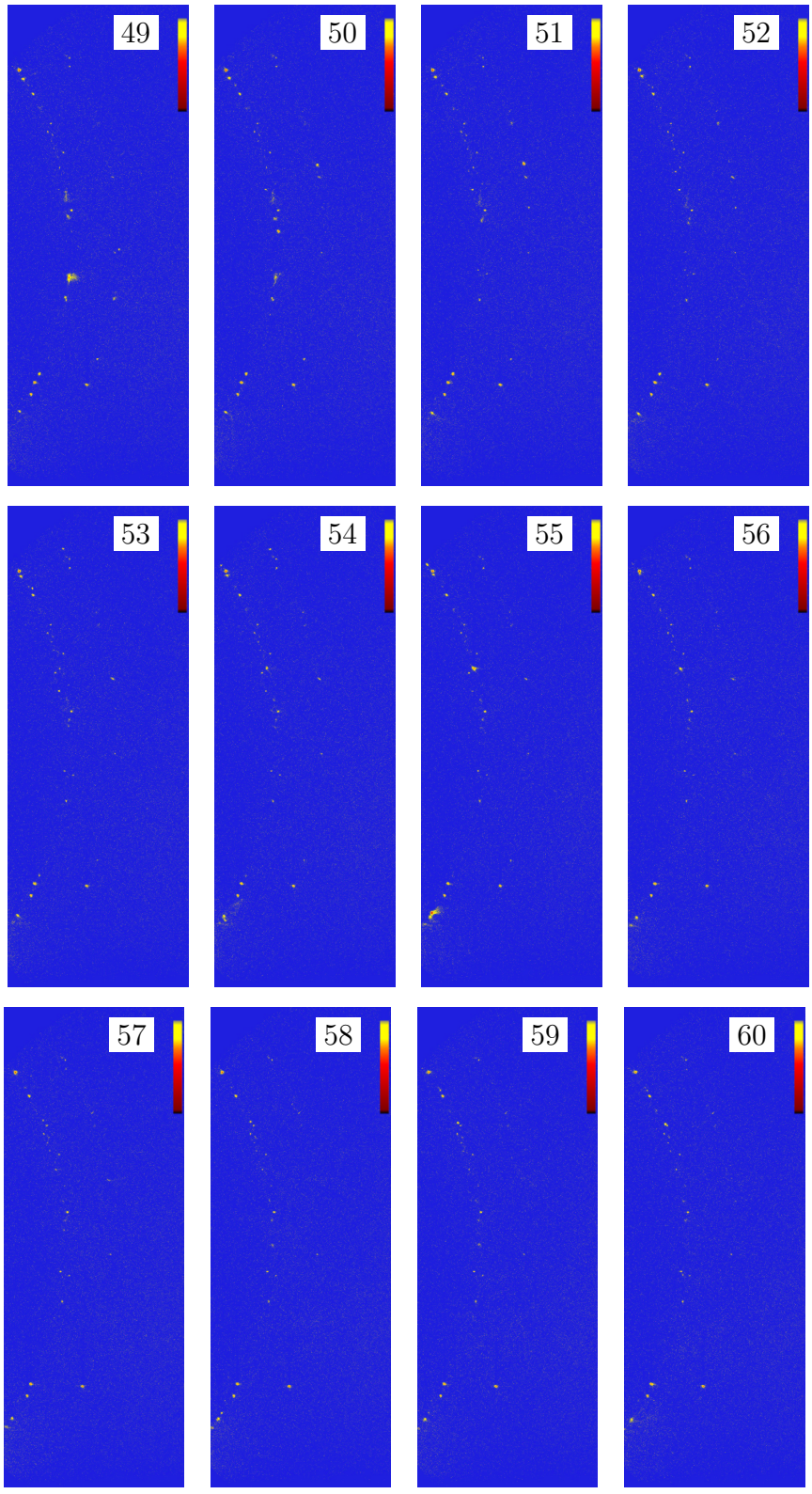


Figure A.5:

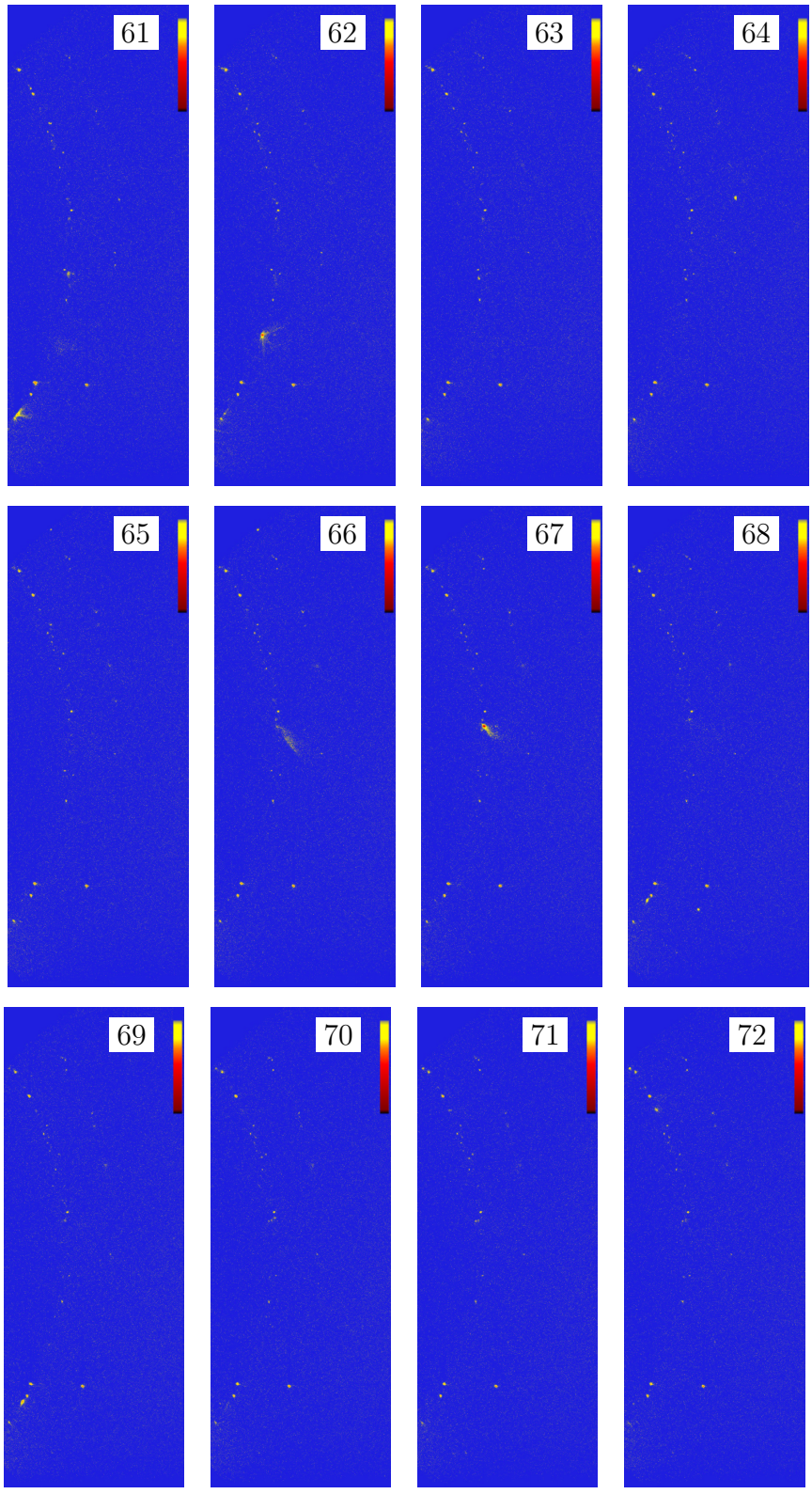


Figure A.6:

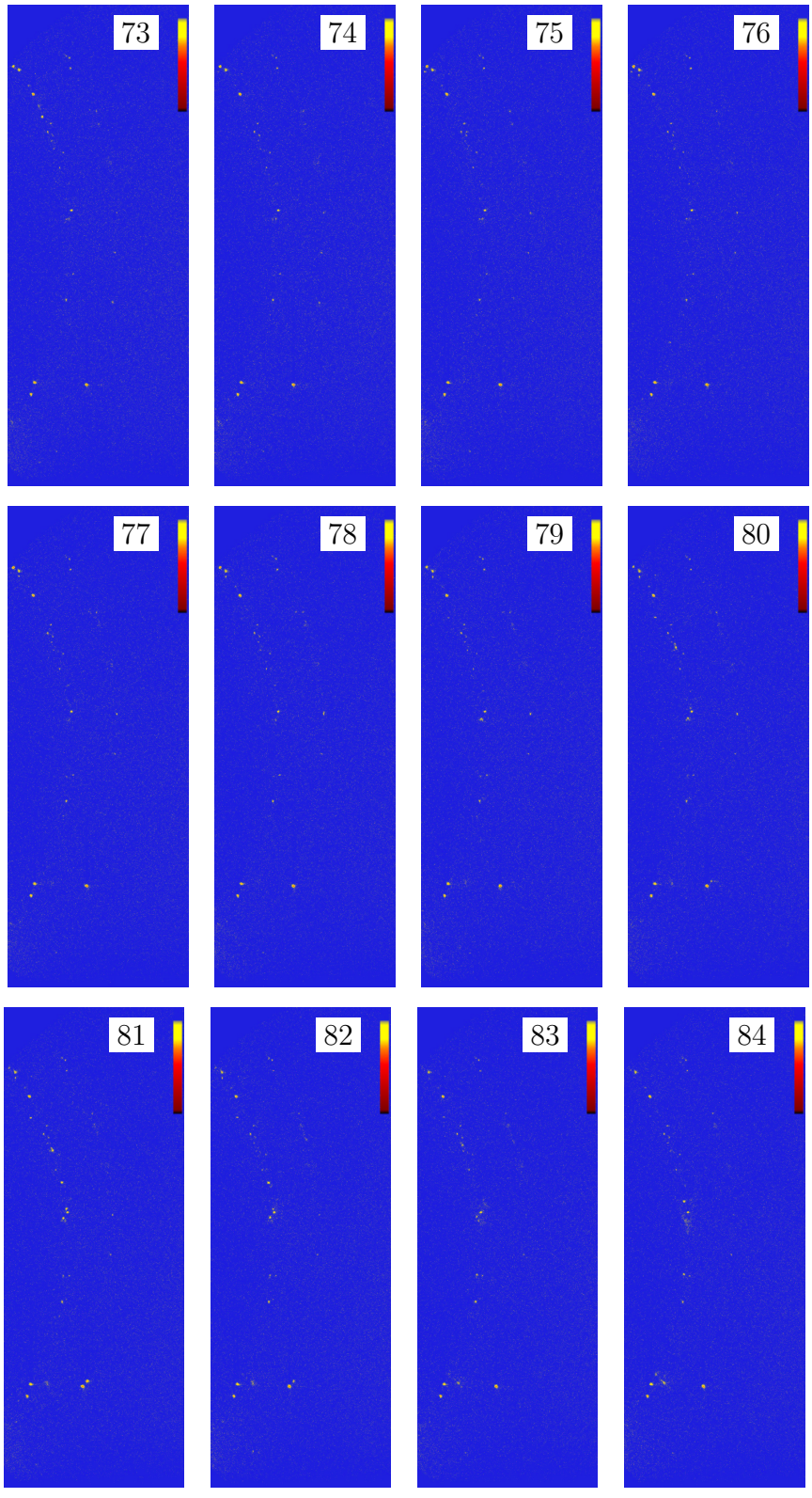


Figure A.7:

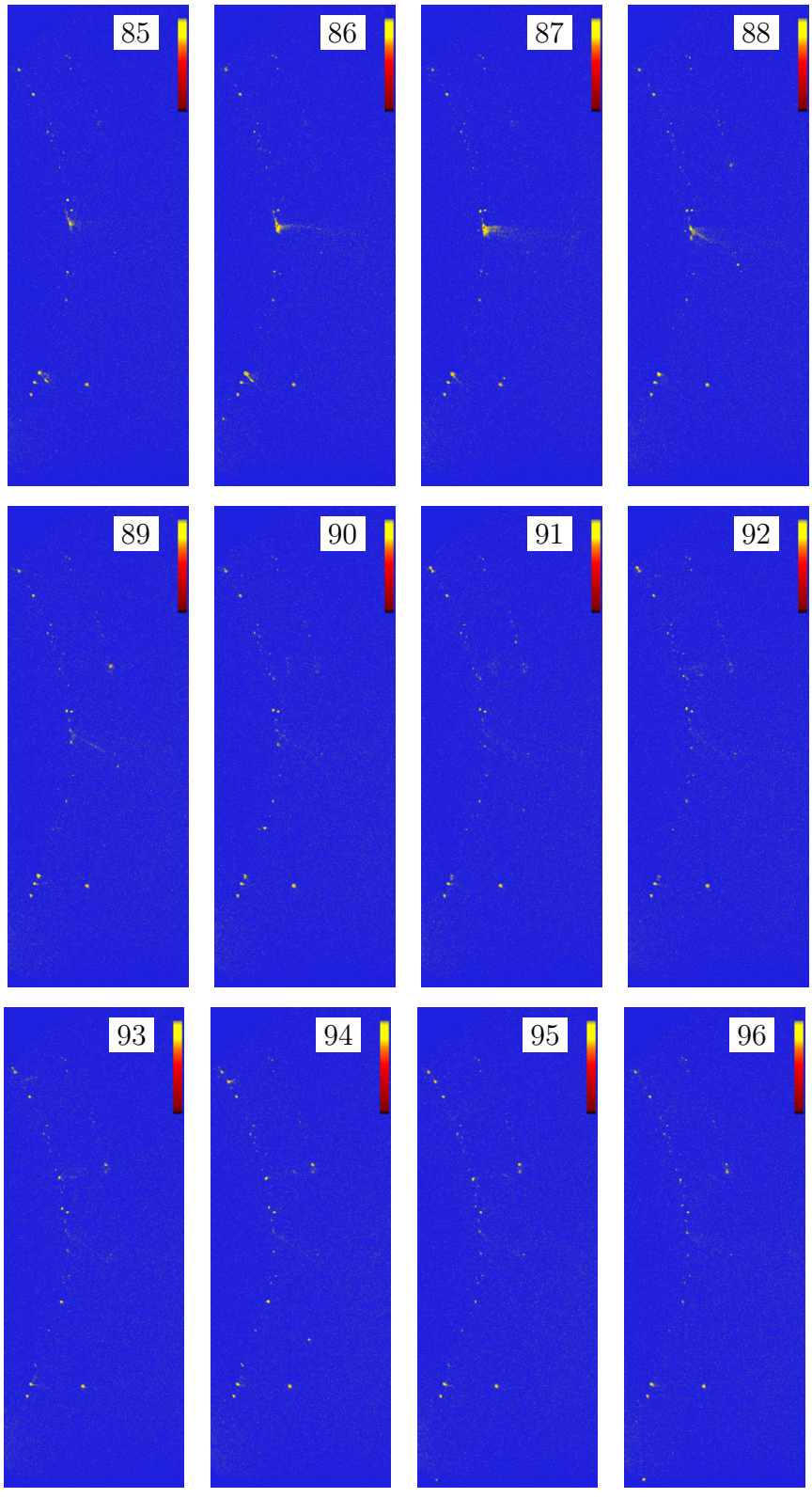


Figure A.8:

References

[1] Direct detection of X-rays (4 keV to $\sim$ 20 keV) using detectors based on deep-depletion CCD technology. Roper Scientific Technical Bulletin 2, 2000.

[2] Jens Als-Nielsen and Des McMorrow. *Elements of Modern X-ray Physics*. Wiley, 2001.

[3] Neil W. Ashcroft and N. David Mermin. *Solid State Physics*. Harcourt Brace, college edition, 1976.

[4] R. Barakat and G. Newsam. Necessary conditions for a unique solution to two-dimensional phase recovery. *J. Math. Phys.*, 25:3190–3193, 1984.

[5] R.H.T. Bates. Fourier phase problems are uniquely solvable in more than one dimension. I: Underlying theory. *Optik*, 61:247–262, 1982.

[6] S. Boutet, I.K. Robinson, Z.W. Hu, B.R. Thomas, and A.A. Chernov. Surface relaxation in protein crystals. *Physical review E*, 66:061914, 2002.

[7] W.P. Burmeister. Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallographica*, D56:328–341, 2000.

[8] Gerald Burns and A.M. Glazner. *Space groups for solid state scientists*. Academic Press, 1990.

[9] M. Casselyn, S. Finet, A. Tardieu, and H. Delacroix. Time-resolved scattering investigations of brome mosaic virus microcrystals appearance. *Acta Cryst. D*, 58:1568–1570, 2002.

[10] A.A. Chernov. *Modern Crystallography III Crystal Growth*. Springer-Verlag, 1984.

[11] A.A. Chernov and H. Komatsu. *Principles of crystal growth in protein crystallization*. Kluwer Academic Publishers, 1995.

[12] P. Debye. *Ann. Physik*, 46:809, 1915.

[13] P.H. Duke. *Synchrotron Radiation; Production and Properties*. Oxford University Press, 2000.

[14] Eric Fanchon, Erik Geissler, Jean-Louis Hodeau, Jean-René Regnard, and Peter A. Timmins. *Structure and Dynamics of Biomolecules: Neutron and Synchrotron Radiation for Condensed Matter Studies*. Oxford, 2000.

[15] J.R. Fienup. Reconstruction of an object from the modulus of its Fourier transform. *Optics Letters*, 3:27–29, 1978.

[16] J.R. Fienup. Phase retrieval algorithms: a comparison. *Applied Optics*, 21:2758–2769, 1982.

[17] J.R. Fienup. Reconstruction of a complex-valued object from the modulus of its Fourier transform using a support constraint. *J. Opt. Soc. Am. A.*, 4:118–123, 1987.

[18] S. Finet, F. Bonneté, J. Frouin, K. Provost, and A. Tardieu. Lysozyme crystal growth, as observed by small angle x-ray scattering, proceeds without crystallization intermediates. *European Biophysics Journal*, 27:263–271, 1998.

[19] F.A. Fischbach, P.M. Harrison, and T.G. Hoy. The structural relationship between ferritin protein and its mineral core. *Journal of Molecular Biology*, 39:235–238, 1969.

[20] J.D. Ng J.A. Gavira and J.M. García-Ruíz. Protein crystallization by capillary counterdiffusion for applied crystallographic structure determination. *Journal of Structural Biology*, 142:218–231, 2003.

[21] K. Gekko. *Water relationships in food*. Plenum Press, 1991.

[22] A. George and W.W. Wilson. Predicting protein crystallization from a dilute solution property. *Acta Crystallographica*, D50:361–365, 1994.

[23] R.W. Gerchberg and W.O. Saxton. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237–246, 1972.

[24] T. Granier, B. Gallois, A. Dautant, B.L. Destaintot, and G. Precigoux. Preliminary x-ray diffraction studies of the tetragonal form of the native horse-spleen apoferritin. *Acta Crystallographica D*, 52:594–596, 1996.

[25] T. Granier, B. Gallois, A. Dautant, B.L. Destaintot, and G. Precigoux. Comparisons of the structures of the cubic and tetragonal forms of horse-spleen apoferritin. *Acta Crystallographica D*, 53:580, 1997.

[26] André Guinier and Gérard Fournet. *Small-Angle Scattering of X-Rays*. Wiley, 1955.

[27] A. P. Hammersley, S.O. Svensson, M. Hanfland, A.N. Fitch, and D. Hausermann. Two dimensional detector software: From real detector to idealised image of two-theta scan. *High Pressure Research*, 14:235–248, 1996.

[28] Pauline M. Harrison and Paolo Arosio. The ferritins: molecular properties, iron storage function and cellular regulation. *Biochimica et Biophysica Acta*, 1275:161–203, 1996.

[29] Eugene Hecht. *Optics*. Addison-Wesley, 1998.

[30] R. Henderson. Cryo-protection of protein crystals against radiation damage in electron and x-ray diffraction. *Proc. R. Soc. Lond. B*, 241:6–8, 1990.

[31] S.B. Hendricks and E. Teller. *Journal of Chemical Physics*, 10:147, 1942.

[32] O. Hignette, P. Cloetens, W.-K. Lee, W Ludwig, and G. Rostaing. Hard X-ray microscopy with reflecting mirrors status and perspectives of the ESRF technology. *Journal De Phsysique IV: JP*, 104:231–234, 2003.

[33] W. Humphrey, A. Dalke, and K. Schulten. VMD - Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14.1:33–38, 1996.

[34] Gene E. Ice, Eliot D. Specht, Jonathan Z. Tishler, Ali Khounsary, Lahsen Assoufid, and Chian Liu. At the limit of nondispersive micro and nanofocusing mirror optics. *Proceedings of the SPIE*, 5347:1–8, 2004.

[35] J.D. Jackson. *Classical Electrodynamics*. Wiley, 3th edition, 1998.

[36] S.H. Kilcoyne, G.R. Mitchell, and R. Cywinski. Temperature dependant SANS from ferritin and apoferritin. *Physica B*, 180-181:767–769, 1992.

[37] R. Klein and B. D'Aguanno. *Light Scattering: Principles and Developement*. Clarendon Press, 1996.

[38] Tom Kupp, Basil Bank, Alex Deyhim, Curtis Benson, Ian Robinson, and Paul Fuoss. Development of a double crystal monochromator. *AIP Conference Procedings*, 705:651–654, 2004.

[39] J.J. Larsen, K. Hald, N. Bjerre, and H. Staplefeldt. Three dimensional alignment of molecules using elliptically polarized laser fields. *Phys. Rev. Lett.*, 85:2470–2473, 2000.

[40] W. Haußler, A. Wilk, J. Gapinski, and A. Patkowski. Interparticle correlations due to electrostatic interactions: A small angle x-ray and dynamic light scattering study. I Apoferritin. *Biochimica et Biophysica Acta*, 1164:331–334, 1993.

[41] H.Y. Li, M.A. Perozzo, J.H. Konnert, A. Nadarajah, and M.L. Pusey. Determining the molecular packing arrangements on protein crystal faces using atomic force microscopy. *Acta Cryst. D*, 55:1023–1035, 1999.

[42] P.J. Loll. Membrane protein structural biology : the high throughput challenge. *Journal of Structural Biology*, 142:144–153, 2003.

[43] S. Marchesini, H. N. Chapman, S. P. Hau-Riege, R. A. London, A. Szoke, H. He, M. R. Howells, H. Padmore, R. Rosen, J. C. H. Spence, and U. Weierstall. Coherent X-ray diffractive imaging: applications and limitations. *Opt. Express*, 11:2344–2353, 2003.

[44] D.W. Marquardt. An algorithm for least-square estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11:431–441, 1963.

[45] Alexander McPherson. *Crystallization of Biological Macromolecules*. Cold Spring Harbor, 1999.

[46] D.A. McQuarrie. *Statistical Mechanics*. University Science Books, 2000.

[47] A. Mhaikesar, M.J. Kazmierczak, and R. Banerjee. Three-dimensional numerical analysis of convection and conduction cooling of spherical biocrystals with localized heating from synchrotron x-ray beams. *Journal of Synchrotron Radiation*, 12:318–328, 2005.

[48] J. Miao, H.N. Chapman, J. Kirz, D. Sayre, and K.O. Hodgson. X-ray pulses and diffraction microscopy of cells with synchrotron radiation. *Annu. Rev. Biophys. Biomol. Struct.*, 33:157–179, 2004.

[49] J. Miao, P. Charalambous, J. Kirz, and D. Sayre. Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature*, 400:342–344, 1999.

[50] J. Miao, T. Ishikawa, B. Johnson, E.H. Anderson, B. Lai, and K.O. Hodgson. High resolution 3D X-ray diffraction microscopy. *Phys. Rev. Lett.*, 89:088303, 2002.

[51] J. Miao, D. Sayre, and H. N. Chapman. Phase retrieval from the magnitude of the Fourier transforms of nonperiodic objects. *J. Opt. Soc. Am. A.*, 15:1662–1669, 1998.

[52] L. Michaelis. Ferritin and apoferritin. *Adv. Prot. Chem.*, 33:53–66, 1947.

[53] R.P. Millane. Phase retrieval in crystallography and optics. *J. Opt. Soc. Am. A*, 7:394–411, 1990.

[54] R.P. Millane. Multidimensional phase problems. *J. Opt. Soc. Am. A.*, 13:725–734, 1996.

[55] R.P. Millane and W.J. Stroud. Reconstructing symmetric images from their undersampled Fourier intensities. *J. Opt. Soc. Am. A.*, 14:568–579, 1997.

[56] V.N. Morozov and T.Y. Morozova. Viscoelastic properties of protein crystals - triclinic crystals of hen egg-white lysozyme in different conditions. *Biopolymers*, 20:451–467, 1981.

[57] V.N. Morozov, T.Y. Morozova, E.G. Myachin, and G.S. Kachalova. Metastable state of a protein crystal. *Acta Crystallographica B*, 41:202, 1985.

[58] Colin Nave and Elspeth F. Garman. Toward and understanding of radiation damage in cryocooled macromolecular crystals. *Journal of Synchrotron Radiation*, 12:257–260, 2005.

[59] Colin Nave and M.A. Hill. Will reduced radiation damage occur with very small crystals? *Journal of Synchrotron Radiation*, 12:299–303, 2005.

[60] Mark A. Pfeifer. *Structural studies of lead nanocrystals using coherent X-ray diffraction.* PhD thesis, University of Illinois at Urbana-Champaign, 2005.

[61] John A. Pitney. *Coherent X-ray diffraction.* PhD thesis, University of Illinois at Urbana-Champaign, 2000.

[62] D. Pontoni and T. Narayanan. Ultra-small angle X-ray scattering studies of attractive colloidal transitions. *Journal of Applied Crystallography*, 36:787–790, 2003.

[63] E. Prince, editor. *International tables for crystallography.* Kluwer, 2004.

[64] R.B.G. Ravelli, P. Theveneau, S. McSweeney, and Martin Caffrey. Unit-cell volume change as a metric of radiation damage in crystals of macromolecules. *Journal of Synchrotron Radiation*, 9:355–360, 2002.

[65] F. Reif. *Fundamentals of Statistical and Thermal Physics.* McGraw-Hill, 1965.

[66] T. Ritz, S. Park, and K. Schulten. Kinetics of excitation migration and trapping in the photosynthetic unit of purple bacteria. *Journal of Physical Chemistry B*, 105:8259–8267, 2001.

[67] I. K. Robinson. Crystal truncation rods and surface roughness. *Phys. Rev. B*, 33:3830–3836, 1986.

[68] I.K. Robinson, P.J. Eng, C. Romainczyk, and K. Kern. *Surface Science*, 1367:105–112, 1996.

[69] I.K. Robinson and I.A. Vartanyants. Use of coherent X-ray diffraction to map strain fields in nanocrystals. *Appl. Surf. Sci.*, 182:186–191, 2001.

[70] I.K. Robinson, I.A. Vartanyants, G.J. Williams, M.A. Pfeifer, and J.A. Pitney. Reconstruction of the shapes of gold nanocrystals using coherent X-ray diffraction. *Phys. Rev. Lett.*, 87:195505, 2001.

[71] F. Rosenberger, P.G. Vekilov, M. Muschol, and B.R. Thomas. Nucleation and crystallization of globular proteins - what we know and what is missing. *Journal of Crystal Growth*, 168:1–27, 1996.

[72] D. Sayre. Some implications of a theory due to Shannon. *Acta Cryst.*, 5:843, 1952.

[73] D. Sayre. The squaring method: a new method for phase determination. *Acta Cryst.*, 5:60–65, 1952.

[74] R.V. Sharma and K.C. Sharma. The structure factor and the transport properties of dense fluids having molecules with square well potential, a possible generalization. *Physica A*, 89:213–218, 1977.

[75] J.-E. Shea and C.L. Brooks. From folding theories to folding proteins: A review and assessment of simulations studies of protein folding and unfolding. *Ann. Rev. Phys. Chem.*, 52:499–535, 2001.

[76] E.H. Snell and J.R. Helliwell. Macromolecular crystallization in microgravity. *Reports on Progress in Physics*, 68:799–853, 2005.

[77] J.C.H. Spence, M. Howells, L.D. Marks, and J. Miao. Lenless imaging: A workshop on 'new approaches to the phase problem for non-periodic objects. *Ultramicroscopy*, 90:1–6, 2001.

[78] R. Strange, S. Morante, S. Stefanini, E. Chiancone, and A. Desideri. Nucleation of the iron core occurs at the three-fold channels of horse-spleen apoferritin: an EXAFS study on the native and chemically-modified protein. *Biochimica et Biophysica Acta*, 1164:331–334, 1993.

[79] T. Teng and K. Moffat. Primary radiation damage of protein crystals by an intense synchrotron X-ray beam. *Journal of Synchrotron Radiation*, 7:313–317, 2000.

[80] B.R. Thomas, D. Carter, and F. Rosenberger. Effect of microheterogeneity on horse spleen apoferritin crystallization. *Journal of Crystal Growth*, 187:499–510, 1998.

[81] UNICAT. 34-ID floor plan. http://www.uni.aps.anl.gov.

[82] Friso van der Veen and Franz Pfeiffer. Coherent x-ray scattering. *Journal of Physics: Condensed Matter*, 16:5003–5030, 2004.

[83] I.A. Vartanyants, J.A. Pitney, J.L. Libbert, and I.K. Robinson. Reconstruction of surface morphology from coherent X-ray reflectivity. *Phys. Rev. B*, 55:13193–13202, 1997.

[84] I.A. Vartanyants and I.K. Robinson. Imaging of quantum array structures with coherent and partially coherent diffraction. *Journal of Synchrotron Radiation*, 10:409–415, 2003.

[85] P.G. Vekilov. Dense liquid precursor for the nucleation of ordered solid phases from solution. *Crystal Growth and Design*, 4:671–685, 2004.

[86] B.E. Warren. *X-ray Diffraction*. Dover Publications, Inc., 1990.

[87] M. Weik, R.G.B. Ravelli, G. Kryger, S. McSweeney, M.L. Raves, M. Harel, P. Gros, I. Silman, J. Kroon, and J.L. Sussman. Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proceeding of the National Academy of Science*, 97:623–628, 2000.

[88] Garth J. Williams. *Microscopy of gold microscrystals by coherent X-ray diffractive imaging*. PhD thesis, University of Illinois at Urbana-Champaign, 2005.

[89] G.J. Williams, M.A. Pfeifer, I.A. Vartanyants, and I.K. Robinson. Three-dimensional imaging of microstructure in Au nanocrystals. *Phys. Rev. Lett.*, 90:175501, 2003.

[90] Z. Wulff. *Zeitschrif für Kristallographie*, 34:449, 1901.

[91] S.-T. Yau and B.R. Thomas P. G. Vekilov. Molecular mechanisms of crystallization and defect formation. *Physical Review Letters*, 85:353–356, 2000.

[92] S.-T. Yau and P. G. Vekilov. Direct observation of nucleus structure and nucleation pathways in apoferritin crystallization. *Journal of the American Chemical Society*, 123:1080–1089, 2001.

[93] J. M. Zuo, I. Vartanyants, M. Gao, and L. A. Nagshan. Atomic resolution imaging of a carbon nanotube from diffraction intensities. *Science*, 300:1419–1421, 2003.