Developing Methods towards the Structure Determination of Biological Particles using Crystallographic and Single Particle Imaging Techniques

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften
am Fachbereich Chemie der Fakultät für Mathematik,
Informatik und Naturwissenschaften
Universität Hamburg

vorgelegt von

Sabine Botha

Februar, 2018

Die vorliegende Arbeit wurde im Zeitraum von April 2015 bis Februar 2018 in der Arbeitsgruppe von Prof. Dr. C. Betzel im Laboratorium für Strukturbiologie von Infektion und Entzündung am Institut für Biochemie und Molekularbiologie des Fachbereichs Chemie der Universität Hamburg durchgeführt.

1. Gutachter: Prof. Dr. Christian Betzel
2. Gutachter: Prof. Dr. Andrew Torda

Datum der Disputation: 13.04.2018

Contents

Methodenweiterentwicklung für das Abbilden einzelner biologischer Partikel durch kohärente Diffraktion und Methodenentwicklung für serielle Kristallografie

Zusammenfassung

Das ultimative Ziel der Strukturbiologie ist es, einzelne, biologische Partikel, wie zum Beispiel Viren, Ribosomen oder sogar Proteinmoleküle, mit atomarer Auflösung abzubilden, ohne die Proben mit aufwendigen Verfahren vorbereiten oder kristallisieren zu müssen. Die Entwicklung und Nutzung von Freien Elektronenlasern ist eine notwendige Bedingung, um dieses Ziel zu erreichen. In den letzten zehn Jahren wurde die Methodenentwicklung zur Einzelmolekülanalyse biologischer Objekte mittels kohärenter diffraktiver Röntgenbildgebung bereits zielorientiert vorangetrieben. Alle Verfahren basieren auf dem sogenannten „Diffraktion-vor-Zerstörung" Prinzip. Durch ultra-kurze, höchst brillante femtosekunden-Röntgenpulse können Diffraktionsdaten aufgenommen werden, bevor Strahlenschäden überhaupt entstehen. Die hohe Strahlensensitivität organischer Materie stellte bisher eine fundamentale Limitierung für die Analyse biologischer Objekte, wie zum Beispiel mithilfe von Röntgenstrukturanalyse oder Elektronenmikroskopie, dar. Die Messungen am FEL finden im Vakuum statt und setzen somit voraus, dass die zu analysierende Probe in Lösung eine homogene Verteilung aufweist und ohne Lösungsmittelrückstände in einer nahezu wasserfreien Umgebung gemessen werden kann.

Im Rahmen dieser Arbeit wurden zwei Virusproben ausgewählt, die für eine weitergehende Methoden- und Algorithmenentwicklung geeignet sind. Es handelt sich um dsRNA-Mycoviren und dsDNA-Adenoviren, die charakterisiert und für zukünftige FEL-Experimente vorbereitet wurden. Untersuchung durch Röntgenkleinwinkelbeugung war hier ein wesentlicher Bestandteil.

Die neu etablierte Methodik der seriellen Kristallografie basiert ebenfalls auf dem Prinzip der „Diffraktion-vor-Zerstörung". Mithilfe der seriellen Kristallografie können am FEL Röntgenlaser sogar Nanokristalle von biologischen Proben vermessen wer-

den, die für das konventionelle Einkristallverfahren mit Synchrotronstrahlung nicht mehr zugänglich sind. Im Gegensatz dazu wird bei der seriellen Methode lediglich ein Diffraktionsbild pro Kristall aufgenommen, und die Diffraktionsbilder von tausenden, einzelnen Kristallen zu einem kompletten Datensatz zusammengefügt. Diese Methode ist inzwischen an FEL Röntgenlasern gut etabliert und darüber hinaus auch für Messungen mit konventioneller Synchrotronstrahlung adaptiert.

Ein weiterer wesentlicher Vorteil der seriellen Kristallografie besteht darin, dass Messungen bei Raumtemperatur, also unter eher physiologischen Bedingungen, durchgeführt werden können. Messungen bei Raumtemperatur ermöglichen zukünftig weiterhin zeitaufgelöste Studien an Proteinen und Enzymen. Allerdings sind solche Experimente, vor allem an Synchrotronstrahlungsquellen, welche zugänglicher für Messzeiten sind als FEL Röntgenlaser, noch nicht hinreichend etabliert, obwohl sich insbesondere moderne Speicherringe für zeitaufgelöste Studien im mikro- und millisekunden Bereich gut eignen.

Das sogenannte Phasenproblem der Kristallografie muss allerdings auch bei dem seriellen Ansatz experimentell gelöst werden. Im Rahmen dieser Arbeit wurde hierzu eine Methode entwickelt, bei der Schwermetall-Ionen besonders schonend in Proteinmikrokristalle über Diffusion eingefügt werden können. Die Mikrokristalle liegen nach Anwendung dieser Methode bereits in einer viskosen, lipidisch-kubischen Phase als Trägermaterial vor und können so direkt für die Messungen genutzt werden. Insbesondere Membranproteine, wichtige Angriffspunkte bei der strukturbasierten Wirkstoffentwicklung, kristallisieren häufig ausschließlich in Form von Mikrokristallen in einer lipidisch-kubischen Phase, da diese ihre natürliche hydrophobe Umgebung artifiziell nachbildet.

Des Weiteren sind in Mikrokristallen die Diffusionszeiten für Substrate und Liganden stark reduziert. Dieses ist für zeitaufgelöste, diffusionsabhängige Enzym-Substrat Studien von Vorteil, weil die erforderliche Diffusionszeit deutlich kürzer sein muss, als die für den Substratumsatz benötigte Zeit. Um Diffusionszeiten in Abhängigkeit von der Kristallgröße zu bestimmen, wurden hier entsprechende Untersuchungen durchgeführt.

Abschließend konnte erfolgreich ein Substrat in einem Proteinkristall quervernetzt werden, die Methodik wurde anschließend auf Mikrokristalle übertragen und seriell unter-

sucht. Somit konnte ein ausgewählter Enzym-Substrat-Komplex mit Synchrotronstrahlung unter Anwendung eines seriellen Ansatzes analysiert werden.

Developing Methods towards the Structure Determination of Biological Particles using Crystallographic and Single Particle Imaging Techniques

Abstract

Structural biologist strive towards the ability of imaging single biological particles such as viruses, ribosomes or even single protein molecules with atomic resolution without the need for tedious sample preparation or crystallisation. Organic matter is however highly sensitive to radiation damage and therefore severely limits the imaging capabilities of available techniques, such as electron microscopy or X-ray diffraction analysis. Since the advent of free electron lasers (FELs), this has become theoretically realizable, and therefore methods towards using coherent diffractive imaging approaches using an FEL have been undergoing development during the past decade. The theory is based upon the "diffraction before destruction" principle, which states that when using an ultra-short, yet highly brilliant X-ray pulse, as generated at an FEL, the diffraction pattern from a single biological particle can be collected before the radiation damage induced by the X-rays traverses and obliterates the sample. These experiments, however, require the experiment to take place under vacuum conditions, and therefore the particles to be analysed need to be highly uniform and homogeneous in solution ans subsequently dried free of remnants, since thousands of individual diffraction patterns in random orientations are merged into a 3D diffraction volume.

To this end, two virus samples which were deemed suitable for method and algorithm development, were characterised and prepared for single particle coherent diffractive imaging experiments at an FEL during the course of this work. The analysed viruses were a dsDNA-Adenovirus, and a ds-RNA Mycovirus. Furthermore small angle X-ray scattering experiments were performed at a synchrotron, and first structural insights into the viruses could be obtained.

A further application of the "diffraction before destruction" approach is serial crys-

tallography. The most commonly used method for solving protein structures is macromolecular crystallography. The long established rotational method poses the barrier of having to grow large protein crystals. For serial crystallography approaches, crystals that are too small for conventional, rotational X-ray diffraction structure solution can be applied. Instead of collecting one complete data set from a single crystal, only a single diffraction pattern is collected per crystal which are then merged into a complete dataset. The successful application of this method for structure solution has been demonstrated numerous times at FELs as well as initial proof of principle experiments at synchrotron X-ray sources.

In addition to the benefits of being able to apply small crystals, data collection is also done at room temperature, which is usually not possible applying single crystal diffraction due to the implications this has for the increase in radiation damage. This makes it possible for scientists to probe proteins close to physiological conditions, and sets the path for time-resolved studies of the proteins. Such experiments are however not yet fully established at synchrotron X-ray sources, despite accessibility for beamtime being far more readily available than at FEL X-ray sources. Furthermore, synchrotron X-ray sources are highly suitable for achieving time resolutions of micro- or milliseconds.

The crystallographic phase problem also needs to be solved experimentally when applying serial crystallography. Through the work done in the course of this thesis, a method for gently introducing heavy atoms into micrometer sized crystals via diffusion was developed. This method negates the need for handling of the subsequent highly fragile, derivatized crystals by incorporating them directly into the viscous, lipidic cubic phase carrier material used to inject the crystals into the X-ray beam. Particularly membrane proteins, common drug targets, do not lend themselves to conventional macromolecular crystallography techniques, due to the fact that they often crystallise in showers of micro-crystals in a lipidic mesophase, as it mimics the native membrane environment.

Due to the associated small crystal size used for serial crystallography experiments, the diffusion pathways through the crystals are decreased. Therefore this method also lends itself to mixing experiments, where a substrate or ligand is diffused into a protein crystal and is then turned over by the individual protein molecules without destroying the

crystal structure. Towards the possibility of successfully applying this concept, temporal analysis of diffusion studies were conducted during the course of this thesis.

Finally, a substrate could be successfully cross-linked into the substrate recognition site of the model protein in an attempt to visualise intermediate states during substrate turnover in the protein. This was then further applied serially, and a selected enzyme-substrate-complex was analysed applying serial crystallography.

Chapter 1

Introduction

In 1901 the first Nobel Prize in Physics was awarded to the German Physicist Wilhelm Conrad Röntgen for the revolutionary discovery of a new type of invisible radiation emanating from a cathode-ray tube. The emitted, penetrating radiation could be detected using a fluorescent screen and for the first time the inner structure of the human body was visualised [73]. Today almost every person is familiar with the concept of X-ray imaging, and the technique has been very widely and extensively explored, well beyond the medical application of using X-rays to image the bones inside of the human body, as Röntgen's first experiments entailed in 1895. Thirteen years after the first Nobel Prize was awarded, a further physicist, Max von Laue, received the physics Nobel Prize for his work with X-rays, specifically for the discovery of their diffraction by crystals [31]. However, Laue's diffraction patterns could only be interpreted later by the Australian born Sir William Lawrence Bragg, who, working in his father's X-ray laboratory, was the first person to interpret the diffraction as *reflections* of a discrete lattice plane, formulating the fundamental law of diffraction theory known as **Bragg's Law**:

$$n\lambda = 2d_{hkl}sin(\theta) \tag{1.1}$$

Unsurprisingly, he jointly received the Nobel Prize in Physics with his father in 1915, and is to date still the youngest physics Nobel Prize laureate [71].

Biomolecular X-ray Crystallography has developed into one of the largest fields for the structure determination of macromolecules in structural biology, with most structural

biology laboratories having in-house X-ray sources that operate using the same principles for X-ray generation as discovered by Röntgen so many years ago. The X-rays are produced by bombarding materials with electrons, knocking core level holes into the electron shell of the atoms. When electrons from the outer shell fall back into these electron holes, they produce X-rays with a wavelength characteristic for the material used. Technical advances have rendered these sources with fluences that enable crystallographic studies of protein crystals in the tens of micrometers range. The major drawback of these sources however is the non-tunable wavelength, as the energy of the X-rays is a characteristic of the material used in the source. Especially when direct phasing methods need to be employed, it is essential to measure near the absorption edge of the corresponding heavy atoms in order to get the maximum anomalous signal for subsequent phase reconstruction. Therefore, synchrotron radiation sources have become indispensable for the protein crystallography field since their advent in the 20th century. Electrodynamics predict that when electrons, positrons or ions are diverted from a linear path onto a circular orbit the electrons experience a centripetal acceleration. In order for energy to be conserved, the electrons need to emit energy and they do this in the form of radiation, a phenomenon known as *Bremsstrahlung*. The first circular electron accelerator exploiting this effect for generating X-rays was built by a further Nobel Prize laureate, Ernest O. Lawrence, in 1929 in Berkeley California. In modern synchrotrons, electrons are accelerated to near relativistic speeds, which results in the emitted X-ray radiation being constrained to the plane of the electron orbit. In these dedicated facilities, an electron gun produces electrons that are then pre-accelerated before being introduced onto the circular path of the so-called electron storage ring. The electrons are forced to maintain this circular motion through *bending magnets*, and any energy lost by the conversion is re-supplied using radio frequency generators. This allows stable operations, and today synchrotron radiation facilities usually run stable operation for 24 hours a day, 7 days a week with the the facilities merely interrupting user operation for machine maintenance, checks and improvements. Synchrotrons produce a polychromatic ”white” X-ray beam, but scientists have developed methods for cutting out a narrow, monochromatic frequency band of X-rays at a variable range of desired wavelengths as required for crystallographic methods. Furthermore, 3rd generation synchrotron light sources offer X-ray beams with

a brilliance allowing data collection from crystals of only a few micrometers in size. However, the downside of these high brilliances is the associated radiation damage induced in the sample by the X-ray beam. Therefore, the benefits in using free electron lasers, X-ray radiation sources that produce highly brilliant yet very short X-ray pulses, for crystallography applications was identified. In an X-ray free electron laser (XFEL) the electrons generated by an electron gun are accelerated linearly close to the speed of light before being introduced into a so-called wiggler. This wiggler consists of a series of magnets with alternating poles, forcing the electrons onto a sinusoidal, ”wiggling” path, transverse to the direction of the beam. Similarly to synchrotron X-ray generation, when being diverted from their linear path, the electrons emit monochromatic photons. Using the appropriate optics and undulator gaps this incoherent beam of photons interferes with itself to form standing electromagnetic waves that are strong enough to interact with the ”wiggling” electron beam, modulating the electron density into longitudinally clustered ”micro bunches”. By design these micro bunches are separated by one optical wavelength and hence the photons emitted by the electron bunches are now in phase, resulting in a coherent X-ray beam. This process is known as self-amplified spontaneous emission. These X-ray radiation sources produce X-ray pulses of a few femtoseconds duration, with peak brilliances exceeding those of a synchrotron radiation source by an order of magnitude. Exploiting these X-ray characteristics have lead to the development of single particle coherent diffractive imaging (SPI) and serial femtosecond crystallography (SFX). Both methods are based on the principle of ”diffraction before destruction” and will be introduced in detail in the following sections.

X-ray radiation sources have undergone tremendous development during the past decades due to their wide applicability in science, and in particular for structural biology they have paved the way for gaining insights into the structures of viruses, DNA, ribosomes and proteins, the building blocks of life. However, given the high complexity of these macromolecular structures, it took a further fifty years after coining Bragg’s law for the field of biomolecular crystallography applying X-rays to reach its first milestone, with Max Perutz solving the first protein structure, haemoglobin [87]. Applying this technique after thirty years of work, he was consequently awarded the the Nobel Prize in chemistry in 1962. Today, X-ray crystallography has become one of the most common

techniques in this ever evolving field of structural biology and serial crystallography has become an established technique for collecting diffraction data from crystals that are too small for conventional, rotational crystallography.

1.1 Sample Preparation for Single Particle Imaging

X-ray free-electron laser (XFEL) radiation sources have expanded the realms of imaging biological particles, and hold the promise of using single-shot coherent diffractive imaging for imaging bioparticles at atomic resolution. Identical copies of a single biological particle are streamed across the X-ray beam, and a single diffraction pattern is collected per particle before the particle is affected by radiation damage and ultimately destroyed. Cryogenic electron microscopy is today well established but limited to cryogenic temperatures and therefore does not lend itself to studying biological systems in their real, native environments nor can it handle time-resolved studies. Therefore, the only approach that holds the promise of imaging single biological particles in their native environment and at room temperature is ultra fast coherent diffractive imaging with free-electron laser radiation. If successful, this could revolutionise structural biology, making many of the currently used, expensive and tedious structure solution methods based on repetitive structures, such as protein crystallography, obsolete. Using the ultra-bright, femtosecond duration X-ray pulses produced by an XFEL, it is believed that radiation damage to the biological sample can be outrun, meaning that the resulting diffraction pattern can be collected before the sample is ultimately destroyed by the X-rays [64]. Theory suggests that at 1 keV the absorption for carbon is saturated at a fluence of 10^9 photons/$(100\text{nm})^2$, however at 8 keV this corresponds to a fluence of 10^{12} photons/$(100\text{nm})^2$ [64]. Once this state is reached, inelastic collisions lead to secondary ionisation processes in the sample [53]. This process results in inner-shell electron vacancies, with electrons from the outer shells filling in through Auger or fluorescence decays. Furthermore, for small particles, higher fluences are required for achieving reasonable signal contrast of the delivered photons, with 10^{12}-10^{13} photons/$(100\text{nm})^2$ being desirable [2]. Scientists believe that the only way of overcoming the sample being damaged by the X-rays is by ”outrunning” the damage, i.e. shortening the X-ray pulses to below the typical Auger lifetimes of a few femtoseconds. This theoretical idea has coined the term ”diffraction before destruction” [64], and is the basis of single particle imaging as well as serial crystallography approaches (which will be discussed in the next section).

Single particle coherent diffractive imaging has been successfully applied at FLASH,

DESY [15] and at the Linac Coherent Light Source (LCLS) in California. Proof of principle, single-shot coherent diffractive imaging experiments have been successfully applied over the past years and these single particle imaging (SPI) techniques have been used to image bioparticles such as viruses [81], bacteriophages [41], cell organelles [36] and cyanobacteria [97]. However, so far this imaging approach has been limited to particles of sizes above 100 nanometers [48]. This is for one due to the possible, achievable resolution applying SPI at the LCLS currently being limited to a few tens of nanometers [2] and various other associated technical, as well as software challenges that are currently being addressed but still need to be overcome. With the European XFEL facility haven taken up operation in September 2017 and the SPB/SFX beamline dedicated to single particle imaging experiments, it is promising to revolutionise this technique with a flexible set-up allowing imaging up to a resolution of 1 Å (verbal communication with Dr. M. Messerschmidt, SPB instrument scientist). It remains to be seen whether this can be implemented successfully once the machine operation has been optimised. A further bottleneck is getting the sample into the X-ray beam. The detectors are located in vacuum, and ideally the particles are streamed through the X-ray beam in single file. To date the most commonly employed method for achieving this is through an aerodynamic lens stack [81, 41, 36, 97]. The general experimental set-up as employed at most of the aforementioned XFEL experiments is shown in Figure 1.1.

An aqueous suspension of the biological particles is aerosolised using a nebuliser or a gas dynamic virtual nozzle (GDVN) operated in the droplet regime, generating droplets of roughly the particle size. This is essential for guaranteeing a single, as opposed to multiple, particles per drop, which would result in clusters that are not useful for single particle imaging. The droplets containing the single particles are then allowed to dry off before being focused into a particle beam using an aerodynamic lens stack. The focus of this particle beam is subsequently aligned with the focus of the X-ray beam and the diffraction patterns are collected using two pnCCD detectors. The X-ray focus is ideally a few nanometers in size and particle beam is focussed down to the same size as the particles to be imaged. Therefore it is crucial for the beamline to operate very stably. Consequently, alignment is a very tedious task, usually requiring multiple beam time shifts for establishing good overlap. Furthermore, the particles need to be dry, as excess water will form

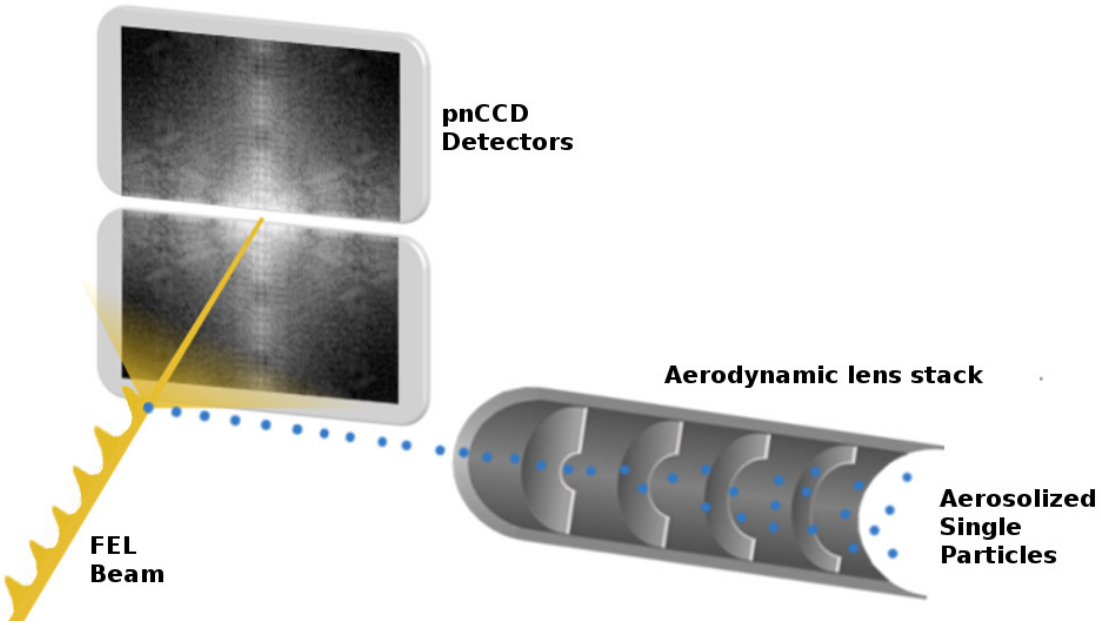


Figure 1.1: General Set-up for Single Particle Imaging Experiments
Schematic of the general experimental set-up for standard single particle imaging experiments at a free electron laser. An aerodynamic lens stack is used for injecting the aerosolised particles into the X-ray beam. The diffraction pattern is collected on a set of pnCCD detectors.

a hydration shell adding unnecessary background to the weak diffraction patterns. Additionally, this also disturbs the size distribution of the particles, introducing even more difficulty to the ongoing algorithm development for retrieving the relative orientations of the particles. Adding to this is the fact that biological samples are most commonly stabilised in a physiologically similar environment using buffers. These are standardly made up of various salts which result in salt crusts surrounding the particles when imaging. Not only does this again interfere with the noise and size distribution, they alter the surface morphology of the virus or particle, rendering the diffraction pattern useless for solving the structure. Therefore, it is essential to try and stabilize the biological particles in water or preferably even in a volatile solution. Thousands of single diffraction patterns in different orientations are required for assembling a full data set for the particle being imaged, meaning that not only is an abundance of the sample required (especially considering the currently achievable hit rate of only a few percent in the best cases), but the particles also have to be perfectly uniform and identical in solution. Therefore, the particles have to be

selected carefully, and the sample homogeneity needs to be confirmed and/or established well in advance prior to imaging at an upcoming beam time. Lastly, a lot of time and effort from scientists all over the world has gone into developing algorithms to orient the diffraction patterns, assembling them into 3D diffraction volumes [51, 54, 41] and performing subsequent phase retrieval [56, 55, 25]. Despite some successes to date, this is by no means standardised and is still undergoing constant development.

In terms of the research activities presented here, two possible virus samples were analysed and prepared for single particle imaging using an X-ray free electron laser, a ds-RNA mycovirus and a doubly inactivated adenovirus. The proposed virus investigations will be focused on a particular ds-RNA mycovirus, which is known to infect the fungus *Fusarium graminearum*, the major causal agent of a worldwide cereal (such as wheat and barley) disease, known as fusarium head blight and is therefore highly relevant in combating this disease. Given the significant impact that depleted wheat stocks have on third world countries, the benefit of gaining the ability to control this disease is evident. The structure of this virus is to our knowledge unknown, therefore the proposed investigations could provide first insights into the structure-function-relationship. The other virus sample, doubly inactivated adenovirus (approx. 100nm diameter), propagated in HEK293 cells, was chosen due to its surface being covered with thin fibrils, making it an interesting sample for algorithm development.

1.2 Serial Crystallography

Serial femtosecond crystallography (SFX) at XFEL radiation sources and serial millisecond crystallography (SMX) at synchrotron radiation sources are relatively new approaches of collecting diffraction data for the structure determination of biological macromolecules. Methods, software and procedures have been undergoing continuous development over the last decade [16, 13, 78, 18]. Suspensions of micro or nano crystals are streamed across the X-ray beam at a free electron laser (XFEL) or high intensity synchrotron radiation source and diffraction patterns are collected in a "diffraction before destruction" approach [64]. The diffraction patterns of thousands of individual crystals in random orientations are collected and the intensities integrated using a Monte-Carlo approach [44, 102]. The main benefits of this method for collecting crystallographic data of biological particles is threefold:

Firstly, conventional crystallographic approaches require protein crystals of at least a few tens of micrometers in size, and often crystals have to be grown to a few hundred micrometers to yield diffraction sufficient for structure solution at high resolution. Growing large protein crystals is often a very tedious and laborious task, so the option of collecting just one diffraction pattern from a micrometer sized crystal and subsequent merging of the data into a 3D diffraction volume as opposed to collecting a whole dataset from a single crystal can be very beneficial. Secondly, the data does not suffer from radiation damage despite utilizing extremely bright pulses due to the "diffraction before destruction" approach and the fact that only one diffraction pattern is collected per image. Thirdly, and also linked to the past two points, small crystals can now be probed at room temperature. Radiation damage is approximately 70 times more severe at room temperature than at cryogenic temperatures [63], and therefore when applying conventional crystallography approaches, crystals are flash-cooled using liquid nitrogen and kept at approximately 100 K during data collection. For every protein crystallisation condition found, cryo-protection needs to be optimised as the protein crystals are very sensitive to freezing and having crystalline ice in the sample can have a severely negative impact on the data quality and the associated completeness of the data. Additionally, the further need to handle the crystals when fishing, cryo-protecting and flash freezing often nega-

tively impacts the integrity of the crystal lattice and sometimes the crystals are even lost completely during handling. Finally, having the crystals in their native, room temperature environment paves the road to time-resolved protein studies with various new possibilities becoming available to structural biology. Smaller crystals mean smaller diffusion pathways, and associated higher spatial and temporal resolution when utilizing mixing for substrate diffusion experiments [79]. Furthermore, pump-probe experiments can be done on a single crystal basis, and photoinisation of all the molecules in a crystal at the same time is sharpened with respect to time when fewer molecules make up the sample volume. This is also highly beneficial when employing caged compounds to trigger a biochemical response of the protein molecules, as for these experiments timing constraints in both diffusion as well as photo-activation have been a severe bottleneck to date. Therefore a lot of resources have been dedicated over the past decade into developing and establishing the concept, methods and software for serial crystallography.

One of the hurdles that has been addressed extensively over the past decade is how to produce a steady stream of microcrystals passing through the X-ray beam. The initially developed and still extensively used method for SFX experiments is the delivery of the crystals via a liquid jet [16, 13, 99] using a gas dynamic virtual nozzle (GDVN) [23], as shown in Figure 1.2.

A GDVN uses gas flow (usually helium) to focus a liquid jet down to a few micrometers in size after being extruded through a capillary with an inner diameter of a few tens of micrometers. This method of liquid focusing makes the nozzles highly resilient and resistant to clogging, and in addition the supersonic expansion of the sheath gas keeps the sample in a liquid state for a few seconds upon being injected into the vacuum chamber. Therefore the crystals in the liquid suspension can be probed by the FEL in an aqueous environment. On the down side, liquid jet delivery requires a relatively high amount of crystal suspension with a large number of crystals going to waste without contributing a diffraction pattern. Despite this it is still the most commonly employed delivery method, since it is robust and particularly lends itself to pump-probe experiments, given the rapid replenishing of fresh sample to the interaction region. Therefore, many room temperature, time-resolved, photo-activated pump-probe experiments have been conducted in this way [3, 7, 43, 42, 47, 93], as well as the majority of the initial proof of principle

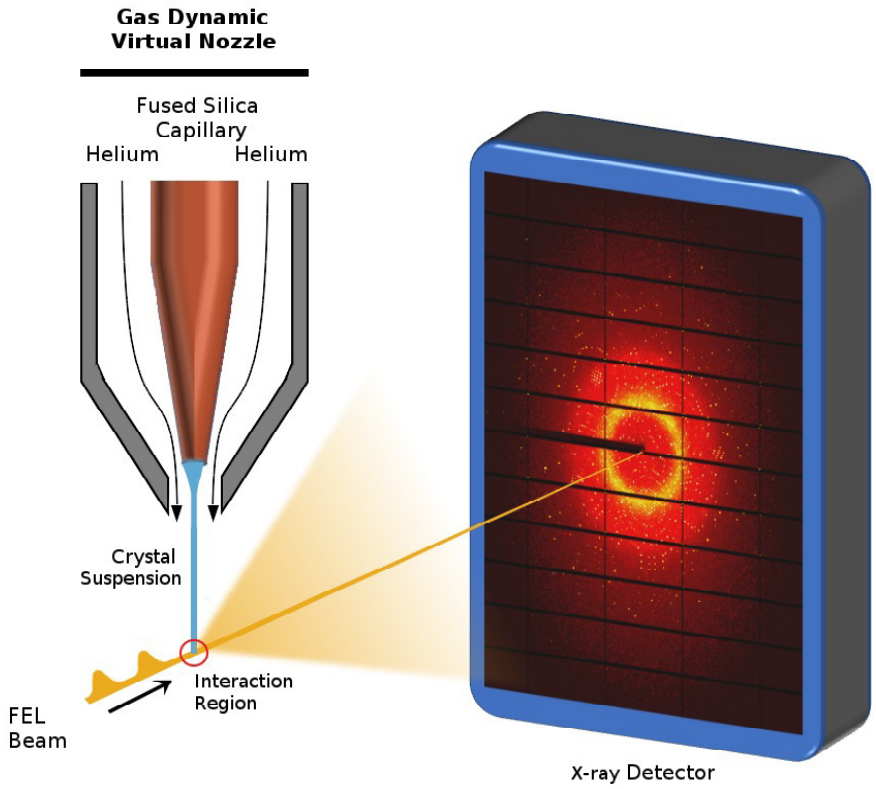


Figure 1.2: Set-up for Serial Crystallography Diffraction Experiments
Schematic of the general experimental set-up for standard serial femtosecond crystallography experiments using a gas dynamic virtual nozzle for sample injection into the free electron laser X-ray beam.

SFX experiments ranging from lysozyme experiments [52, 13, 5, 27, 86] to ribosomal subunits [22], *in vivo* grown protein crystals [68] and bacterial cells [77], amongst many others. In an attempt to reduce sample consumption, many other sample delivery methods into the FEL have been underway, including an electro-spinning flow focused jet [84] or acoustic droplet injection [70, 85]. Since these are not of interest to the work done in this thesis, they will not be discussed further. A further method that has been extensively researched for reducing the sample consumption however that has also found wide applicability due to its adaptability to synchrotron operation is the implementation of severely slowing down the jet by embedding the samples into a viscous carrier medium such as a lipidic cubic phase (LCP) [100], vacuum grease [90] or agarose gels [19]. A further major drawback to using a GDVN as the delivery method are the shortcomings for synchrotron radiation sources due to the jet speed of many m/s not rendering the crystals in

the beam long enough to yield useful diffraction as well as having been developed for in vacuum application. Subsequently the viscous, slow-jet delivery method was found to be easily adaptable to atmospheric operation for synchrotron radiation beamlines and it was shown that serial diffraction data collected from crystals in the micrometer size range at a synchrotron have sufficient quality to be used for phasing calculations. Serial crystallography has therefore been adapted to synchrotron application during the past years [11, 94, 33, 88]. The ability to collect room temperature crystallographic data opens up the possibility of time resolved studies, as bio-macromolecules can now be probed close to their native environments at a synchrotron radiation source and using crystals several magnitudes smaller than required for conventional methods. It has further been shown that exposure times of less than 3 ms are sufficient for collecting serial room-temperature data at a synchrotron radiation source, for example using a lysozyme microcrystal suspension [94]. SFX diffraction data are of sufficient quality for *de novo* phasing approaches [8, 62, 105, 60, 35, 104]. However systematic inaccuracies and variances in the data resulting from experimental factors, such as the wide spectral distribution of the XFEL, as well as altering sample to detector distance when exchanging the sample delivery nozzle and variations in the size of the microcrystal distribution or liquid jet width, have been shown to severely hamper phasing attempts applying heavy atom data sets [62]. These factors are however not as present for serial data collected at a synchrotron, but *de novo* phasing methods using SMX data hat not yet been widely explored, with only one proof of principle study [94] having been conducted to date.

Multiple software packages have been developed for online data monitoring during collection [57], such as crystal hit rate monitoring, subsequent off-line hit finding [9, 28], as well as for indexing and integrating the diffraction intensities from the thousands of randomly oriented crystals traversing the X-ray beam [102, 101, 37, 76, 40]. In CrystFEL [102, 101], the software used in this thesis, crystal hits are identified and indexed using MOSFLM [67] and DIRAX [24], and subsequently the intensities are integrated using a Monte-Carlo Integration approach [44, 45] and post-refined to account for the partiality of the respective reflections. The integrated intensities can then be converted to a format compatible with all the well established software suites used for phase retrieval and refinement using conventional, rotational macromolecular crystallography techniques.

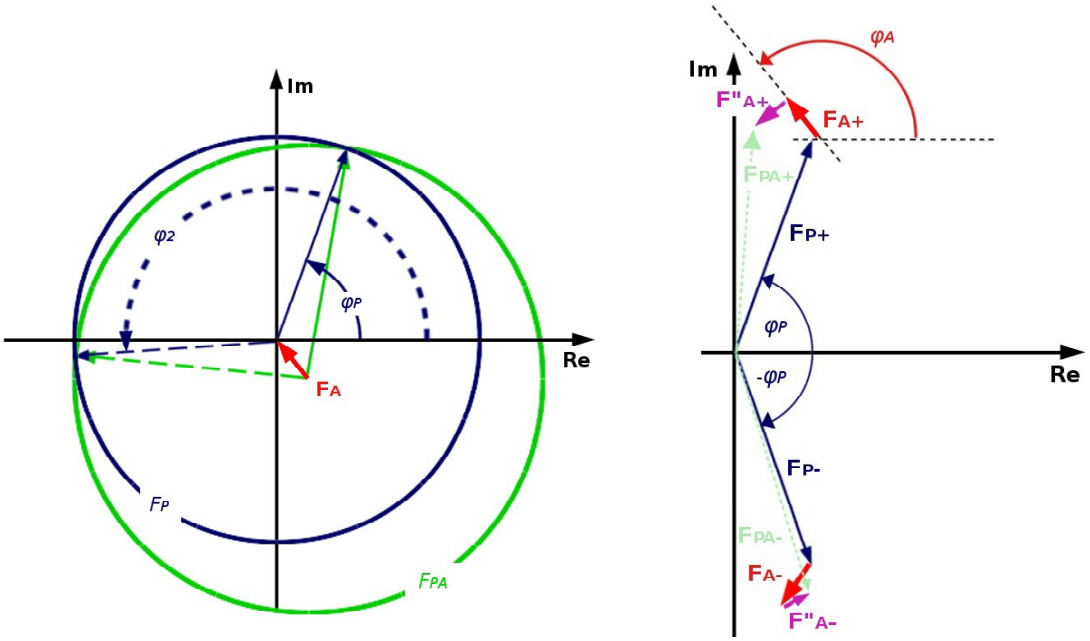
1.3 Single Isomorphous Replacement with Anomalous Scattering (SIRAS)

When collecting diffraction data, only the intensities of the diffracted X-rays are recorded and the phases are lost, which is commonly known in crystallography as the phase problem. Therefore, an essential task in macromolecular crystallography for solving the structure is phase retrieval. Today, the most common method employed is *molecular replacement*, a comparatively simple method where the lost phase information is simply substituted with the phases calculated from a similar, homologous structure. This method is however very biased towards this model structure, and more importantly, is only applicable to structures where there is a homologous structure available. Therefore, experimental phasing methods for phase retrieval are required for solving new, unknown structures and the underlying concepts will briefly be introduced here. Particular focus will be directed towards SIRAS phasing, as this is the method used for the work described in this thesis.

For SIRAS phasing, two datasets are required. Firstly, the magnitudes of the structure factors for the native protein crystals ($\mathbf{F_P}$) are collected and then an additional dataset of a heavy atom derivative of the protein is recorded ($\mathbf{F_{PA}}$) at the absorption edge of the heavy atom. The relationship of the structure factors themselves is shown in Equation 1.2 where $\mathbf{F_A}$ are the complex structure factors of the heavy atom. The positions of the heavy atoms can be calculated from the difference data of F_{PA} and F_P. These provide differences and, provided the native and derivative crystals are isomorphous, the positions of the (few) marker atoms can be calculated using Patterson methods [71]. Therefore the complex structure factor $\mathbf{F_A}$ is known. This is termed marker atom substructure phasing, and is based on the differences introduced in the electron density of the protein by introducing marker atoms.

Figure 1.3a shows the relationship between these complex structure factors for a generic, acentric reflection *hkl* and how the phase can be reconstructed, provided the structure factors of the heavy atom, as well as the magnitudes of the structure factors for the protein and its derivative are known.

However, as can be seen from Figure 1.3a, this gives rise to two possible solutions



(a) Harker diagram showing the phase reconstruction applying single isomorphous replacement

(b) Vector relationship between the structure factor contributions for a *Bijvoet* pair of reflections

Figure 1.3: Harker Diagram for Single Isomorphous Replacement and Structure Factor Relationship for the Anomalous Scattering Contribution
a) Harker diagram, showing how the protein phase, φ_P for a generic, acentric reflection can be reconstructed from the structure factors of the heavy atom $\mathbf{F_A}$ and the magnitudes of the structure factors of the protein F_P and its heavy atom derivative F_{PA}. b) Vector relationship of the anomalous scattering contribution of a *Bijvoet* pair **F+"** and **F-"**.

for the protein phase, φ_P and φ_2. Therefore, in SIRAS phasing, anomalous scattering is employed in addition to single isomorphous replacement. Heavy atoms scatter X-rays anomalously if irradiated at an absorption edge of the respective heavy atom. This gives rise to intensity differences between two reflections of a *Bijvoet* pair. Therefore anomalous scattering data contains pairs of reflections that have different intensities. This anomalous scattering contribution is imaginary and therefore always delayed by 90° relative to the component obtained from single isomorphous replacement, as shown in Equation 1.2 [71]:

$$\boldsymbol{F}_{PA} = \boldsymbol{F}_P + \boldsymbol{F}_A + i \cdot \boldsymbol{F}_A'' \tag{1.2}$$

where $\mathbf{F_A}$**"** is the anomalous contribution to the structure factor. The anomalous com-

ponents $\mathbf{F_{A+}}$'' and $\mathbf{F_{A-}}$'' therefore represent the structure factors of a *Bijvoet pair*. However, similarly to above, only the magnitudes of the structure factors can be measured, therefore only F_{PA+} and F_{PA-} are known. The relationship between all the complex structure factors is shown in Figure 1.3b. Given the 90° phase difference of the single isomorphous replacement and anomalous difference heavy atom structure factors, the phase ambiguity can be resolved by combining the two methods (Figure 1.4).

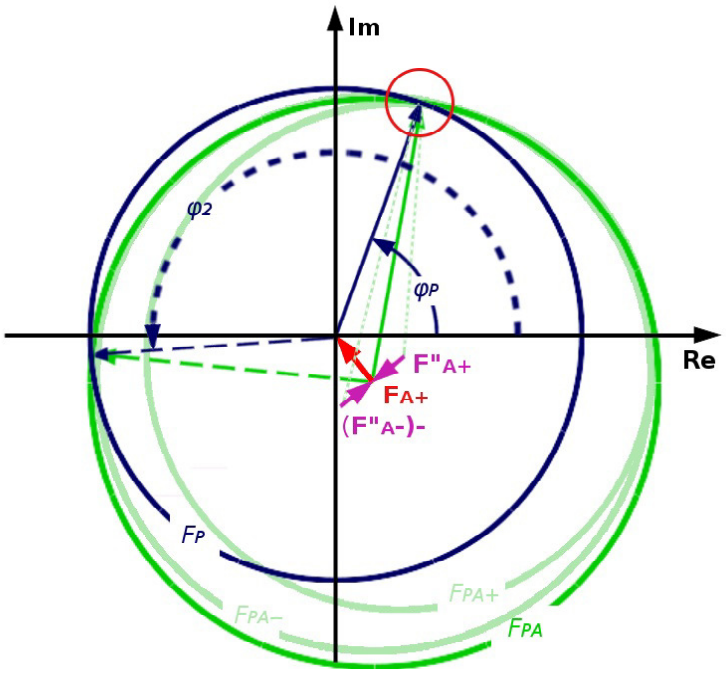


Figure 1.4: Harker Diagram Demonstrating the Principle of Solving the Phase Ambiguity applying SIRAS
Harker diagram, showing how the protein phase, φ_P for a generic, acentric reflection can be reconstructed from the structure factors of the heavy atom $\mathbf{F_A}$ and the magnitudes of the structure factors of the protein F_P and its heavy atom derivative F_{PA}. Perpendicular to $\mathbf{F_A}$, $\mathbf{F_{A+}}$'' and $\mathbf{F_{A-}}$'' represent the anomalous contributions to the heavy atom structure factors. F_{PA+} and F_{PA-} are the magnitudes of the anomalous scattering factors of the heavy atom protein derivative. The correct phase can be identified where all the circles representing the structure factor magnitudes overlap.

To overcome the aforementioned inaccuracies when phasing SFX data, a large number of diffraction patterns is required and Yamashita et al. (2015) demonstrated that single-wavelength anomalous diffraction (SAD) phasing with approximately 80 000 patterns as well as SIRAS phasing with approximately 20 000 patterns is feasible. Nass et al. (2016)

showed that by systematically addressing some of the experimental inaccuracies during SFX data processing, SAD phasing applying gadolinium was possible with only approximately 10 000 patterns in the case of lysozyme and native sulphur SAD was successfully applied to phase the crystal structure of thaumatin with approximately 125 000 images. In a different study Nakane et al. (2016) systematically reduced the number of images used for successful SIRAS phasing of the membrane protein bacteriorhodopsin, using an iodine-labelled detergent heavy atom additive and determined the minimal number of SFX patterns required for obtaining meaningful phases. They show that approximately 12 000 patterns were sufficient for successful SIRAS phasing and a further reduction of required images was achieved by a combination of single isomorphous replacement (SIR) and SIRAS [60]. Therefore, SIRAS phasing has been shown to be the method of choice for reconstructing the phase information for SFX data, it has however not yet been applied to SMX data. The only study where SMX data were successfully phased *de novo* applied multiple isomorphous heavy-atom replacement using anomalous scattering (MIRAS) phasing [94], and therefore in the work presented in this thesis the first successful SIRAS phase reconstruction of SMX data will be presented.

Chapter 2

Motivation and Outline

In this thesis, the mycovirus and adenovirus samples were extensively characterised regarding monodispersity and homogeneity in preparation for potential single particle imaging experiments. They were further assessed in various buffer compositions with the aim of finding a low to preferably no salt condition in which they remained stable or possibly even a volatile buffer. The sample preparation, as well as the classification will be presented in this thesis, however unfortunately it was not possible to perform single particle coherent diffractive imaging experiments at an XFEL due to instrumentational difficulties during the beamtime. In addition to preparing these two virus samples as well as classifying them using DLS, TEM, Nanosight and SAXS, due to their unknown structure crystallisation trials were performed in parallel to preparing the samples for FEL SDI beam time. Unfortunately no crystals large enough for diffraction data could be obtained.

The main part of this thesis aims at applying and contributing to developing serial crystallography techniques that have been adapted for both FEL and synchrotron data collection. Firstly, the possibilities and limitations of retrieving *de novo* phases of serially collected synchrotron data were further investigated and a novel heavy atom soaking method for microcrystals for serial collection, using slow lipidic cubic phase injection into the X-ray beam at a synchrotron, was developed. It was found that diffraction data collection of a single derivative was sufficient for *de novo* structure determination of the model system proteinase K. Experimental phases were determined by single isomorphous replacement with anomalous scattering (SIRAS). The resulting electron density maps were

of good quality, allowing for automatic model building of the entire structure, albeit the two terminal residues. Furthermore, the crystals used in this study were all grown natively and the derivative diffraction data were collected by introducing these crystals into lipidic cubic phase containing mercury. Applying this procedure was sufficient for reaching an occupancy of 0.4 for both possible (mutually exclusive) mercury sites after minutes of in situ soaking and proved to be a very gentle way of obtaining derivative crystals, with almost no variation in the unit cell constants of the derivative crystals compared to the native cell dimensions. A lipidic cubic phase injector was used to stream proteinase K micron-sized crystals embedded in lipidic cubic phase across the X-ray beam in a serial crystallography approach. Data were collected in continuous, shutterless operation at room temperature at the PETRA III synchrotron radiation source. The mercury derivative data were collected by supplementing the lipidic cubic phase with mercury before embedding native crystals and streaming them across the beam, making the need to handle the derivatized and usually fragile crystals obsolete. The data were subsequently phased using SIRAS and the structure could be built using automatic model building tools.

Further, experiments towards developing time-resolved serial crystallography at room temperature were performed. For the selected model system, proteinase K, diffusion controlled substrate mixing was investigated to gain insights into the timing constraints implied by this method. Therefore, the inhibitory compounds boric acid and phenylmethane sulfonyl fluoride were soaked into variously sized crystals for various time points, and subsequently analysed crystallographically. The occupancies of the bound inhibitory compounds were then refined isotropically, and it was investigated whether any time dependence could be determined. Analogously, a mercury cage was synthesized, with the intention of investigating whether a triggered release of mercury could be achieved, and whether this was faster than the simple diffusion of mercury into the protein crystal. Should the time constants of both approaches be equal, this could determine that there may be no advantage for using caged compounds in small crystals for time-resolved, mixing studies. Finally, different substrates were screened for further substrate (not inhibitory) studies, and the oxidized chain B of insulin was identified to be a suitable candidate. Efforts were then made to slow down the enzymatic reaction time of proteinase K until an electron density of the insulin peptide could be observed in the active site of

proteinase K. This was successfully achieved using cross-linking techniques for larger crystals collected using standard rotational crystallographic techniques. A further attempt utilising microcrystals in a similar fashion and collecting diffraction data in a serial approach at the PETRA III, P11 beamline was attempted, but the clear density observed for the larger crystal could not be reproduced in this manner.

Chapter 3

Materials and Methods

3.1 Chemicals

All chemicals used for the work in this thesis were acquired in analytical grade from the companies *AppliChem, Fluka, Merck, Thermo Scientific, Roth, Serva* or *Sigma Aldrich* if not directly specified in the methods below. For a complete list of all the chemicals used along with their specific hazards and precautions, please refer to Appendix A.1.

3.2 Consumables

Table 3.1: Consumables Overview
Overview of the consumables used for the work in this thesis along with the respective manufacturers.

Description	Manufacturer
Amicon Ultra 4, Ultra 15, Ultra 0.5, MWCO 3 kDa, 10 kDa	Merck Millipore
Dialysis Tubing	Roth
Syringes 1 mL, 5 mL, 10 mL, 20 mL	VWR
Syringe filters	VWR
96-well SWISSCI MRC2 plates	Hampton Research
96-well Vapour Batch Diffusion Plates	Douglas Instruments
24-well Linbro Plates	Jena Bioscience
Highly Viscous Vacuum Grease	GE Bayer Silicones

The consumables used for the work in this thesis, including Eppendorf tubes, pipette

tips, serological pipettes, and Falcon tubes, were obtained from the company *Sarstedt*. All other consumables are listed in Table 3.1 along with the respective manufacturers.

3.3 Equipment

A complete list of all the equipment used for the work in this thesis along with the respective manufacturers can be found in Appendix B.

3.4 General Sample Characterisation and Sample Preparation Techniques

3.4.1 Buffer and Solution Preparation

Buffers and solutions were prepared using de-ionized water from a *SG Wasseraufbereitung und Regenerierstation GmbH* cartridge. Buffer pH was measured using a SevenEasy (*Mettler-Toledo*) pH-meter and monitored while adjusting. All buffers and solutions were filtered through a 0.2 µm filter (*Sartorius Stedim*) prior to use.

3.4.2 Sample Concentration

Samples were concentrated using an Amicon Ultra-0.5 centrifugal filter unit with a molecular weight cut-off (MWCO) of 3 kDa or 10 kDa as appropriate. The particle solutions were then centrifuged at 4 °C at 4 000 g until the desired concentration was reached. Protein concentrations were verified photometrically using a *Thermo Fischer* Nanodrop device which uses the specific absorption of aromatic amino acids at a wavelength of 280 nm to calculate the concentration according to equation (3.1).

$$c = \frac{A_{280} M_w}{\varepsilon d} \tag{3.1}$$

In the above equation A_{280} is the measured specific absorption at 280 nm, M_w the molecular weight of the protein, d the sample depth and ε the extinction coefficient of

the protein at 280 nm. A list of the parameters used for the individual proteins can be found in Table 3.2.

The concentration for the virus samples was always used as a relative value and not an absolute one, as the molecular weight had not yet been determined (see Section 5.1) and the amino acid sequence is not known.

Table 3.2: Protein Properties used for Concentration Determination
Protein molecular weights and extinction coefficients used for the protein concentration determination.

Protein	Molecular Weight [kDa]	Extinction Coefficient E1%
Proteinase K	28.9	14.3
Insulin	3.5	9.13
Mycovirus	-	1.0*
Adenovirus	-	1.0*

3.4.3 Dynamic Light Scattering

Dynamic Light Scattering Measurements were performed using the *SpectroSize 300* cuvette system from *Xtal Concepts*. This particular device has machine parameters of 100 mW laser output power, 660 nm wavelength, and scattered light is detected at a 90° angle. The refractive index was $n = 1.33$ (water) and unless stated otherwise the viscosity was taken to be $\eta = 1016\ cP$. The hydrodynamic radius R_H was calculated by the Xtal Concepts software from the decay time constant of the auto-correlation function, as described by the Stokes-Einstein-Equation (3.2), where k is Boltzmann's constant, T the temperature, and D the diffusion constant [4, 80].

$$R_h = \frac{kT}{6\pi\eta D} \tag{3.2}$$

An approximation of the particle molecular weight can additionally be calculated from the hydrodynamic radius using equation (3.3) [80],

$$R_h = \left(\frac{3M_w(V_p + V_h)}{4\pi N_A}\right)^{\frac{1}{3}} \tag{3.3}$$

where M_w is the molecular weight of the protein, V_p the specific particle volume, V_h the hydration volume and N_A Avogadro's constant.

Prior to performing a DLS measurement, samples were centrifuged at 16 000 *g* for at least 10 min, and 10-15 µL were pipetted from the top meniscus for analysis into a quartz cuvette. For standard measurements, 20 measurements were taken, each of 20 s duration and for long-term measurements a standard measurement was supplemented with further, less frequent measurements over the desired time frame. The temperature during all measurements was kept constant at 20°*C* using the built-in Peltier element.

3.4.4 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE is a method that separates biological macromolecules according to their molecular weight and electrical charge. Protein samples are pipetted into pockets of a vertically orientated polyacrylamide gel and then an electrical current is applied. Due to the varying sizes of individual molecules, the velocity at which they traverse the gel varies, and therefore they travel different distances along the gel in a specified amount of time.

Adding SDS to the gel results in all molecules acquiring a negative charge so that they are now only separated according to they respective molecular weight. The gel itself comprises two layers, the first of which is the stacking gel and is the layer onto which the proteins are applied. This smaller layer is used to focus the proteins and the subsequent separating gel through which the proteins are then separated according to their respective sizes. A list of the buffers used for doing the SDS-PAGE can be found in Table 3.3. Prior to being applied to the gel, 7.5 µL protein sample where mixed with 12.5 µL of 1x sample buffer and heated to 96°*C* for 5 min. The marker used was the *Unstained Protein Molecular Weight Marker* from *Fermentas*.

3.4.5 Thin-layer Chromatography (TLC)

Thin layer chromatography is a technique used to separate and visualize non-volatile mixtures. A drop of the mixture is applied to a plate of glass coated with aluminium oxide, known as the stationary phase. The glass plate is then stood upright in a container filled with a small volume of a solvent mixture, the mobile phase, which is drawn up through the plate due to capillary forces. The different components in the mixture to be analysed

Table 3.3: List of Buffers and Solutions used for SDS-PAGE
A complete list of all the buffers and solution compositions and pH values used for SDS-PAGE.

Buffer/Solution	Composition
1x Electrode buffer	192 mM Glycine 25 mM Tris-HCl 0.1% (w/v) SDS
Stacking Buffer	0.5 M Tris-HCl pH 6.8 0.4% (w/v) SDS
Separating Buffer	1.5 M Tris-HCl pH 8.8 0.4% (w/v) SDS
5x Sample Buffer	50 mM Tris-HCl pH 6.8 10% (w/v) SDS 50% (v/v) Glycerine 125 mM DTT 0.1% (w/v) Bromophenol blue
Coomassie Solution	0.25% (w/v) Coomassie Brilliant Blue G-250 25% (v/v) 2-Propanol 10% (v/v) Acetic Acid
De-staining Solution	20% (v/v) Acetic Acid

ascend the plate along with the mobile phase, but due to their differing solubility properties at different rates, whereby separation is achieved. The plate is then stained with different dyes to visualise the separate bands. TLCs were performed 10 min, 2 h, 2.5 h, 3 h and 4 h after initiating the reaction, or until the educt with the lower molar mass was no longer detectable.

3.5 Sample Preparation for Single Particle Imaging

The mycovirus samples were supplied courtesy of PD Dr. Cornelia Heinze from the Biozentrum Klein Flottbek and the adenovirus samples were kindly provided by Dr. Thomas Speiseder from the Heinrich Pette Institut.

3.5.1 Nanosight

Nanosight experiments were performed at the Universitätsklinikum Eppendorf, courtesy of Alexander Hartmann. The viruses were diluted to a concentration of 6 µg/mL and then subjected to 10 measurements using a Nanosight microscope set-up. For Nanosight experiments, the particles flowing through a liquid cell are illuminated with a laser and the scattered light is captured by a camera installed onto the optical microscope. The images are sent to the connected desktop computer and using tracking software the particle velocity, size and density is then calculated.

3.5.2 Transmission Electron Microscopy

The transmission electron microscopy (TEM) images of the mycovirus particles were performed at the Centre for Free Electron Laser Science with the assistance of Dr. Carolin Seuring and Thomas Seine. The sample was negatively stained for improved contrast of the images. A standard TEM copper grid from *Electron Microscopy Science* was glow discharged before applying 2 µL of virus solution and then left to incubate for at least 1 min before blotting off the excess liquid. A drop of 2% uranyl acetate staining solution was then prepared, onto which the grid was inverted and left to incubate for a further 30 s. The grids were then washed multiple times using ultra pure water and were subsequently imaged using a *FEI Tecnai G20* transmission electron microscope.

3.5.3 Small Angle X-ray Scattering

Small Angle X-ray Scattering is a well-established technique for acquiring structural information from particles in solution. The elastic scattering of macromolecules is measured for scattering angles of 0.1° to 10°. It is typically used to determine the molecular

weight of mono-disperse particle suspensions as well as for resolving the size and shape of macromolecules.

Samples were dialized overnight into water 24 h before the scheduled beamtime. This was done using dialysis tubing with a MWCO of 10 000 Da. The sample was subsequently concentrated and mono-dispersity of the sample was checked with DLS.

SAXS measurements were conducted at the P12 EMBL beamline at the PETRA III storage ring at the *Deutsches Elektronensynchrotron (DESY)* in Hamburg in July 2016. 20 µL of 0.01 mg/mL, 0.09 mg/mL and 0.17 mg/mL concentrated suspensions for the adenovirus samples and 0.10 mg/mL, 0.21 mg/mL and 0.42 mg/mL concentrated suspensions for the mycovirus samples were measured as well as interspersed blank measurement using the dialysis buffer to be subtracted from the data. The X-ray wavelength was 1.24 Å, the exposure period was 50 ms and the detector was positioned at a distance of 2 m. The data was processed using the software packages PRIMUS [46] and GNOM [91] from the ATSAS program suite [29]. These packages were used to calculate the Kratky- and Guinier plots from which the radius of gyration, R_G, as well as the maximal diameter, D_{max}, could be determined. *Ab initio* models were calculated using the packages DAMMIN [92] and DAMMIF [30], also from the ATSAS suite, or submitted to the online EMBL server.

3.5.4 Crystallisation Experiments

For initial crystallisation trials the mycovirus samples were applied using concentrations in the range of 4.0-12.0 mg/ml, and screened using the commercially available screens listed in Table 3.4.

Sitting drop vapour diffusion plates were set up in Maxiwell plates using the Honeybee robot, with 30 µL reservoir volume, and drop volumes of 300 nL protein with 300 nL reservoir solution. In parallel batch under oil plates were set up in Tegysaki plates, with drop volumes of 300 nL protein combined with 300 nL reservoir solution. The plates were then covered in Al's oil (1:1 mixture of silicon oil and paraffin oil). During the first two weeks the plates were checked daily for possible hits using an Olympus microscope, and after this the Al's oil was exchanged with paraffin oil and checked for crystal hits once a month.

Table 3.4: Crystallisation Screens
List of the commercially available crystallisation screens along with the respective manufacturers.

Screen	Supplier
$AmSO_4$-Suite	Qiagen, Germany
ComPAS-Suite	Qiagen, Germany
JCSG-plus	Molecular Dimensions, UK
MORPHEUS	Molecular Dimensions, UK
PACT premier	Molecular Dimensions, UK
Structure	Molecular Dimensions, UK
SturaFootprint&Macrosol	Molecular Dimensions, UK

Additives were also applied to the mycovirus using the additive screening kit from *Molecular Dimensions*. The monodispersity of the virus particles supplemented with the various additives were determined using DLS. Subsequently further crystallisation screening plates were set up with the mycovirus dialysed into the buffer and additive conditions determined to be the most stabilising. Additional plates were also set-up after the viruses had been dialized into water.

The adenovirus sample was applied as provided using concentrations in the range of 2.0-10.0 mg/ml and then screened using the kits listed in Table 3.4 in a sitting drop set-up as outlined above.

In both cases possible hits were tested under a fluorescence microscope, or tested for diffraction using our in house IμS X-ray source. Conditions that tested positive were further investigated in a Linbro plate hanging drop set up, using 1 mL reservoir volume, and 2 μL protein as well as reservoir solution drop volumes. The conditions giving positive hits were varied in pH, protein concentration and precipitant concentration to sample a larger area of the solubility phase diagram.

3.6 Serial Crystallography

3.6.1 Proteinase K Crystallisation

Proteinase K crystals were grown using the batch method, with initial conditions taken from Betzel et al. [10]. Protein was obtained from *Merck KGaA* and 10-30 mg/mL were dissolved in 50 mM Tris-HCl pH 7.0 containing 10 mM $CaCl_2$. 50-100 µL of protein were combined with 15-50 µL of precipitant, containing 2 M $NaNO_3$ and 50-100 µL of the protein buffer. Depending on the final concentrations of protein and precipitant, crystals grew in various sizes ranging from 2-5 µm to multiple mm sized crystals. Smaller crystals grew overnight, and larger crystals grew in 1-2 days. The crystal suspensions were washed with crystal storing solution consisting of 75% (v/v) buffer and 25% (v/v) precipitant immediately after crystallisation and then once a month. The differently sized crystal suspensions were stored at 20°C. For applications requiring single crystals, a drop of crystal suspension was pipetted onto a cover slide and single crystals were fished out and placed into an MRC plate well that had been prepared with 200 µL crystal storing solution in the reservoir and 2 µL crystal storing solution in the well. For cryogenic measurement, the wells were supplemented with 2 µL of cryo-protectant, consisting of crystal storing solution supplemented with 40% (v/v) glycerol.

3.6.2 SIRAS Phase Determination using LCP Soaking and Crystal Delivery

Crystal Suspension Preparation

LCP was mixed with DL-α-monoolein, 9.9 MAG, obtained from *Fuka* and crystal storing solution. Monoolein was mixed with crystal storing solution 60:40% (w/v) using coupled Hamilton syringes [17] and then the respective volumes were adjusted until a clear, isotropic phase formed. For the mercury derivative the crystal storage solution was supplemented with 1 mg/ml $HgCl_2$ prior to combining with the lipid. Immediately prior to injection 8 µL of crystal pellet (crystal size 5 x 10 µm) were embedded into 20 µL LCP as described in Botha et al [94] and the crystal containing LCP was then transferred to the sample reservoir of the injector [100].

Experimental set-up for LCP Crystal Delivery

The injector developed at the Arizona State University, USA, [100] was mounted using a *Thorlab* mounting stage so that the tip of the nozzle was immediately above the X-ray beam interaction region. A *Shimadzu 20AD* HPLC system was used to drive the injector and helium sheath gas was introduced at the nozzle tip to constrain the LCP from curling back on itself after extrusion. A nozzle with a 50 µm inner diameter (ID) capillary was used to inject the crystals, and the helium pressure was set to 240-470 psi and the flow rate on the HPLC pump varied from 2.5-4 µL/min with an associated pressure of 37-48 psi, corresponding to an LCP sample flow rate of 73-113 nL/min. The hit rate was monitored online using OnDA [57] and the interaction overlap was optimised by moving the jet horizontally through the X-ray beam. A schematic of the experimental set-up is shown in Figure 3.1.

Data Collection

Data were collected at the P11 beamline at PETRA III, Deutsches Elektronen Synchrotron (DESY) in May 2016. The PILATUS 6M detector was operated in shutterless mode at 20 Hz. The exposure time per image was set to be 20 ms, the detector distance was kept constant at 200 mm and the beam energy was set to 12,0 keV. This corresponds to an X-ray wavelength of 1.033 Å. The crystal hit rate was monitored online using OnDA [57]. The theoretical L-III absorption edge for mercury is at 1.009 Å [21], however due to the non-standard experimental set-up it was not possible to adjust the X-ray wavelength to the optimal value. Furthermore, the experimental set-up also did not allow for a fluorescence scan to determine the real optimal wavelength for mercury absorption. The impact this may have had on the strength of the anomalous signal during phasing will be discussed in Section 5.2.1.

Data Processing and Refinement

Hit finding, indexing and integration were done using CrystFEL version 0.6.1 [102], and phased using the SHELX package version 2016/1 [95, 82, 83]. Subsequently density modification was performed with DM [20] and the electron density calculated from the

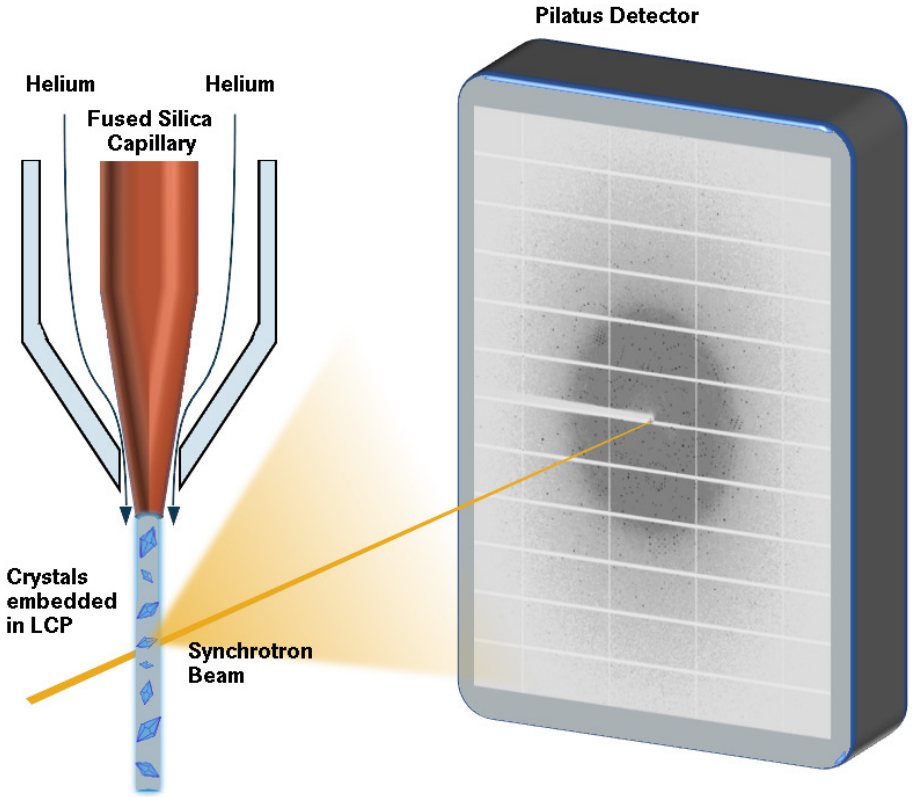


Figure 3.1: Experimental Set-up for Serial Data Collection in LCP
Schematic of the experimental set-up used for serial data collection. The protein crystals are embedded into LCP and streamed through the synchrotron X-ray beam. Helium sheath gas is introduced at the nozzle tip to keep the extruded LCP jet from curling up on itself.

phased data was passed on to automatic model building with ARP/wARP [49]. The resulting structure was refined using alternate cycles of Refmac5 [59] of the CCP4i program suite [103] and manual refining in Coot [26]. To test the robustness of the data the raw integrated intensities of the native and derivative data were also passed into AutoRickshaw [66, 65] for automatic phasing and model building.

For investigating the diffusion of the mercury through the LCP into the crystals, the derivative data were binned into different time intervals, defined by the time lapsed between introducing the crystal into the LCP and the diffraction pattern being collected. Time intervals of 7-30 min, 30-50 min, 50-75 min, 75-100 min, 100-150 min and images taken after 150 min were applied and processed individually. Subsequently the structures

were refined with alternate cycles of Refmac5 [59] and Coot [26] and ultimately the occupancies of the two mercury atoms were refined using Phenix [1]. During this procedure the respective B-factors were set to 20.00 and kept constant during the occupancy refinement.

3.6.3 Inhibitory Molecule Diffusion Study

Soaking Concentration Calculation

To calculate the minimum soaking concentration needed, it is first necessary to calculate the number of protein molecules present in the crystal given its particular size. Given the diamond shape of the crystals (see Figure 4.11), this was approximated as two pyramids with square bases joined at the base, and the volume of which can be calculated using Equation 3.4, where A is the area of the pyramid base (the square of the crystal width) and h the height of one pyramid and therefore half the length of the crystal. The crystal parameters were determined using an optical microscope.

$$V_{crystal} = 2 \cdot \frac{1}{3} Ah \tag{3.4}$$

Proteinase K crystallises in spacegroup $P4_32_12$, from which it can be inferred that there are 8 protein molecules present per unit cell, the volume of which can be calculated with Equation 3.5 from the unit cell constants a, b and c.

$$V_{unitcell} = a \cdot b \cdot c \tag{3.5}$$

The total number of protein molecules present in the whole crystal is therefore the crystal volume $V_{crystal}$ divided by the unit cell volume $V_{unitcell}$ and multiplied by 8, the number of molecules present in each unit cell (Equation 3.6).

$$N_{molecules} = \frac{V_{crystal}}{V_{unitcell}} \cdot 8 \tag{3.6}$$

Finally the number of protein molecules can be converted into a molar protein mass using Avogadro's constant $N_{Avogadro} = 6.022140857 \cdot 10^{23}\ 1/mol$ as shown in Equation 3.7.

$$n_{protein} = \frac{N_{molecules}}{N_{Avogadro}} \tag{3.7}$$

From this molar protein mass, the molar mass of the inhibitor solution was chosen to be $n_{inhibitor} = n_{protein} \cdot 500$ to guarantee that the inhibitor was in excess. The molar mass was calculated for a 500 µm sized crystal (Table 3.5), and this inhibitor concentration was then used for all crystal sizes in the following studies. A summary of the calculated concentrations (using Equations 3.4-3.7) for both boric acid and PMSF is shown in Table 3.6. The inhibitor concentrations were calculated according to the fact that 1 µL of inhibitor solution would be added to each crystal (see next section), and therefore this 1 µL inhibitor solution has to contain a molar mass of $n_{inhibitor}$. For conversion to g/L, the molecular weights were $M_{\text{Boric acid}}$=61.83 g/mol and M_{PMSF}=174.2 g/mol, as provided by the supplier.

Table 3.5: Protein Soaking Concentrations
Crystal volume, unit cell volume, total number of protein molecules in the crystal and total molar mass of a 500 µm proteinase K crystal, calculated with Equations 3.4-3.7.

Protein			
V_{crystal} [µm^3]	$V_{\text{unit cell}}$ [µm^3]	$N_{\text{molecules}}$	n_{protein} [µmol]
$1.04 \cdot 10^7$	$4.91 \cdot 10^{-7}$	$1.69 \cdot 10^{14}$	0.00028
$1.04 \cdot 10^7$	$4.91 \cdot 10^{-7}$	$1.69 \cdot 10^{14}$	0.00028

Table 3.6: Inhibitor Soaking Concentrations
Corresponding inhibitor soaking concentrations to ensure a 500x molar excess of inhibitor molecules in 1 µL solution, referenced to a 500 µm proteinase K crystal.

	Inhibitor		
	$n_{\text{inhibitor}}$ [mmol]	$c_{\text{inhibitor}}$ [mol/L]	$c_{\text{inhibitor}}$ [g/L]
Boric Acid	0.00014	0.14	8.66
PMSF	0.00014	0.14	24.39

Crystal Preparation

Proteinase K crystals were grown in variously sized suspensions as described previously. The longest crystal axis was measured in a standard optical microscope and used to classify the crystal size. The 500 µm, the 200 µm and the 30 µm crystals were prepared in an MRC plate as also described in Section 3.6.1. The 2 µL crystal drop was then supplemented with 1 µL of 8.7 mg/mL of boric acid in mother solution or 25.0 mg/mL PMSF in isopropanol. The timer was started and after the allocated amount of time had passed, the crystals were cryo-protected as outlined before and then flash frozen. The total time from adding the inhibitor solution to flash cooling was noted down and taken to be the total soaking time.

The 5 µm crystals were collected in a semi-serial approach. 1 µL of crystal suspension was pipetted onto a *suna precision* micro-patterned silicon chip and then supplemented with 1 µL 8.7 mg/mL boric acid or 25.0 mg/mL PMSF solution. After the allocated soaking time had passed, the excess solution was blotted off by applying a Kim wipe to the bottom of the chip.

Data Collection

A summary of the data collection strategies for all of the crystals can be found in Table 3.7.

The 200 µm and the 500 µm crystals were collected at the IµS home source from *Incoatec*, using a standard rotational set-up. The 30 µm as well as the 5 µm sized crystals were collected at the P11 beamline at PETRA III, DESY in April 2017 at an X-ray energy of 12.3 keV (corresponding to a wavelength of 1.0089 Å). The 30 µm crystals were also collected in a standard rotational set-up under cryo-conditions. The 5 µm crystals were collected at room temperature using the humidifier set-up [69, 14] provided at P11. 0.5 µL crystal suspension were pipetted onto a micro-patterned silicon chip [69] with 8x8 μm^2 windows and supplemented with 0.5 µL inhibitor solution. After the allotted soaking time excess liquid was blotted off by contacting an absorbent paper wipe to the bottom of the chip. A single crystal was identified through diffraction, and then 5 images were collected while rotating 1° per image.

Table 3.7: Diffusion Data Collection Strategies
X-ray diffraction data collection strategies for crystals with different dimensions, along with the collected crystal soaking times for boric acid and PMSF.

Crystal Size [µm]	Boric Acid Soaking Time [s]	PMSF Soaking Time [s]	Collection Strategy
500	47	50	X-ray Wavelength: 1.5418 Å
	143	131	Rotation Angle: 1°
	326	322	Rotation Range: 60-90°
	625	620	Exposure Time: 30 s
200	52	45	X-ray Wavelength: 1.5418 Å
	177	142	Rotation Angle: 1°
	330	365	Rotation Range: 60-90°
	86400	86400	Exposure Time: 90 s
30	43	38	X-ray Wavelength: 1.0089 Å
	47	72	Rotation Angle: 0.5°
	139	135	Rotation Range: 90°
	142	151	Exposure Time: 100 ms
	314	322	
	349	613	
	613		
	645		
5	10	10	X-ray Wavelength: 1.0089 Å
	30	30	Rotation Angle: 1°
	120	120	Rotation Range: 3°
	300	300	Exposure Time: 60 ms

Data Processing and Refinement

The rotational datasets (30 µm, 200 µm and 500 µm sized crystals) were processed using XDS Version Nov 1, 2016 [39]. The chip data was first manually analysed and only datasets that were inspected to be single crystal hits were processed individually with XDS Version Nov 1, 2016 [39] and then all datasets belonging to the same soaking time were merged using xscale. The structure was solved using pdb entry 1IC6 (after having removed all water het atoms) [96] and subsequently refined using alternate cycles of Refmac5 [59] and Coot [26]. The boric acid or PMS molecules were built into the residual density at Ser224, and the covalent and hydrogen bond lengths were fixed to

comply with the ones identified by LigPlot+ [98]. Once the structure was fully refined and all water molecules added, the individual B-factors of the inhibitor molecule were set to 20.0 and the respective occupancies were refined using Phenix [1].

3.6.4 Caged Mercury De-caging Study

Cage Synthesis

The mercury containing cage was synthesised as described in [106]. A solution of 0.526 mL salicylaldehyde, 0.611 mg, 5.0 mmol were soluted in 20 mL of absolute ethanol and supplemented with 0.594 mL 2-aminothiophenole, 0.687 mg, 5.5 mmol while stirring. This reaction solution was heated to 85°C in an oil bath for 2 hours. The reaction was monitored every hour by thin-layer chromatography using a mixture of of 1:5 ethyl acetate/petroleum ether as the flow control agent. When salicylaldehyde could no longer be detected, the mixture was cooled to 5°C until a colourless, solid substance precipitated. These crystals were filtered with approximately 10 mL of ethanol and a further 10 mL petroleum ether and dried by applying a vacuum. Weighing the dried product gave 467.0 mg, 2.037 mmol. 114.6 mg, 0.5 mmol of the obtained colourless product was suspended in 8 mL methanol and stirred with the addition of 0.5 mmol, 159.4 mg mercury acetate. A yellow, solid substance precipitated. The mixture was boiled in a 85°C oil bath for 15 min. The crystallisation was completed by subsequent cooling to -18°C, and the solid material was filtered and washed with a small amount of methanol. The final product was again dried by applying an ultra high vacuum and weighed to give a final 139.8 mg, 0.314 mmol of the mercury cage. Both reaction steps were monitored using thin layer chromatography.

The first reaction step product was checked with nuclear magnetic resonance (NMR) and the final mercury-containing product was analysed with electrospray ionization mass spectrometry (ESI-MS).

Crystal Preparation and Mercury De-caging

For this study 200 µm proteinase K crystals were used that had been prepared as individual crystals in an MRC plate as outlined in section 3.6.1. The synthesised cage was dissolved

at a concentration of 2.5 mg/mL in a 1:3 (v/v) DMSO:Ethanol mixture. Crystals were placed in 3 µL drops of crystal storing solution and prepared in the following three ways and measured. Firstly, 1 µL of dark, unirradiated peptide was added to a crystal well and left to soak for 90 min before being flash-frozen for subsequent X-ray data collection. Secondly, a well with a crystal was prepared identically to the one before, but irradiated with UV light for the duration of the 90 min soak before being flash-cooled for diffraction data collection. Lastly, 10 µL of the cage solution were irradiated for 90 min before adding 1 µL of this solution to a crystal well and again leaving it to soak (unexposed to UV) for a further 90 min before flash-cooling and measuring the crystal.

Data Collection

Data were collected for the 200 µm crystals under cryo conditions at the IµS home source from *Incoatec* with the strategy outlined in Table 3.8.

Table 3.8: Mercury Cage Soaked Crystal Collection Strategy
Data collection strategy for all the proteinase K crystals after soaking with the caged mercury solution and the various UV irradiation procedures.

X-ray Wavelength	1.5418 Å
Rotation Angle	1°
Rotation Range	60°
Exposure Time	45 s

Data Processing and Refinement

All datasets were processed using XDS Version Nov 1, 2016 [39]. The integrated intensities were then used to solve and refine the protein structure as described previously in Section 3.6.3. Once the structure was fully refined, the expected mercury atoms and all water molecules added, the individual B-factors of the mercury atoms were set to 20.0 and the occupancies refined using Phenix [1].

3.6.5 Substrate Identification Experiments

Flowcell Experimental Set-up, Crystal Preparation and Mounting

A flowcell was constructed and the experimental set-up is shown in Figure 3.2. An X-ray transparent glass capillary (*Mark* capillary from *GLAS, Germany*) was mounted to the goniometer stage of a conventional, rotational crystallography set-up and connecting a *Bio-Rad ECONO* pump to the leading edge of the capillary. For this a commercially available T-connector from *Upchurch* was modified as displayed in Figure 3.2.

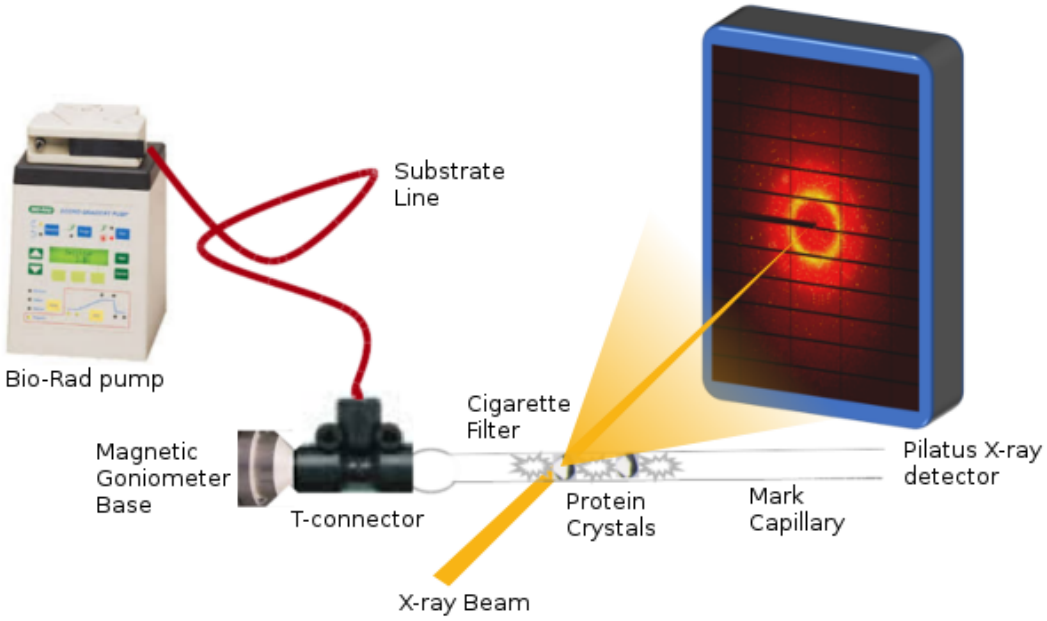


Figure 3.2: Experimental Set-up for the Flowcell Experiments
Schematic of the experimental set-up used for the flowcell experiments. Proteinase K crystals are fixated in a *Mark* capillary between pieces of cigarette filters. Substrate solution is continuously streamed through the crystals in the capillary during diffraction data collection using a *Bio-Rad* pump.

A standard magnetic goniometer base was epoxied into the near side of the T-connector, being careful to seal this exit but leaving the flow path between the remaining two exits clear. The *Bio-Rad* tubing was exchanged with silicon tubing with an inner diameter of 1.3 mm, so that the lowest possible flow rate setting resulted in an actual flow rate of 31 µL/min. The pump configured in this way was connected to the top inlet of the T-connector using 2 m of transparent tubing with an inner diameter of 1 mm, therefore providing an available substrate solution sample volume of 1.57 mL, sufficient for a run-

ning time of 50 min. The inner diameter of the *Mark* capillary was chosen to best fit the crystal size, i.e. for a 500 µm crystal a 0.7 mm inner diameter capillary was chosen. The capillary was filled with proteinase K crystal storing solution and then a small piece of fibre glass cigarette filter was inserted with a syringe needle and positioned slightly downstream of where the crystal was intended to be positioned. A crystal was then fished and deposited into the top of the capillary and left to settle on the cigarette filter through gravitational forces. A further piece of cigarette filter was now inserted and once again using a syringe needle positioned so that it just came into contact with the crystal, without crushing the crystal, to fixate it and restrain it from tumbling during data collection. Now another crystal was inserted then another piece of cigarette filter and so forth until a tandem of 3-4 crystals were fixated inside a single capillary. The capillary prepared in this way was now mounted to a short piece of peek tubing connected to the far side of the T-connector using glass wax from *HARVARD*, again making sure to completely seal the transition but not obstructing the flow path. For data collection the far tip of the capillary was cut off and the BioRad pump was switched on. The peptide solutions listed in Table 3.9 were all collected, with a fresh capillary with fresh crystals being mounted for each new solution. The flow-through from the end of the *Mark* capillary was caught in a plastic container fixed immediately below the capillary outlet.

Table 3.9: Peptide Solutions Screened in the Flowcell Set-up
List of all the peptide solutions streamed through the flowcell set-up for data collection along with the concentrations used and the respective suppliers.

Peptide	Concentration [mg/mL]	Supplier
Ac-Pro-Ala-Pro-Phe-Pro-Ala-NH_2	3	China Peptides
Suc-Ala-Ala-Pro-Phe-pNa	3	Bachem
H-Thr-Pro-Arg-Lys-OH	3	Sigma Aldrich

Flowcell Data Collection

The flowcell was mounted onto the goniometer at the P11 beamline of PETRA III, DESY in November 2016 for room temperature data collection. Standard rotational data were collected on the crystals, using 0.2° rotation per 100 ms exposure at 12.0 keV. Due to the

supply line constricting the goniometer's movement, only 20-40° could be collected per crystal. Therefore multiple datasets were collected for each substrate, facilitated by the tandemly mounted crystals.

Flowcell Data Processing

For the flowcell data, the individual datasets were processed using XDS Version Nov 1, 2016 [39] and all datasets collected for the same substrate were merged using xscale. The integrated intensities were then used to solve and refine the protein structure as described previously in Section 3.6.3. The proteinase K active sites and substrate recognition sites were visibly inspected for residual electron density before adding water molecules to the structure and completing the refinement procedure.

Crystal Preparation for Substrate Screening through Soaking

200 µm crystals were prepared in an MRC pate as outlined in Section 3.6.1. The individual 2 µL drops were then supplemented with 1 µL of one of the substrate solutions listed in Table 3.10. The drops were left to incubate for 5 min, cryo-protected as described in Section 3.6.1, fished and flash frozen in liquid nitrogen.

Table 3.10: Peptide Solutions Screened using a Soaking Approach
List of peptide solutions along with the respective concentrations used and the manufacturers screened for possible substrates using crystal soaking and subsequent crystallographic structure determination.

Peptide	Concentration [mg/mL]	Supplier
Ac-Pro-Ala-Pro-Phe-Pro-Ala-NH_2	15	China Peptides
Suc-Ala-Ala-Pro-Phe-pNa	15	Bachem
Kentsin (H-Thr-Pro-Arg-Lys-OH)	8.5	Sigma Aldrich
Insulin (chain B oxidized)	12	Sigma Aldrich
Myoglobin (horse heart)	62.5	Sigma Aldrich
Caspase Substrate (Ac-Ala-Glu-Val-Asp-pNa)	50	Bachem
Fibrinogen Binding Peptide (H-Glu-His-Ile-Pro-Ala-OH)	3	Bachem
Arg-Gly-Asp-Ser	1	Sigma Aldrich
Neuropeptide Y (H-His-Arg-Leu-Arg-Tyr-NH_2)	1	Bachem

Data Collection

Data were collected conventionally, with 0.5° rotation per 100 ms exposure, with a beam energy of 12.3 keV at the P11 beamline of PETRA III, DESY, in the standard cryo set-up. For data collection crystals were rotated 180° for each substrate.

Data Processing

The individual datasets were processed using XDS Version Nov 1, 2016 [39]. The integrated intensities were then phased and refined as described previously (Section 3.6.3). The proteinase K active site and substrate recognition site were visibly inspected for residual electron density before adding water molecules to the structure and completing the refinement procedure.

3.6.6 Substrate Cross-linking into the Substrate Recognition Site of the Enzyme

Establishing the Cross-linking Procedure

Proteinase K crystals were grown and prepared as outlined in Section 3.6.1. Initial studies were done using 200 µm sized crystals prepared in MRC plates. Glutaraldehyde solutions were prepared in different concentrations by diluting the 50% glutaraldehyde solution in water (as provided) with proteinase K crystal storing solution. Since the crystals dissolved for concentrations above 0.1%, this concentration was not to be exceeded during the following experiments.

Crystal Preparation for Conventional X-ray Data Collection

Bovine insulin, chain B oxidised, is only soluble at low pH, therefore crystal storing solution was prepared with pH 2.0 and the insulin was dissolved in this at a concentration of 3.3 mg/mL. To a 1.5 µL sitting drop of crystal storing solution containing a single proteinase K crystal, 1.5 µL of the insulin solution were added, and after an incubation time of 5 min, a further 1.5 µL of a 0.2% glutaraldehyde solution were added (final drop concentration of 0.07%). The crystal was fished, cryo-protected and measured seven days

later.

Data Collection

For the 200 µm crystal, data were collected under cryo conditions at the IµS home source from *Incoatec* applying the same strategy as outlined for the 200 µm crystals in Table 3.11.

Table 3.11: Cross-linked Insulin Soaked Crystal Collection Strategy
Data collection strategies for all the 200 µproteinase K crystal after soaking with insulin solution and subsequent cross-linking with glutaraldehyde.

X-ray Wavelength	1.5418 Å
Rotation Angle	1°
Rotation Range	60°
Exposure Time	60 s

Data Processing

The 200 µm crystal datasets were processed with XDS Version Nov 1, 2016 [39]. The integrated intensities were then phased and refined as described previously (Section 3.6.3). The proteinase K active site and substrate recognition site were visibly inspected for residual electron density before adding water molecules to the structure.

Crystal Preparation for Serial Data Collection using a Tape-drive Set-up

For the serial data collection crystals of 10 µm size were grown as outlined in Section 3.6.1. The crystals were pooled and washed so that for the tape drive experiment 4 mL of crystal suspension, of which 1/4 was crystal pellet, if left to settle overnight, was prepared. This suspension was filtered through a 20 µm stainless steel frit to remove any larger crystals or crystal clusters. Three different samples were prepared and collected during the beamtime: 1.) 500 µL crystal suspension and 50 µL of 2.5 mg/mL insulin solution were combined and left to incubate for 5 min before the addition of 500 µL 0.2% glutaraldehyde solution (sample **C1**). 2.) 400 µL crystal suspension and 200 µL of

2.5 mg/mL Insulin solution were combined and left to incubate for 5 min before the addition of 400 µL 0.2% glutaraldehyde solution (sample **C2**). 3.) 400 µL crystal suspension were combined with 200 µL Insulin and the sample was **not** subsequently cross-linked with glutaraldehyde (sample **N1**).

Tape-drive Experimental Set-up

The Tapedrive set up, as described in [11], was kindly set-up at the P11 beamline [14] by Dominik Oberthür and his team. A schematic of the working principle is shown in Figure 3.3.

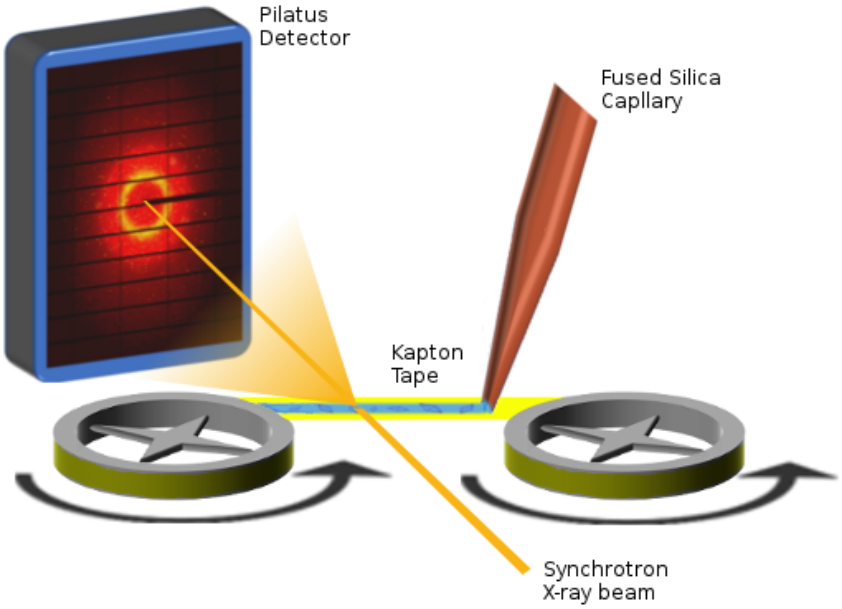


Figure 3.3: Experimental Set-up for Serial Data Collection using the Tape-drive Instrumentation
Experimental set-up for serial data collection of the insulin soaked and subsequently cross-linked proteinase K crystals using the tapedrive instrumentation. Crystal suspension is extruded through the silica capillary onto kapton tape that then traverses the X-ray beam.

A rotating beam chopper was placed upstream of the focusing optics in the X-ray path at P11 to generate X-ray pulses of 4.73 ms duration at a repetition rate of 25 Hz. The Pilatus 6M detector was triggered in sync with the X-ray pulses so that a single diffraction pattern was collected per pulse. As described in [11], this chopper configuration is

required to avoid the sample boiling due to a continuous, intense X-ray exposure as well as to establish control over the exact exposure time. The general concept of the tapedrive is based on a standard, traditional audio tape. The polyimide tape from the right supply roll is wrapped off and onto the left roll. In between the two rolls, a fused silica capillary deposits a thin stream of crystal containing solution onto the detector side of the tape directly upstream of the X-ray interaction region. The crystal suspension was loaded into a sample reservoir that was continuously rocked back and forth by 350° by a motor to keep the crystals suspended and to avoid them settling in the reservoir during data collection. The flow of the crystal suspension was controlled using an *Elveflow* OB1 controller in a hydraulic set-up with water pushing a steel plunger into the reservoir and thus expelling the sample. Pressures of 1900-3000 mbar were used to achieve sample flow rates of 2-5 µL/min. The tape speed was set to 1 mm/s for the duration of this experiment. The hit rate was monitored online during the experiment using OnDA [57].

Tape-drive Data Collection

Proteinase K crystal samples were prepared during the beamtime as outlined in Section 3.6.6. The tape drive data were collected at P11, PETRA III, DESY, in October 2017, using the set-up outlined above, with 12.0 keV X-rays. The two cross-linked samples were both collected immediately after mixing, as well as during the subsequent night, corresponding to a soaking time of 24 hours. In total 325 341 images were initially collected for sample **C1** and a further 137 118 images after 24 h. For sample **C2**, 243 955 images were collected and a further 80 773 after 24 h. A total of 163 140 images were finally collected for sample **N1**.

Tape-drive Data Processing

For the tape drive data, hit finding, indexing and integration were done using CrystFEL version 0.6.1 [102]. The integrated intensities were then phased and refined as described previously (Section 3.6.3). The proteinase K active site and substrate recognition site were visibly inspected for residual electron density before adding water molecules to the structure.

Using the tape drive set-up described in Section 3.6.6, out of the 243 955 images col-

lected for sample **C2**, a total of 46214 patterns were identified as crystal hits by CrystFEL and could be indexed successfully. Since there was no significant difference to be seen between sample **C1** and sample **C2**, only the latter will be presented here due to the higher insulin and glutaraldehyde concentrations used compared to sample **C1**. Out of the 55 773 images collected the following night on the same sample, 2130 patterns could be indexed. Since no significant residual electron density was visible in the substrate recognition site upon first inspection, the data were investigated further by calculating electron density maps at different time points. Runs 1-2, runs 3-4, runs 5-6, runs 7-8, runs 9-10 and runs 11-12 were merged into individual datasets in order to evaluate any time dependant effects. Here, the following time points will be presented and discussed, as they constitute the extremes: 22 min - 52 min (runs 1-2, 8335 indexed images), 3 h - 3 h 34 min, (runs 11-12, 5563 indexed images) and the data collected the following night (21 hours later, 2130 indexed images).

For sample **N1** 7302 patterns out of the 163 140 images collected could be indexed. These patterns were also divided into two datasets, a transient (11 min - 1 h 37 min, 3396 indexed images) and a long-term data set (1 h 37 min - 2 h 57 min, 3906 indexed images).

Chapter 4

Results

4.1 Sample Preparation for Single Particle Imaging

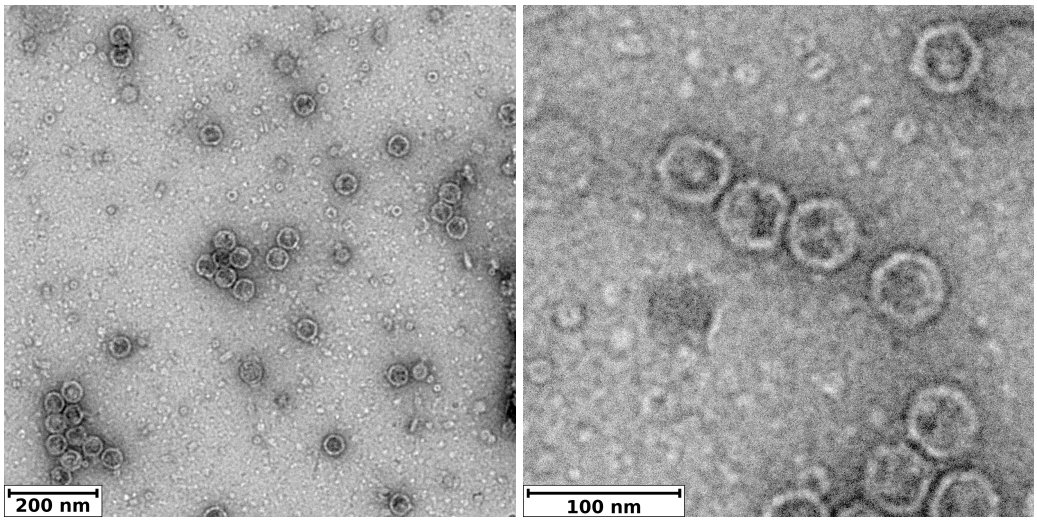
4.1.1 Mycovirus

Sample Characterisation

The Mycovirus samples were provided by Dr. Cornelia Heinze from the University of Hamburg Biozentrum in Klein Flottbeck in a 100 mM sodium phosphate buffer, pH 6.5. The samples were checked by DLS in the buffer provided (Figure 4.3a). Transmission electron microscopy images (Figure 4.1) of the viruses showed a size of approximately 30 nm, which correlates well with the DLS size distribution.

Both the DLS and TEM also confirmed that the viruses display a very homogenous size distribution. Using a Nanosight microscope setup, the particle concentration was calculated to be $17.07 +/- 0.74 \cdot 10^8$ particles/mL, corresponding to the mean integral of the curves shown in Figure 4.2. The particle size determined by Nanosight however had a mean of 169 nm with a standard deviation of 67 nm, therefore resulting in a gross over-estimation of the particle size using Nanosight, when compared with the TEM and DLS results. Furthermore, the Nanosight traces do not show any evidence of a peak around 30 nm. Possible reasons for this will be discussed is Section 5.1.

For single particle imaging experiments at an FEL, it is required to remove as much of the salt in the buffer as possible to eliminate salt crust formation during drying. For application with an aerodynamic lens system a volatile buffer would be beneficial to avoid



(a) 90000x Magnification **(b)** 300000x Magnification

Figure 4.1: Mycovirus TEM Images
TEM images of the Mycovirus at magnification levels 90000x and 300000x.

disruptions from insufficient drying. Therefore, different volatile solutions at various concentrations, including an ammonium acetate buffer, ethanol, isoproponal and pure water were tested by overnight dialysis and subsequent DLS analysis. The mycovirus particles only stayed intact for the case where the particles were dialysed into pure water, the DLS results of which are displayed in Figure 4.3b as well as in 100 mM Ammonium Acetate buffer, pH 6.5 (Figure 4.3c).

The stability of the mycovirus particles when dialysed into water and ammonium acetate was tested by daily DLS measurements over the course of 4 days. The water samples remained stable and mono-disperse with no visible changes in the DLS results after 24 hours (data not shown). The ammonium acetate samples however clearly degraded after 24 h (Figure 4.3d), with a stronger peak occurring at 13.4 nm, half the expected size of the virus particles. The samples prepared by dialysing into de-ionized H_2O were used for later SAXS measurements as well as for additional crystallography trials. Despite the viruses showing a monodisperse size distribution in water (Figure 4.3b), the DLS analysis also shows that they are stable in the volatile buffer ammonium acetate (Figure 4.3c), which is ideal for single particle coherent diffraction experiments. Unfortunately it was not possible to image the virus particles at FLASH during the course of this work due to extensive technical issues during the beamtime.

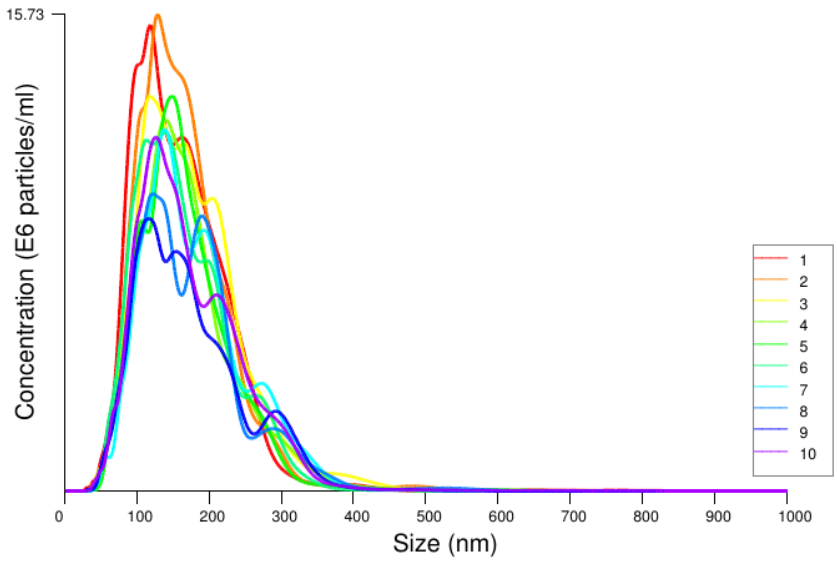


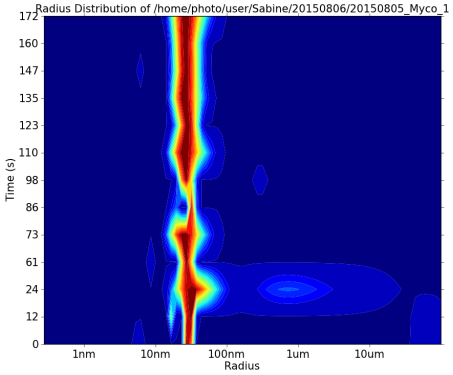
Figure 4.2: Mycovirus Nanosight Results
Concentration versus size curves for all 10 Nanosight measurements of the mycovirus sample.

SAXS

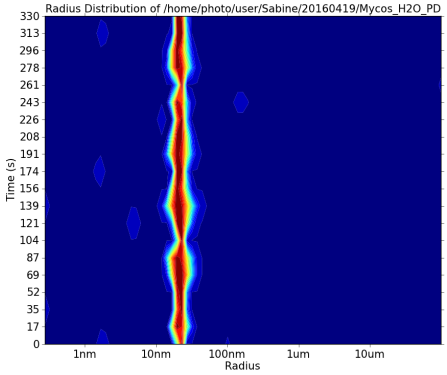
The SAXS data were analysed using *Primusqt* from the *ATSAS* package. An overview of this data analysis is shown in Figure 4.4.

The scattering curves for the Mycovirus sample in de-ionized water were collected for the three protein concentrations of 0.42 mg/mL, 0.21 mg/mL and 0.10 mg/mL. After buffer subtraction the data was scaled and merged, resulting in the curve displayed in Figure 4.4a. From subsequent Guinier analysis (Figure 4.4c), a radius of gyration of 16.49 nm was calculated from the slope of the fit. The distance distribution plot was calculated (Figure 4.4b), and gave a resulting D_{max} of 47.45 nm and an associated porod volume of 24 540 nm^3, corresponding to a molecular weight MW_{Porod} of 14.4 MDa. From the parabolic shape of the curve, it can further be inferred that the particles are mono-dispersively globular in shape, as would be expected for a virus particle.

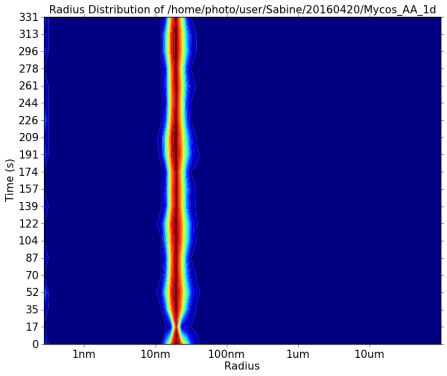
The *ab initio* model calculated from DAMMIN is depicted in Figure 4.5. The dimensions of the particles were measured to be 38.5 nm wide and 31.2 nm long. This again correlates well with the sizes gained from TEM and DLS, and the overall shape of the *ab initio* model bears a high resemblance to what is shown in the TEM images (Figure 4.1).



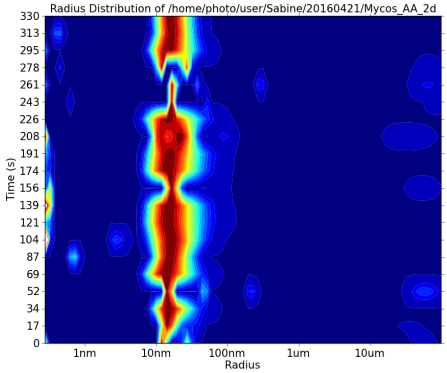
(a) 100 mM sodium phosphate, pH 6.5



(b) De-ionized Water



(c) 100 mM Ammonium Acetate, pH 6.5



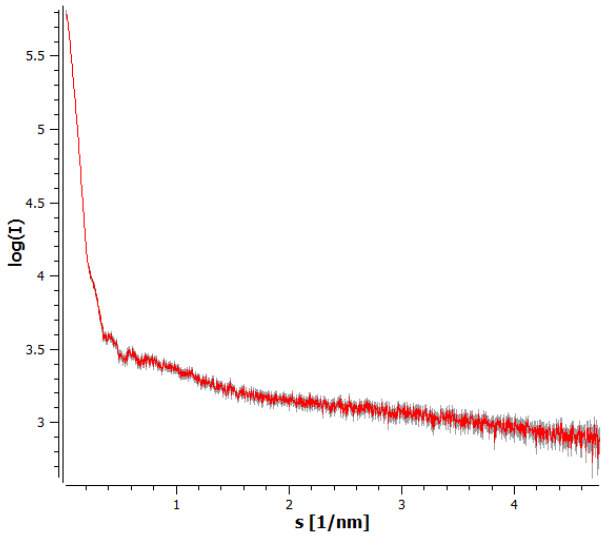
(d) 100 mM Ammonium Acetate, pH 6.5, after 24 h

Figure 4.3: Mycovirus DLS Results
DLS size distribution of the Mycovirus particles in native buffer (a), water (b) an ammonium acetate buffer (c) and 24 h after dialysis into the ammonium acetate buffer (d), along with the respective hydrodynamic radii.

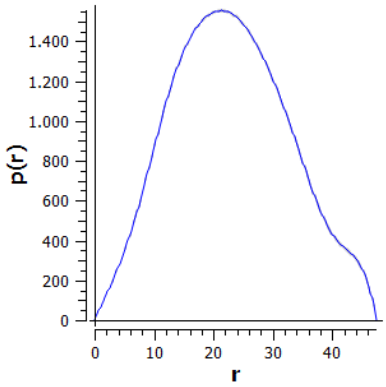
Crystallisation Experiments

Initial crystallisation trials were set up using the virus samples in their supplied 100 mM sodium phosphate buffer, pH 6.5, and concentrations ranging from 4-12 mg/mL as described in Chapter 3.5.4. The most promising condition, B1 of the ammonium sulfate screen (0.2 M cadmium sulfate, 2.2 M ammonium sulfate) yielded spherulites (Figure 4.6a) after 2 days. However all attempts made to sample around this condition did not lead to crystals.

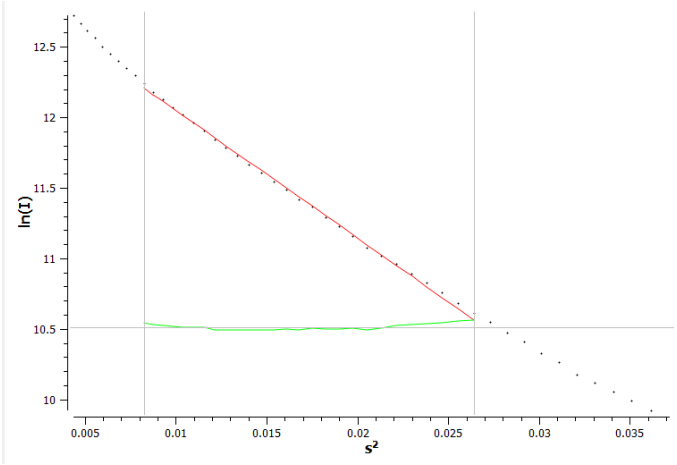
The spherulites were also crushed (Figure 4.6b) and used as a seed stock but to no avail. It was further not possible to verify that the spherulites indeed stem from the my-



(a) Averaged and merged mycovirus scattering data



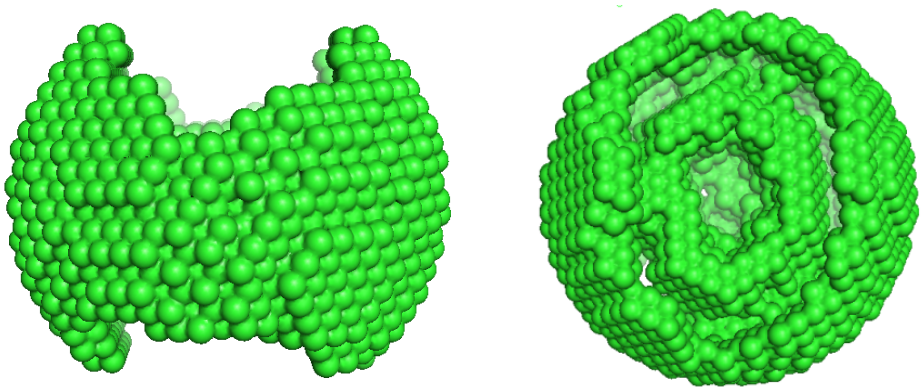
(b) Distance distribution plot



(c) Guinier plot

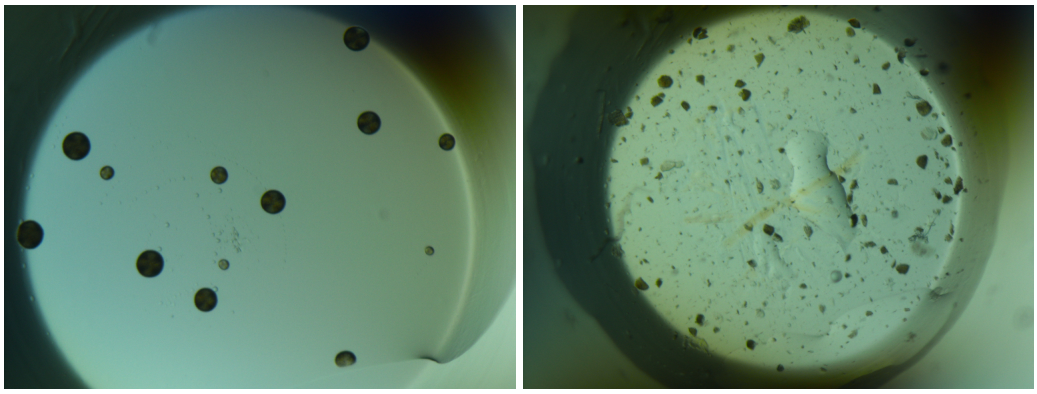
Figure 4.4: Mycovirus SAXS Results Overview
Mycovirus SAXS analysis applying *Primusqt*. (a) shows the averaged and merged scattering intensities versus the scattering angle, (b) the calculated distance distribution plot and (c) the Guinier plot for small scattering angles.

covirus particles and not from a buffer or precipitant component, as intrinsic fluorescence gave no measurable signal, which could however also have been due to the oil covering the plate for this crystallisation procedure. In a further attempt to optimise the mycovirus sample for crystallisation, buffers were screened using the appropriate kits including additives. The optimal buffer condition as determined by DLS was found to be 50 mM bicine buffer, pH 8.5, and 120 mM TCEP and was shown to have a stabilising effect



(a) side view **(b)** top view

Figure 4.5: SAXS *ab initio* Model for the Mycovirus
Ab initio model for the Mycovirus calculated from the SAXS data using DAMMIN viewed (a) from the side and (b) from the top.



(a) Spherulites **(b)** Crushed spherulites for seeding

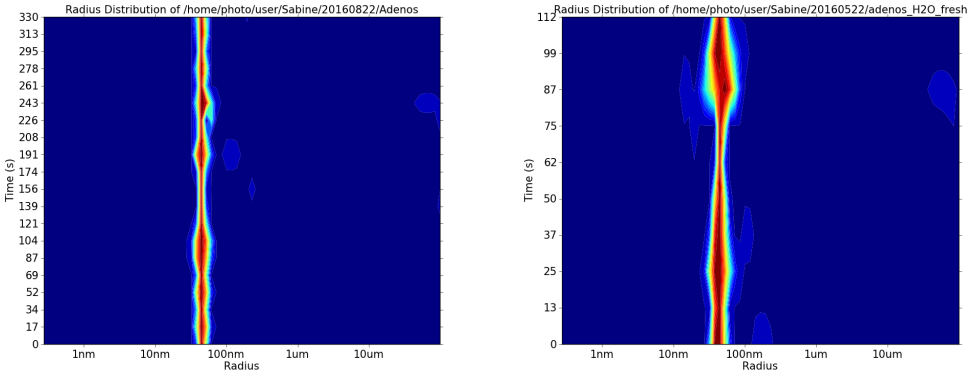
Figure 4.6: Mycovirus Crystallisation Results
Spherulites grown in 0.2 M cadmium sulfate and 2.2 M ammonium sulfate (a) that were crushed (b) to make a seed stock for further crystallisation experiments.

on the virus. Therefore, the virus was dialysed into this 50 mM bicine buffer, pH 8.5, containing 120 mM TCEP and further crystallisation plates were set up. However this did not result in any distinguishable virus crystals. In a final attempt the virus dialysed into de-ionized H_2O was also set up in crystallisation trials but this sample preparation unfortunately did not lead to virus crystals either.

4.1.2 Adenovirus

Sample Characterisation

Adenovirus samples were provided courtesy of Dr. Thomas Speiseder from the Heinrich Pette Insitute. Similarly to the mycovirus sample, the mono-dispersity as well as the particle size distribution was checked by DLS, as provided in a 25 mM Hepes buffer, pH 7.4, containing 137 mM NaCl as well as 3 mM KCl (Figure 4.7a). Similarly, efforts were undertaken to dialyse the viruses into solutions more applicable for single particle imaging experiments. Again, the viruses were shown to be stable in de-ionized H_2O (Figure 4.7b), however no volatile buffer could be identified in which the adenoviruses remained intact. Unfortunately here too, it was not possible to image the virus particles at FLASH during the course of this work due to extensive technical issues during the beamtime.



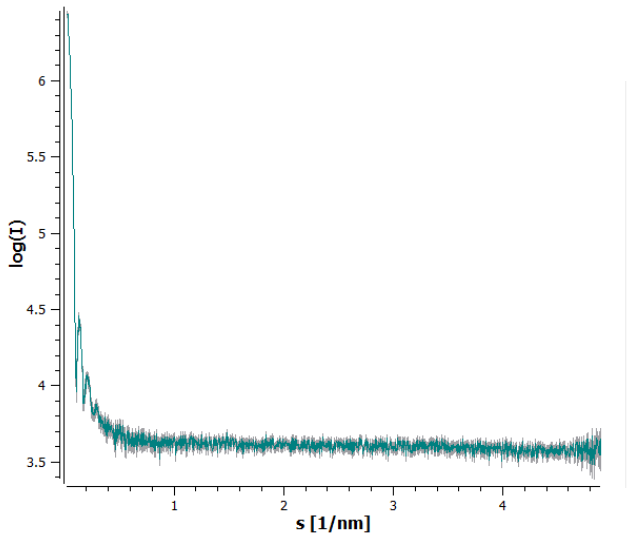
(a) 25 mM Hepes buffer, pH 7.4 **(b)** De-ionized Water

Figure 4.7: Adenovirus DLS Results
DLS size distribution of the adenovirus particles in its native buffer (a) and in water (b), along with the respective hydrodynamic radii.

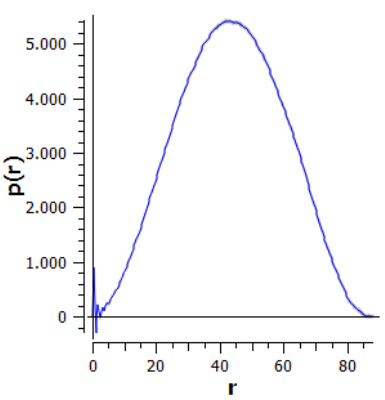
SAXS

The SAXS data collected from the adenovirus particles dialysed into water were analysed as before using *Primusqt* from the *ATSAS* package. An overview of the data analysis is shown in Figure 4.8. The scattering curves for the adenovirus samples in de-ionized water were collected for the three protein concentrations of 0.01 mg/mL, 0.09 mg/mL

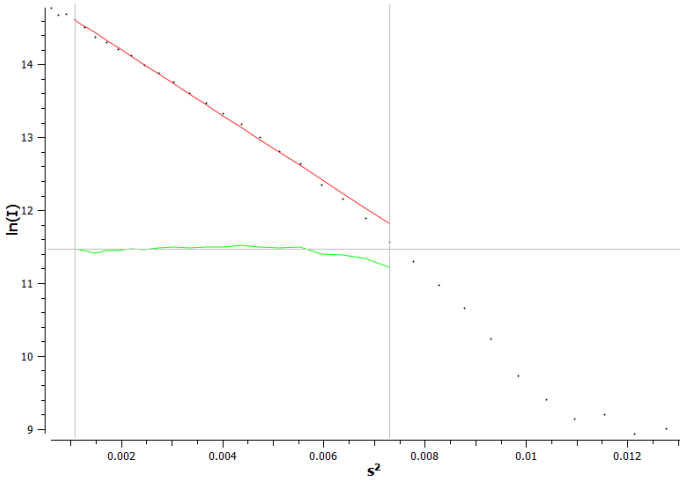
and 0.17 mg/mL. After buffer subtraction the data were scaled and merged, resulting in the curve displayed in Figure 4.8a. From subsequent Guinier analysis (Figure 4.8c) a radius of gyration of 36.5 nm was calculated from the slope of the fit. Kratky plot analysis (Figure 4.8b) gave a resulting D_{max} of 88.1 nm and an associated porod volume of 279 395 nm^3, corresponding to a molecular weight MW_{Porod} of approximately 164 MDa.



(a) Averaged and merged adenovirus scattering data



(b) Distance distribution plot



(c) Guinier plot

Figure 4.8: Adenovirus SAXS Results Overview
Adenovirus SAXS analysis applying *Primusqt*. (a) shows the averaged and merged scattering intensities versus the scattering angle, (b) the calculated distance distribution plot and (c) the Guinier plot for small scattering angles.

The *ab initio* model calculated from DAMMIN is depicted in Figure 4.9. The dimensions of the particles were measured to have a diameter of 88.1 nm, which correlates well with the sizes shown in the DLS analysis. Here too the parabolic shape of the curve, indicates a mono-dispersively globular shape distribution, as would be expected for a virus particle. Furthermore, the *ab initio* model is clearly hexagonal in shape, re-affirming what has been reported previously for this kind of virus [89, 72, 75].

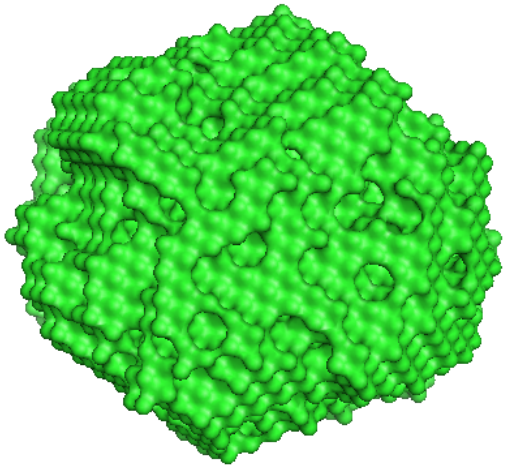
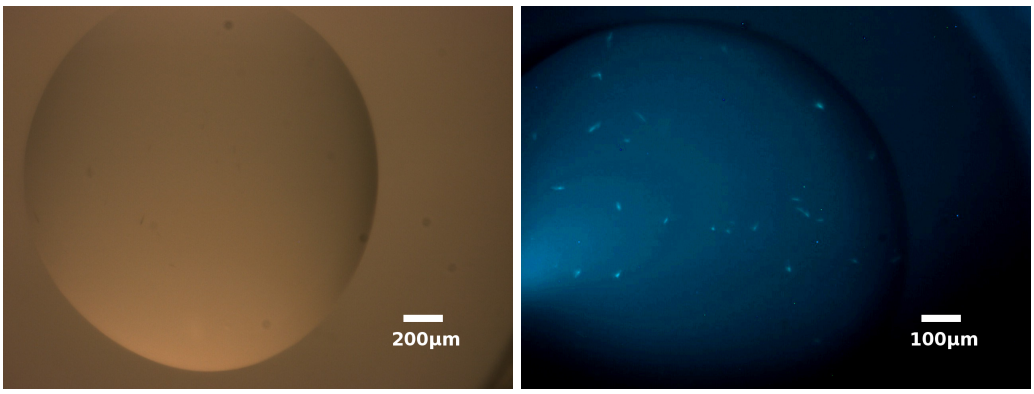


Figure 4.9: SAXS *ab initio* Model for the Adenovirus
Ab initio model for the adenovirus calculated from the SAXS data using DAMMIN.

Crystallisation Experiments

Initial crystallisation trials were set up using the virus samples in their supplied 25 mM Hepes buffer, pH 7.4, containing 137 mM NaCl as well as 3 mM KCl and a concentration of 2-10 mg/mL as described in Chapter 3.5.4. Many of the conditions in the Stura Footprint screen showed showers of tiny needles (Figure 4.10a) that fluoresced when checking for intrinsic fluorescence under a fluorescent microscope (Figure 4.10b) and turned blue when adding blue Izit crystal dye from *Hampton Research* (results not shown).

Both of these results are positive indications of protein being present, and that therefore the microcrystal needles that grew under the conditions in these particular drops do indeed consist of protein and are not salt crystals. The most promising condition, B8, containing 2.0 M potassium/sodium phosphate, pH 7.0, was further optimised by screening around this condition. Both protein concentration, buffer concentration and the pH were



(a) Adenovirus crystals **(b)** Fluorescent microscope image

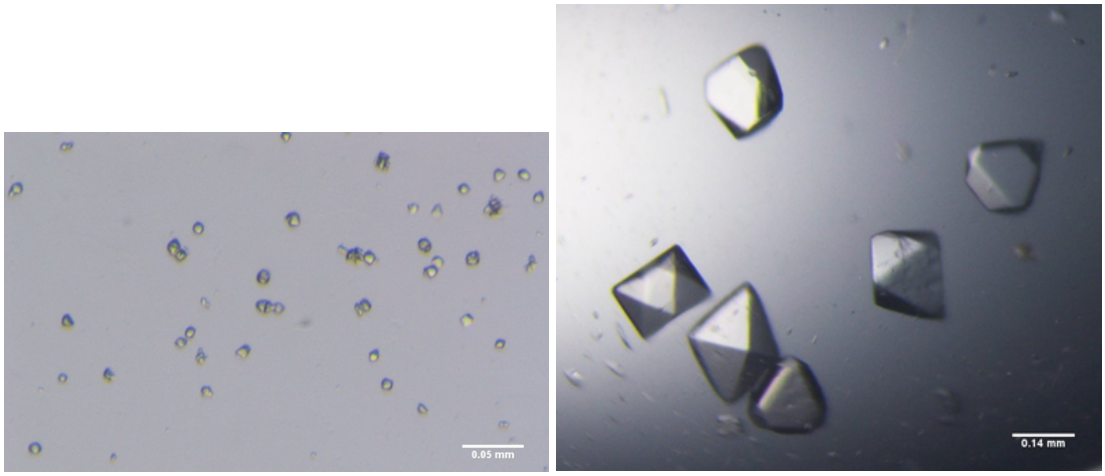
Figure 4.10: Adenovirus Crystal Images
Adenovirus crystals grown in 2.0 M potassium/sodium phosphate, pH 7.0 viewed under the (a) light microscope and (b) fluorescent microscope where clear intrinsic fluorescence is detected.

varied systematically. It was however not possible to grow crystals of sufficient size for testing diffraction. In a final attempt the virus samples dialysed into de-ionized H_2O were applied to crystallisation experiments but this sample preparation unfortunately did not lead to crystals large enough for diffraction studies either.

4.2 Serial Crystallography

4.2.1 Proteinase K Crystallisation

By varying the ratio of protein to buffer to precipitant solution, as described in Section 3.6.1, it was possible to grow homogeneously sized proteinase K crystals in batch, in any desired crystal size. The crystals, when washed multiple times with crystal storing solution, were stable for multiple months without any noticeable deterioration in diffraction quality. The largest crystal grown in this range was 3 mm in length, and the smallest detectable crystals were 1-2 µm. The crystals always displayed the same tetragonal shape, regardless of the crystal size, and light microscope images of two exemplary crystal suspensions are shown in Figure 4.11.



(a) 5 µm sized crystals **(b)** 200 µm sized crystals

Figure 4.11: Microscope Images of Proteinase K Crystal Suspensions
Light microscope images of two proteinase K crystal suspensions, (a) with 5 µm sized crystals and (b) with 200 µm sized crystals.

4.2.2 SIRAS Phase Determination using LCP Soaking and Crystal Delivery

Data were collected serially from a 2-5 µm crystal containing LCP jet streamed through the X-ray interaction region as outlined in Section 3.6.2 at the P11 beamline, PETRA III, DESY.

Mercury Diffusion Analysis

The diffusion of mercury through the LCP and into the protein crystals was analysed by studying the increase in occupancy of the two mercury atoms over time. The the collected data was binned into different time bins, classified by how many minutes post embedding of the crystals into the Hg-containing LCP the diffraction pattern was collected.

The bins were chosen to be 7-30 min, 30-50 min, 50-75 min, 75-100 min, 100-150 min and after 150 min, so that each time bin contained enough images for a complete dataset. Table 4.1 shows the temporal evolution of the occupancies of the two mercury atoms, from which it can be taken, that there is no significant increase in occupancy of the two mercury atoms over the duration of the data collection and that both mutually exclusive mercury sites are nearly fully occupied during the first 30 minutes of data collection.

Table 4.1: Mercury Occupancies at Given Time Points
Mercury occupancies of the two mutually exclusive mercury binding sites after 30, 50, 75, 100 and 200 minutes of introducing native proteinase K crystals into LCP that had been supplemented with $HgCl_2$.

Time [min]	Occupancy Hg 1	Occupancy Hg 2	Combined Occupancy
30	0.41	0.36	0.77
50	0.43	0.36	0.79
75	0.38	0.45	0.83
100	0.39	0.44	0.83
200	0.42	0.43	0.85

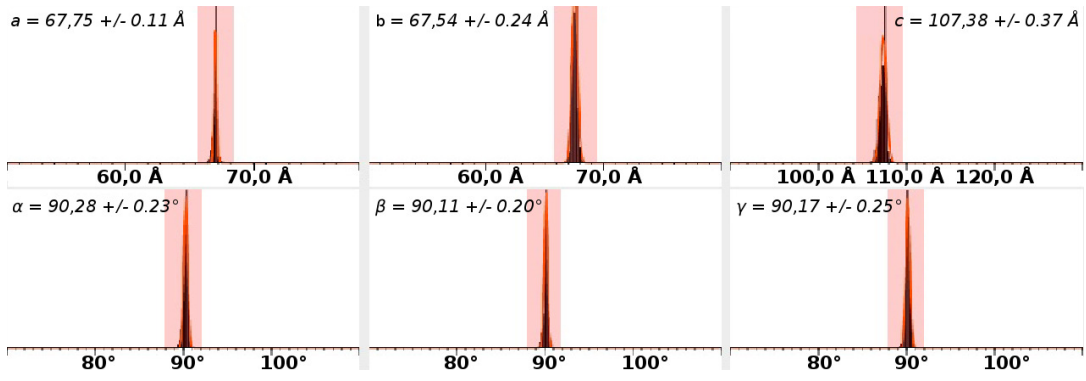
Therefore, the mercury diffuses very efficiently through the LCP and into the crystals, demonstrating that this is a very simple and effective method for soaking heavy atoms into micro-crystals for serial data collection and alleviating the need to handle the often more fragile derivative crystals after soaking. For the occupancy refinement the associated B-factors of the mercury atoms were kept constant at 20.0.

Native and Mercury Derivative Serial Data Analysis

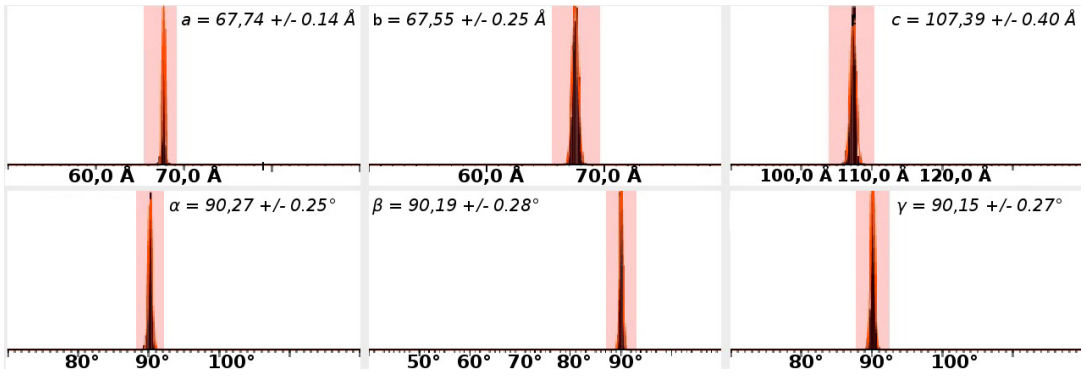
In total 28 674 of the collected 162 702 native Proteinase K patterns were crystal hits and were identified and indexed using CrystFEL [102] as described in Section 3.6.2. Table 4.2 provides the data statistics for this native data that diffracted to a resolution of

1.89 Å. For the mercury derivative, 64 665 patterns of the 397 626 patterns collected in total were identified as crystal hits and could be indexed, similarly to the native data, using space group $P4_32_12$ with cell constants a = b = 67.55 Å, c = 107.40 Å and α=β=γ=90°. The derivative data were processed to a resolution of 1.89 Å and no loss in diffracting power between the two datasets was observed. The latter could merely be processed to a slightly higher resolution, due to the higher redundancy of the data (Table 4.2).

The Gaussian distribution of the cell constants for both data sets of the serially collected images is shown in Figure 4.12, as found and fitted by the *cell_explorer* program from CrystFEL [102].



(a) native data



(b) derivative data

Figure 4.12: Unit Cell Constant Distributions from CrystFEL
Unit cell constants calculated from the unit cell distributions, calculated using the *cell-explorer* routine from CrystFEL for the (a) native and (b) derivative crystal data.

When evaluating the distribution of cell constants from all the crystals used for the mercury derivative, as well as the native data sets in CrystFEL, it can be seen that the mean of the distributions show a maximal deviation of 0.01% for all cell constants be-

Table 4.2: Summary of the Data Statistics for the Native and Derivative Proteinase K Data
Data collection statistics for the native and derivative, serially collected proteinase K datasets that were successfully used to calculate *de novo* phases applying SIRAS, along with the refinement statistics for the final, fully-refined structure.

	Native data	Derivative data
Wavelength (Å)	1.033	1.033
Resolution range (Å)	19.91-1.90 (1.93-1.90)	19.91-1.89 (1.92-1.89)
Space group	$P4_32_12$	$P4_32_12$
Unit-cell parameters		
a / b / c (Å)	67.55 / 67.55 / 107.40	67.55 / 67.55 / 107.40
α / β / γ (°)	90.00 / 90.00 / 90.00	90.00 / 90.00 / 90.00
Total reflections	8029942 (34674)	21365495 (45300)
Unique reflections	37456 (3731)	38258 (3824)
Multiplicity	214.6 (9.4)	558.6 (11.9)
Completeness (%)	99.90 (99.04)	99.97 (99.71)
Mean I/σ(I)	10.06 (2.83)	15.89 (3.10)
Rsplit	8.66 (38.43)	5.68 (35.99)
$CC_{1/2}$	0.998 (0.921)	0.999 (0.921)
Rwork / Rfree	14.23 / 17.14	
No. of atoms		
Total	2199	
Protein	2077	
Het atoms	2	
Water	120	
R.m.s.d., bonds	0.019	
R.m.s.d., angles	1.784	
Coordinate error (Luzzati) (Å)	0.126	
Ramachandran favored (%)	96.90	
Ramachandran outliers (%)	0.39	
Average B factor ($Å^2$)	16.6	
Overall	16.6	
Protein	16.4	
Solvent	26.0	

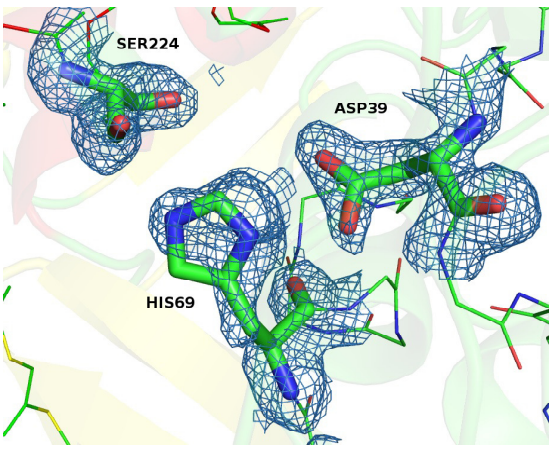
tween the native and derivative crystals, Figure 4.12. This indicates that this method for derivatizing the crystals is exceptionally gentle which is also indicated by the fact that there is no difference in resolution between the data sets (Table 4.2). It should be noted that CrystFEL calculates all the unit cell dimensions individually, and therefore it is up to the user to input the correct cell constants since, in this case as can be seen from Figure 4.12, the value determined from the Gaussian distribution for the cell length a is not exactly the same as for cell length b. The unit cell dimensions were therefore chosen to be a = b = 67.55 Å, c = 107.40 Å , as this gave the best data statistics for all the data, and these cell constants were then also used to calculate the data statistics for the reduced datasets presented below.

SIRAS Phasing and Structure Solution

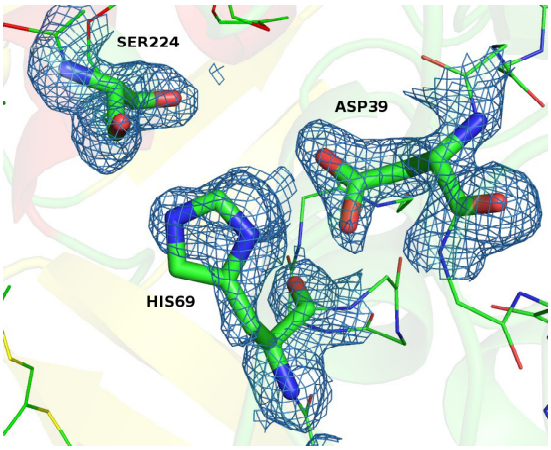
The data was prepared for SIRAS phasing, as outlined in Section 3.6.2, using SHELXC and subsequently two potential heavy atom sites were located with SHELXD. It has been reported previously, that two mercury atoms bind exclusively to the active site of proteinase K [74], which is in good agreement with this observation of the heavy atom sites.

The subsequent 20 cycles of SIRAS phasing and density modification of the serial data with SHELXE, gave the best solution with a CC_{all}=44.9, a CC_{weak}=34.40 and a PATFOM of 64. Finally, the resulting SHELXE phases of the best solution were subjected to another 10 cycles of density modification by DM. The inverse and original hand solutions were inspected visibly in Coot of which the original hand was deemed correct and used for further processing. During these 10 cycles of DM density modification the FOM improved from 0.64 to 0.84 and the electron densities calculated at the different stages of the phasing process are shown in Figure 4.13.

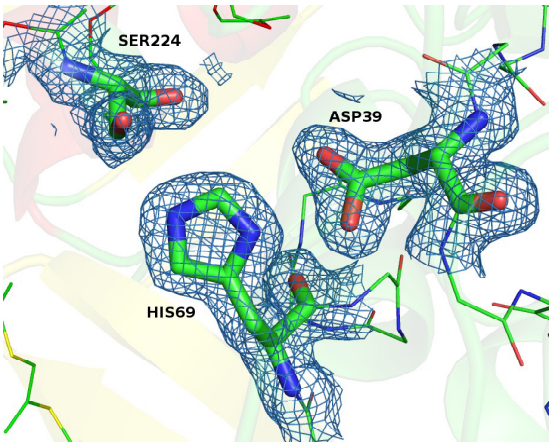
This final density, along with the native proteinase K sequence was submitted to ARP/wARP for 10 rounds of automatic model building. ARP/wARP correctly built 99.28% of the structure during these 10 rounds, leaving only the two terminal alanines to be built manually in Coot. Both termini showed clear electron density in Coot, as shown in Figure 4.14, and the final, complete protein structure was subsequently refined using alternate cycles of automatic and manual refinement with Refmac5 and Coot respectively.



(a) 2Fo-Fc after SHELXE



(b) 2Fo-Fc after DM



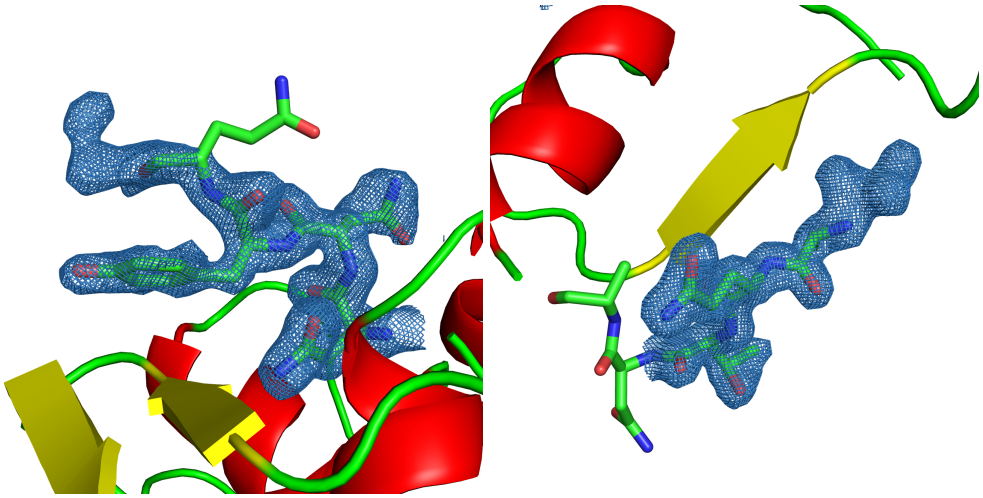
(c) 2Fo-Fc after refinement

Figure 4.13: Electron Density Maps at Different Stages of the Phasing Process
2Fo-Fc electron density maps calculated using the phases after applying (a) SHELXE, (b) DM and (c) the final refined phases all contoured at 1.0σ and overlaid with the final refined structure. The active site is shown and the residues forming the catalytic triad are labeled.

The overall structure was refined to a final R_{work} / R_{free} of 15.69 / 19.52.

An anomalous difference map was calculated using these final phases from the refined model and all of the 64 665 indexed patterns collected from the mercury derivative and is shown in Figure 4.15 contoured at 5σ. From this it is clearly visible that the anomalous density is constrained around the two mercury atoms.

In a further attempt to determine the robustness of this data for phasing, the full native and derivative data from CrystFEL were submitted to the EMBL-HH Automated Crystal



(a) N-terminus density **(b)** C-terminus density

Figure 4.14: Electron Density Maps of the Terminal Residues
2Fo-Fc electron density maps of the (a) N-terminal and (b) C-terminal alanine residues, not identified by ARP/wARP, contoured at 1.0σ.

Structure Determination Platform Auto-Rickshaw along with the protein sequence. This resulted in a structure that was 99% complete with one further residue not being built compared to phasing the data manually, demonstrating the high phasing quality of this data. The structure in this case was refined to a final R_{work} / R_{free} of 15.69 / 19.52.

It was further investigated what the lowest number of images used for calculating the phases and for automatically building the structure could be. When using only 12 000 randomly selected patterns (native/derivative: 5 000/7 000, multiplicity: 41.8/64.9 or 4 000/8 000, multiplicity: 33.3/73.4), the phases obtained from SHELXE were sufficient for ARP/wARP to successfully trace 95% of the protein backbone structure in 10 cycles with the longest gap consisting of 3 amino acids in a loop section. A summary of the data statistics is given in Table 4.3 and Table 4.4. Further reducing the number of images used to 11 000 images in total (native/derivative: 3 000/8 000, 4 000/7 000, 5 000/6 000, 6 000/5 000) despite being sufficient for substructure determination, did not yield electron density maps sufficient for automatic building of the structure with ARP/wARP. The therefore required 12 000 images, when assuming a realistic combined hit and indexing rate of approximately 30%, could be collected in as little as 30 minutes at a synchrotron radiation source beamline such as P11 at PETRA III, limited only by the 25 Hz read-out rate of the Pilatus detector.

Figure 4.15: Proteinase K Anomalous Difference Density Map
Phased refined model of proteinase K with the anomalous difference density map contoured at 5.0σ. The active site residues Asp39, His69 and Ser224 are depicted as sticks. The blue spheres are bound calcium ions. The two mercury atoms (purple) covalently bind to Cys73 close to the protolytic triad.

Table 4.3: Data Collection Statistics for the Native Proteinase K Datasets
Data collection statistics for the native proteinase K crystals using 6000, 5000, 4000 and 3000 indexed diffraction images.

	6000 images	5000 images	4000 images	3000 images
Wavelength (Å)	1.033	1.033	1.033	1.033
Resolution range (Å)	19.91-2.04 (2.11-2.04)	19.91-2.04 (2.11-2.04)	19.91-2.04 (2.11-2.04)	19.91-2.04 (2.11-2.04)
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$
Unit-cell parameters				
a / b / c (Å)	67.55 / 67.55 / 107.40	67.55 / 67.55 / 107.40	67.55 / 67.55 / 107.40	67.55 / 67.55 / 107.40
α / β / γ (°)	90.00 / 90.00 / 90.00	90.00 / 90.00 / 90.00	90.00 / 90.00 / 90.00	90.00 / 90.00 / 90.00
Total reflections	1605620 (25039)	1339372 (21218)	1065881 (18213)	801425 (14459)
Unique reflections	32117 (3211)	32117 (3211)	32117 (3211)	32117 (3211)
Multiplicity	50.0 (7.9)	41.8 (6.7)	33.3 (5.9)	25.2 (4.9)
Completeness (%)	99.93 (99.28)	99.81 (98.16)	99.67 (96.95)	99.20 (92.84)
Mean I/σ(I)	5.35 (2.81)	4.96 (3.23)	4.49 (3.09)	4.03 (3.57)
Rsplit	17.14 (41.84)	18.73 (44.41)	20.93 (46.02)	24.30 (49.58)
$CC_{1/2}$	0.987 (0.903)	0.985 (0.892)	0.982 (0.879)	0.974 (0.868)

Table 4.4: Data Collection Statistics for the Derivative Proteinase K Datasets
Data collection statistics for the derivative proteinase K crystals using 6000, 5000, 4000 and 3000 indexed diffraction images.

	8000 images	7000 images	6000 images	5000 images
Wavelength (Å)	1.033	1.033	1.033	1.033
Resolution range (Å)	19.91-2.04 (2.11-2.04)	19.91-2.04 (2.11-2.04)	19.91-2.04 (2.11-2.04)	19.91-2.04 (2.11-2.04)
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$
Unit-cell parameters				
a / b / c (Å)	67.55 / 67.55 / 107.40	67.55 / 67.55 / 107.40	67.55 / 67.55 / 107.40	67.55 / 67.55 / 107.40
α / β / γ (°)	90.00 / 90.00 / 90.00	90.00 / 90.00 / 90.00	90.00 / 90.00 / 90.00	90.00 / 90.00 / 90.00
Total reflections	2348178 (18845)	2071662 (16608)	1753861 (14174)	1474620 (11560)
Unique reflections	32117 (3211)	32117 (3211)	32117 (3211)	32117 (3211)
Multiplicity	73.4 (6.1)	64.9 (5.5)	55.1 (4.8)	46.6 (4.2)
Completeness (%)	99.60 (95.98)	99.43 (94.27)	99.11 (91.22)	98.54 (85.92)
Mean I/σ(I)	6.82 (5.51)	6.26 (3.27)	6.06 (5.58)	5.53 (4.05)
Rsplit	13.86 (42.18)	14.65 (43.10)	15.73 (45.16)	16.92 (47.85)
$CC_{1/2}$	0.992 (0.907)	0.991 (0.905)	0.989 (0.892)	0.988 (0.885)

4.2.3 Inhibitory Molecule Diffusion Study

Boric Acid Soaking

The results from the binding study of boric acid to proteinase K with LigPlus [98] are shown in Figure 4.16.

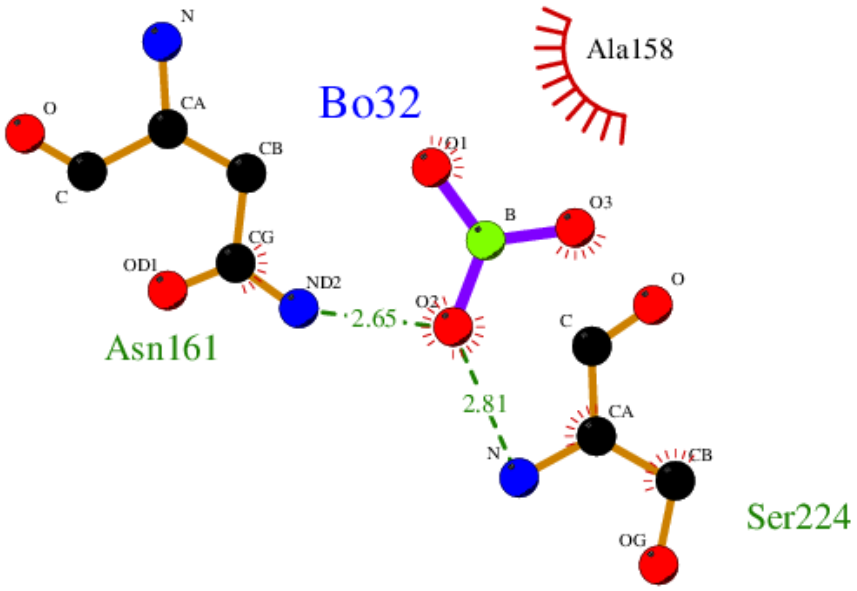


Figure 4.16: Boric Acid and Proteinase K Ligplus Binding Results
Ligplus binding results of proteinase K with boric acid. It is predicted that one of the oxygen atoms of the boric acid forms hydrogen bridges with the ND2 nitrogen of Asn161 and the N nitrogen of Ser224.

From this it can be seen that one oxygen atom of the boric acid molecule forms two hydrogen bonds with the protein, one to the N nitrogen of the active serine 224 residue and one with the ND2 nitrogen of the asparagine 161 residue.

Table 4.5 shows an overview of all the time points collected for the differently sized crystals as outline in Section 3.6.3. For the 5 µm, 120 s crystals unfortunately not enough data was collected during the beamtime for a complete dataset to yield a structure solution and was therefore not processed further. From the occupancy values it can be seen that contrary to the expected increase of the occupancy with time and disproportionate dependance on crystal size, the numbers scatter over a range of 0.6-1.0. There is no visible correlation between occupancy and crystal size or soaking time. Further it should be noted that for the standard rotational cryo-collection (30 µm, 200 µm and 500 µm crystals) soaking times below 38 s could not be achieved due to the experimental handling of the crystals.

Table 4.5: Boric Acid Soaking Study Results Overview
Overview of the boric acid soaking time points collected for the various crystal sizes along with the resulting mean boric acid atom occupancy.

Crystal Size [µm]	Soaking Time [s]	Occupancy
500	47	0.89
	143	0.94
	326	0.77
	625	1.0
200	52	0.87
	177	0.88
	330	1.0
	86400	1.0
30	43	0.75
	47	0.86
	139	0.95
	142	0.57
	314	0.87
	349	0.60
	613	0.72
	645	0.57
5	10	1.0
	30	0.63
	120	-
	300	0.89

PMSF soaking

Phenylmethanesulfonyl fluoride (PMSF) disassociates into phenylmethanesulfonyl (PMS) and a fluoride ion, with the PMS sulphur covalently binding to the OG oxygen of the 224 serine residue. The binding study results of proteinase K and PMS calculated with LigPlus [98] is shown in Figure 4.17. In addition to the covalent bond, a further hydrogen bond is predicted between the nitrogen NE2 of the residue histidine 69 and one of the PMS oxygen atoms.

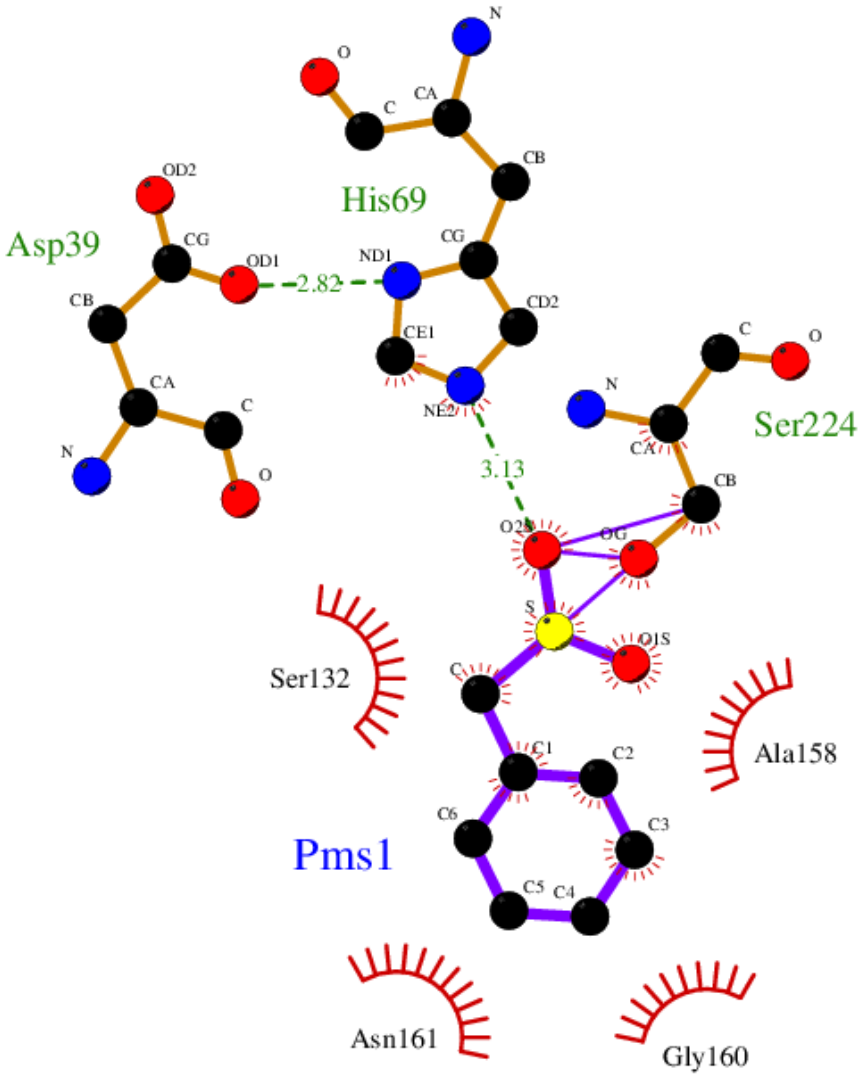


Figure 4.17: PMS and Proteinase K Ligplus Binding Results
Ligplus binding results of proteinase K with PMS. It is known that the OG atom of the active serine residue, in this case Ser224, covalently binds to the PMS sulphur atom.

Table 4.6 shows an overview of all the time points collected for the differently sized crystals as outlined in Section 3.6.3. The 5 µm, 10 s and 120 s datasets were not complete enough to yield a structure solution and were therefore not processed further. From the occupancy values it can be seen that contrary to the expected increase of the occupancy with time and disproportionate dependance on crystal size, the numbers scatter over the a range of 0.1-0.7. Similarly to the boric acid case outlined before, there is no visible correlation between occupancy and crystal size or soaking time. Further it should be noted that for the standard rotational cryo-collection (30 µm, 200 µm and 500 µm crystals) soaking times below 38 s could not be achieved due to the experimental handling time required for the larger crystals.

Table 4.6: PMS Soaking Study Results Overview
Overview of the PMS soaking time points collected for the various crystal sizes along with the resulting mean PMS atom occupancy.

Crystal Size [µm]	Soaking Time [s]	Occupancy
500	50	0.50
	131	0.50
	322	0.56
	620	0.48
200	45	0.44
	142	0.50
	365	0.52
	86400	0.57
30	38	0.13
	72	0.17
	135	0.44
	151	0.37
	322	0.38
	613	0.32
5	10	-
	30	0.69
	120	-
	300	0.69

4.2.4 Caged Mercury De-caging Study

The cage was synthesised with the assistance of Lena Carstensen at the Anorganic Chemistry Institute at the University of Hamburg as outlined in Section 3.6.4.

Cage Synthesis

The salicylaldehyde reacted with the 2-aminothiophenole as shown in Figure 4.18 to compound **1** with a subsequent yield of 41%. Compound **1** was further supplemented with mercury acetate ($Hg(OAc)_2$) to obtain the mercury containing product (Figure 4.19). The yield of this final mercury cage was determined to be 63% and therefore the overall yield of the complete synthesis was 26%.

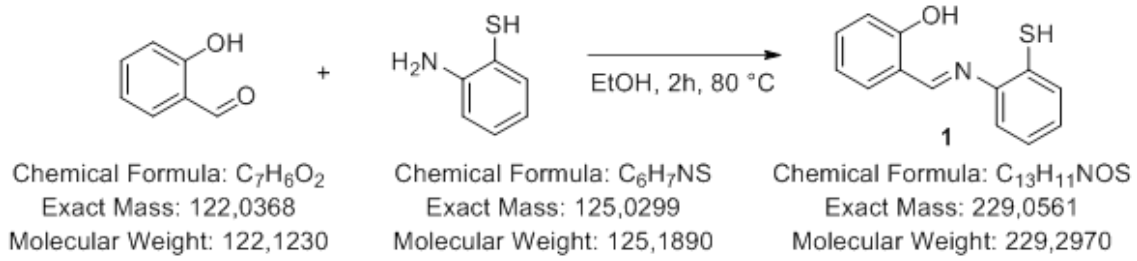


Figure 4.18: Reaction Step 1, Hg-Cage Synthesis
First reaction step of the Hg-cage synthesis. Salicylaldehyde reacts with 2-aminothiophenole to compound **1**.

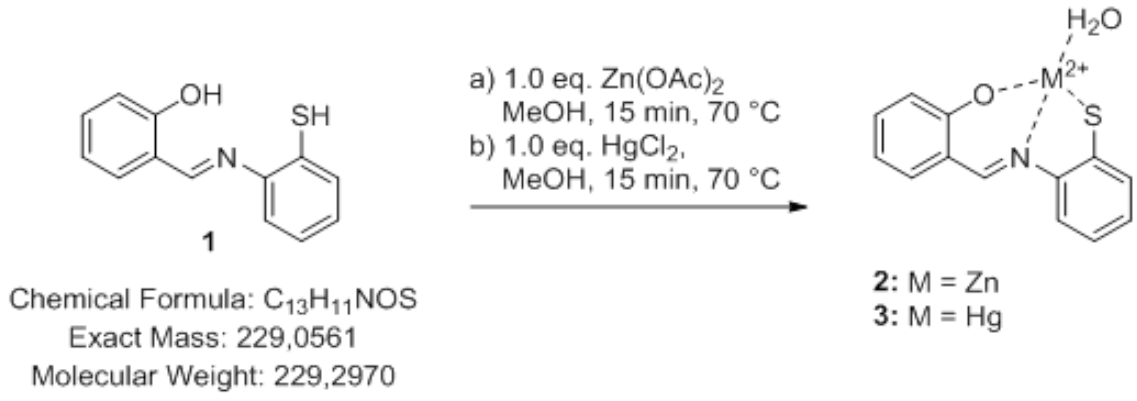


Figure 4.19: Reaction Step 2, Hg-Cage Synthesis
Final reaction step of the Hg-cage synthesis. Compound **1** is supplemented with $Hg(OAc)_2$ to obtain the mercury containing product, compound **2**.

The reaction steps were monitored with thin-layer Chromatography during the synthesis and the TLC results are shown in Figure 4.20. **E1** and **E2** correspond to the two educts, the salicylaldehyde and the 2-aminothiophenole, respectively. **E1** is UV active at 254 nm, and therefore the non-visible trace outline is indicated with gray. The TLC was used to follow the reaction, and therefore each educt was dappled along with the reaction mixture onto the respective **X** positions. This procedure ensures that should two spots ascend to the same height, despite not being the same substance, they can still be differentiated. The final product that was later isolated corresponds to the green or accordingly the blue spots.

(a) Reaction Step 1 **(b)** Reaction Step 2

Figure 4.20: Thin Layer Chromatography Results
Both Reaction steps were monitored with thin layer chromatography at distinct time intervals, where **E** indicates the educt(s), **S** the reaction mixture, and **X** corresponds to cross-dappling the respective educt and the reaction mixture. In reaction step 1 (a), the green spot corresponds to the product, compound **1** and in reaction step 2 (b), the spot dyed blue indicates compound **2**.

The successful synthesis of compound **1** was further verified with NMR, giving the following results:

1H-NMR (300 MHz, 25 °C, CDCl3, TMS): δ [ppm] = 12.51 (s, 1H, H8), 7.99 (d, 1H, H13), 7.90 (d, 1H, H6), 7.70 (dd, 1H, H3), 7.51 (ddd, 1H, H12), 7.40 (m, 2H, H1, H10), 7.11 (d, 1H, H11), 6.96 (t, 1H, H2).

13C-NMR (101 MHz, 25 °C, CDCl3, TMS): δ [ppm] = 132.90 (C3), 128.55 (C1), 126.83 (C13), 125.69 (C11), 122.32 (C12), 121.66 (C2), 119.66 (C10), 118.01 (C6).

These results correlate well with the expected product as described in [106]. The final reaction product, compound **2** was subjected to ESI-MS (HR-ESI+) and it was calculated that for the expected product [$C13H11NO2SHg^{2+}$] m/z should be 447.02. The ESI-MS however gave an m/z of 228.05, probably corresponding to the compound [$C13H10NOS^{-}$]. However, considering that the Hg release is designed to be triggered by photo activation, it is possible that it simply was not detected in the ESI-MS. This was checked by redoing the final reaction step but substituting the $Hg(OAc)_2$ with $Zn(OAc)_2$ as described specifically in [106], and this compound also gave an m/z of 228.05 by ESI-MS.

Mercury De-caging Study

For the de-caging study three test datasets were collected for investigating whether the synthesized cage was reacting as expected. These were as follows: a) 90 min UV-irradiated peptide soaked into a native proteinase K crystal, b) non UV-irradiated peptide soaked into a native proteinase K crystal which was then UV-irradiated for 90 min and the negative control c) neither the peptide nor the soaked crystal were irradiated with UV light. The data and refinement statistics for these datasets are summarised in Table 4.7.

The mercury atom in the two, mutually exclusive, possible binding sites was refined anisotropically, keeping the Bfactor of the respective mercury atoms constant at 20.0. The refined occupancies for all three datasets is summarised in Table 4.8. From this it can be seen that there is no significant difference between the occupancies of the mercury atoms depending on when, or even if, the UV radiation takes place.

Table 4.7: Diffraction Data Statistics for the De-caging Study
Summary of the data collection and refinement statistics for proteinase K crystals after a) soaking with UV-irradiated cage, b) cage soaked crystal with subsequent UV-irradiation and c) cage soaked crystal without any UV exposure.

	UV-irradiated peptide	UV-irradiated, peptide soaked crystal	peptide soaked crystal, no UV
Wavelength (Å)	1.5418	1.5418	1.5418
Resolution range (Å)	20.0-2.29 (2.42-2.29)	20.0-2.15 (2.28-2.15)	25.0-2.19 (2.43-2.29)
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$
Unit-cell parameters			
a / b / c (Å)	67.48 / 67.48 / 107.04	68.13 / 68.13 / 108.14	67.48 / 67.48 / 107.03
α / β / γ (°)	90.00 / 90.00 / 90.00	90.00 / 90.00 / 90.00	90.00 / 90.00 / 90.00
Total reflections	106831 (12585)	160644 (19908)	11649 (1823)
Unique reflections	11576 (1700)	14328 (2111)	11162 (1707)
Multiplicity	9.23 (7.4)	11.21 (9.43)	7.2 (7.3)
Completeness (%)	98.6 (92.4)	98.7 (92.8)	95.8 (93.6)
Mean I/σ(I)	25.6 (11.4)	21.4 (10.1)	22.4 (9.6)
R_{merge}	9.1 (20.3)	11.7 (23.3)	7.7 (19.4)
$CC_{1/2}$	99.8 (97.8)	99.8 (98.5)	99.8 (98.7)
Rwork / Rfree	17.5 / 22.2	14.6 / 17.2	14.2 / 17.1
No. of atoms			
Total	2188	2178	2190
Protein	2051	2079	2031
Het atoms	3	3	4
Water	134	96	155
R.m.s.d., bonds	0.008	0.007	0.008
R.m.s.d., angles	1.077	1.095	1.185
Coordinate error (maximum likelihood) (Å)	0.21	0.15	0.22
Ramachandran			
favored (%)	99.25	99.65	99.06
outliers (%)	0.75	0.35	0.94
Average B factor (Å^2)			
Overall	8.3	11.27	14.75
Protein	8.1	10.97	14.57
Solvent	11.8	17.44	18.81

Only for the first case, where the peptide was UV irradiated before being soaked into the crystal, a slight increase in the occupancy for the first mercury site is noticeable. This could be due to a more effective de-caging, but could also be explained by the fact that given the small irradiated volume of peptide, a comparatively large amount of water evaporated and thereby increased the mercury concentration. Furthermore, the cage is only soluble in DMSO which made the crystals very fragile when soaking. Therefore fishing and mounting the crystals resulted in them breaking and falling apart, making a direct comparison of the crystals difficult, as can be seen from Table 4.7.

Table 4.8: Mercury Occupancy after De-caging
Mercury occupancy of the two mutually exclusive mercury binding sites in the proteinase K crystals after soaking with UV-irradiated cage, cage soaked crystal with subsequent UV-irradiation and cage soaked crystal without any UV exposure.

	UV-irradiated peptide	UV-irradiated, peptide soaked crystal	peptide soaked crystal, no UV
Occupancy HG1	0.24	0.04	0.08
Occupancy HG2	0.06	0.01	0.08

4.2.5 Substrate Identification Experiments

Flowcell Substrate Identification

Since the aim of time-resolved studies is capturing intermediate enzyme states while turning over a substrate, as opposed to being inhibited, the focus of the following work was on capturing such states for which a suitable substrate needed to be identified. Using the flowcell set-up described in Section 3.6.5, three different peptide substrates were investigated, and enough data were collected for complete diffraction datasets. These peptides included the standard proteinase K substrate Suc-Ala-Ala-Pro-Phe-pNa, which is extensively used for monitoring the activity of the enzyme for enzymatic studies. During the experiment it could be seen that the substrate solution was clear when entering the crystal containing *Mark* capillary, but yellow after having traversed the crystal. This is a known indication for this peptide that it has been cleaved. The electron density was calculated for the crystals constantly being replenished with fresh substrates and the substrate recognition site (Gly100-Tyr104 and Ser132-Gly136), as well as the catalytic active site (Asp39, His69 and Ser224) of proteinase K were inspected. However no density that could be related to the substrate located in the regions of interest could be found for any of the three peptides.

Substrate Identification by Soaking

As will be discussed in detail in Section 5.2.4, it was decided that the crystals were possibly too large resulting in only the surface molecules of the crystal turning over the substrate, resulting in a too low occupancy of the substrate in the crystal to yield visible

electron density or that the substrate could possibly be hindered from penetrating into the crystals by the channels being blocked by the outside molecules binding the substrate. It was therefore decided to soak proteinase K crystals with a variety of substrates in an attempt of finding a suitable substrate for further investigations, as described in Section 3.6.5. In particular, some longer peptides were to be employed, possibly giving rise to a ”pass the parcel” principle, where the surface molecules cleave the peptide, but the resulting fragments then diffuse further into the crystal before binding to an enzyme molecule and being cleaved further and so forth.

As described in Section 3.6.5, high-resolution datasets of proteinase K crystals soaked in the peptide solutions listed in Table 3.10 were collected at the P11 beamline of PETRA III, DESY. In particular, not only were short peptides in the single digit amino acid length range used, but crystals were also soaked in solutions with far longer peptide chains, including insulin and myoglobin. Positive electron density was found in the substrate recognition site for both, the crystal soaked with chain B oxidised insulin, as well as for the horse-heart myoglobin soaked crystal. Since insulin, chain B oxidised was chosen as the substrate for all the following experiments, only the results for this case will be shown here. The myoglobin results were similar.

The diffraction data collected from the insulin soaked crystal were indexed and integrated. The structure was then refined to a final R_{work}/R_{free} of 14.91 / 16.68, and a summary of the data statistics can be found in Table 4.9.

Upon inspection of the substrate recognition site in particular, some difference density can clearly be identified that does not correspond to water molecules (Figure 4.21), especially when considering the 1.15 Å resolution of the maps.

In order to ascertain that proteinase K was indeed turning over the insulin chain, biochemical studies were performed. Pure insulin solution (3.3 mg/mL), pure proteinase K solution (0.9 mg/mL) as well as a mixture of the two (3:1 (v/v)) that had been incubated for two hours were applied to SDS-Page, the results of which are shown in Figure 4.22. From this it can be seen that despite the insulin band clearly being visible when applied individually, it completely disappears when having been incubated with proteinase K for two hours. The proteinase K band is however visible when applied individually, as well as in combination with the insulin.

Table 4.9: Data Collection and Refinement Statistics for the Insulin Soaked Proteinase K Crystal
Summary of the data collection and refinement statistics of the flash-frozen proteinase K crystal soaked with insulin solution.

	Insulin soaked dataset
Wavelength (Å)	1.0089
Resolution range (Å)	50 - 1.15 (1.22-1.15)
Space group	$P4_32_12$
Unit-cell parameters	
a / b / c (Å)	67.81 / 67.81 / 107.23
α / β / γ (°)	90.00 / 90.00 / 90.00
Total reflections	547255 (78686)
Unique reflections	89274 (14218)
Multiplicity	6.13 (5.53)
Completeness (%)	99.5 (98.8)
Mean I/σ(I)	10.37 (1.91)
R_{merge}	10.2 (68.9)
$CC_{1/2}$	99.7 (99.5)
Rwork / Rfree	11.2 / 14.0
No. of atoms	
Total	2500
Protein	2149
Het atoms	1
Water	350
R.m.s.d., bonds	0.030
R.m.s.d., angles	2.462
Coordinate error (maximum likelihood) (Å)	0.018
Ramachandran favored (%)	99.6
Ramachandran outliers (%)	0.4
Average B factor (Å^2)	
Overall	10.3
Protein	9.0
Solvent	25.2

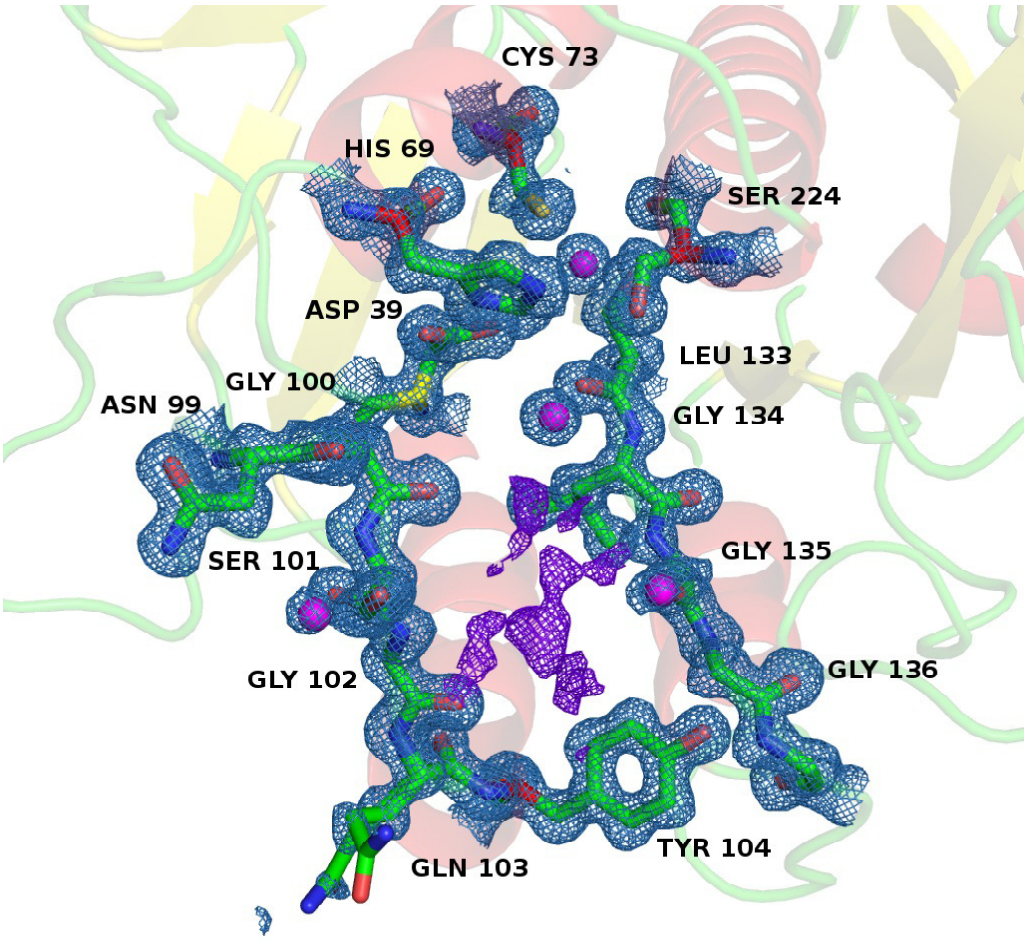


Figure 4.21: Electron Density Map of the Substrate Recognition and Active Site of Proteinase K after Soaking the Crystal with Insulin
2Fo-Fc map (blue) and nF1-mF2 difference density map (purple) of the substrate recognition (Asn99-Tyr104 and Leu133-Gly136) and active site (Asp39, His69 and Ser224), depicted as sticks, of the proteinase K crystal soaked with insulin. The electron density maps are contoured at 1.0σand the magenta spheres are water molecules.

In a further classification experiment, DLS was employed to investigate whether a proteinase K crystal remained intact while turning over insulin. For this the lowest insulin concentration that still yielded reasonable DLS signal (determined to be 3.3 mg/mL) 15 µL were filled into a quartz cuvette and subjected to a fast, characterising DLS measurement. Subsequently 5 µL of proteinase K solution at a concentration that no longer gave rise to a detectable DLS signal (determined to be 0.9 mg/mL) was added to the insulin solution in the cuvette and a long term DLS measurement was taken. The results of this study are depicted in Figure 4.23.

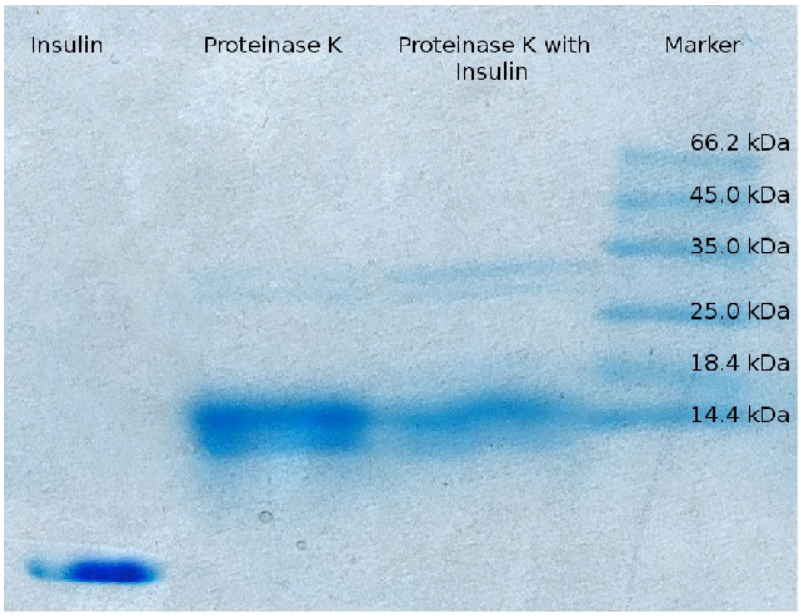
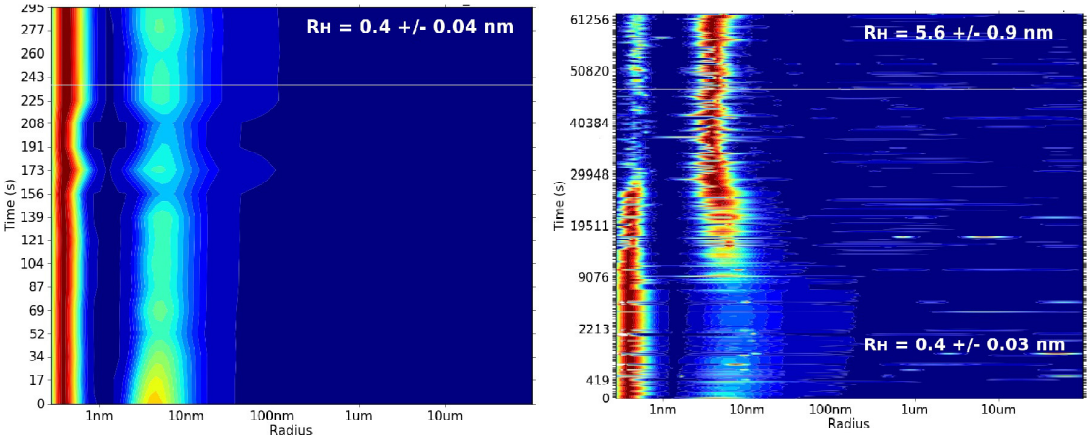


Figure 4.22: SDS-page Results of Insulin, Proteinase K and a Proteinase K and Insulin Cocktail
SDS-page of pure insulin solution, proteinase K solution and a mixture of proteinase K and insulin solution that was incubated at room temperature for 2 h. The protein marker was applied to the far right well, and the bands are labeled with the respective molecular weights, as supplied by the provider.



(a) Insulin

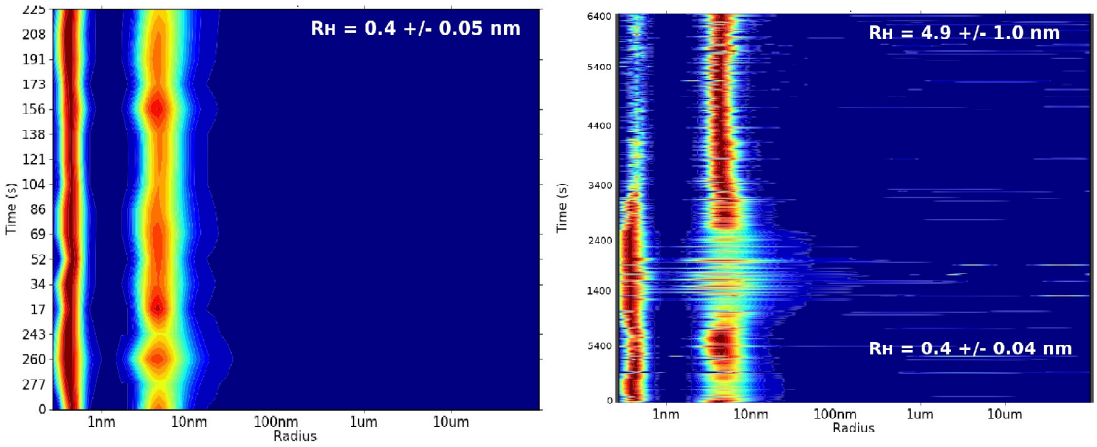
(b) Insulin with Proteinase K

Figure 4.23: Long term DLS of Insulin and Proteinase K Solution
Temporal evolution of the DLS size distribution for the insulin solution as well as for the insulin solution after a few micro-liters of proteinase K solution have been added, along with the respective hydrodynamic radii.

At the beginning of the measurement, the DLS signal from the insulin can still clearly be detected. Over time however it disappears, and a signal at a much larger size dominates. This signal could be solely noise (given that the minimal insulin concentration that still

gave DLS signal was used), or aggregates of the insulin fragments clustering together.

Concurrently, the same measurement procedure was applied, but instead of adding a few micro-liters of proteinase K in solution, a 500 µm proteinase K crystal was fished and added to the insulin solution in the cuvette (Figure 4.24).



(a) Insulin

(b) Insulin with Proteinase K

Figure 4.24: Long term DLS of Insulin Solution with a Proteinase K Crystal
Temporal evolution of the DLS size distribution for the insulin solution as well as for the insulin solution after a proteinase K crystal has been added, along with the respective hydrodynamic radii.

For the pure insulin solution, the insulin signal can also clearly be seen, as well as a slight peak at the ”aggregate formation peak”. This is probably due to the fact that for cost reasons, the same insulin sample was used as for the previous measurement, but a day later and so the sample had slightly degraded overnight. What can be seen during the long term DLS however is a very similar evolution of the insulin aggregate formation as in the above case where proteinase K solution, as opposed to a proteinase K crystal, was added. The slight spike in aggregate formation at the start of the measurement is due to the crystal coming from a pH 6.5 solution, a pH at which insulin is not soluble.

4.2.6 Substrate Cross-linking into the Substrate Recognition Site of the Enzyme

Conventional X-ray Data Collection

For determining conditions under which the crystals remained stable during cross-linking, different cross-linking concentrations were tested. It was found that a 200 µm crystal, prepared as described in Section 3.6.6, after 7 days of soaking in a 1 mg/mL insulin and subsequently 0.1%(v/v) glutaraldehyde (Figure 4.25) solution remained stable. Even when transferring the crystals into pure water prior to cryo-protecting and testing diffraction, diffraction did not degrade, indicating that cross-linking had been successful. A further crystal prepared in this way was subjected to conventional diffraction data collection at the IµS X-ray source and the structure was solved and refined. A summary of the data and refinement statistics is shown in Table 4.10. Figure 4.26 shows the electron density in the active and substrate recognition site of proteinase K.

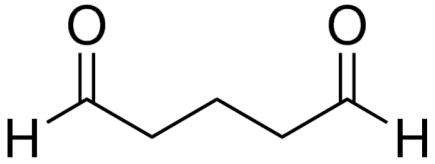
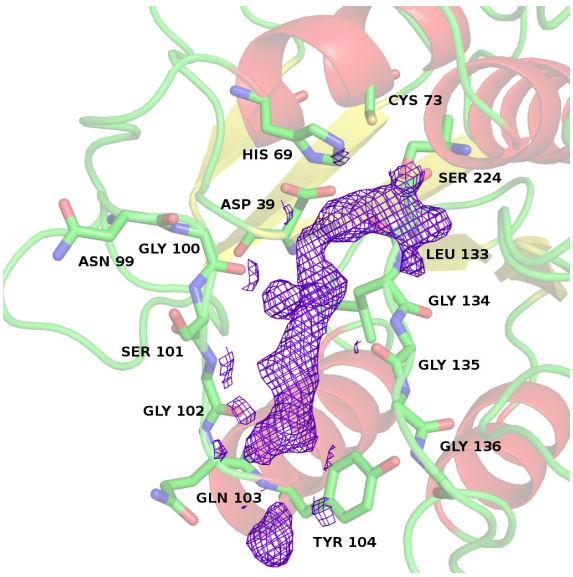


Figure 4.25: Chemical Formula for Glutaraldehyde
Chemical formula of glutaraldehyde as specified by the provider [38].

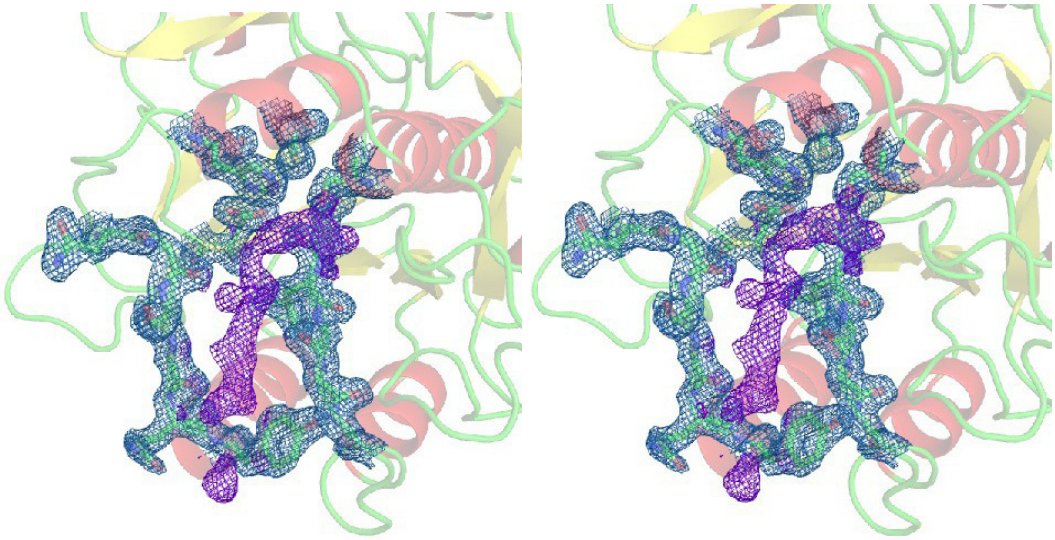
From this clear, continuous difference density can be seen in the substrate recognition site, that is too long to correspond to a glutaraldehyde molecule (Figure 4.25). Therefore it is clear that this density can only correspond to the backbone of an insulin fragment that has successfully been cross-linked into the active site of the enzyme.

Table 4.10: Data Collection and Refinement Statistics for the Insulin Soaked Proteinase K Crystal
Summary of the data collection and refinement statistics of the flash-frozen proteinase K crystal soaked with insulin solution.

	Insulin soaked, cross-linked Proteinase K crystal
Wavelength (Å)	1.5418
Resolution range (Å)	23.1-2.15 (2.28-2.15)
Space group	$P4_32_12$
Unit-cell parameters	
a / b / c (Å)	68.27 / 68.27 / 107.00
α / β / γ (°)	90.00 / 90.00 / 90.00
Total reflections	62451 (9586)
Unique reflections	13490 (2086)
Multiplicity	4.6 (4.6)
Completeness (%)	99.2 (98.6)
Mean I/σ(I)	11.29 (3.67)
Wilson B factor (Å^2)	20.47
R_{merge}	15.2 (48.2)
$CC_{1/2}$	99.3 (90.0)
Rwork / Rfree	21.5/27.8
No. of atoms	
Total	2200
Protein	2078
Het atoms	1
Water	121
R.m.s.d., bonds	0.016
R.m.s.d., angles	1.66
Coordinate error (Luzzati) (Å)	0.16
Ramachandran favored (%)	99.61
Ramachandran outliers (%)	0.39
Average B factor (Å^2)	16.7
Overall	16.83
Protein	13.02
Solvent	21.33



(a) Close-up view of the active and substrate recognition sites of the proteinase K model with the associated nF1-mF2 electron density map



(b) Stereo image of the active and substrate recognition sites of the proteinase K model with the associated electron densities (to be viewed with stereo glasses)

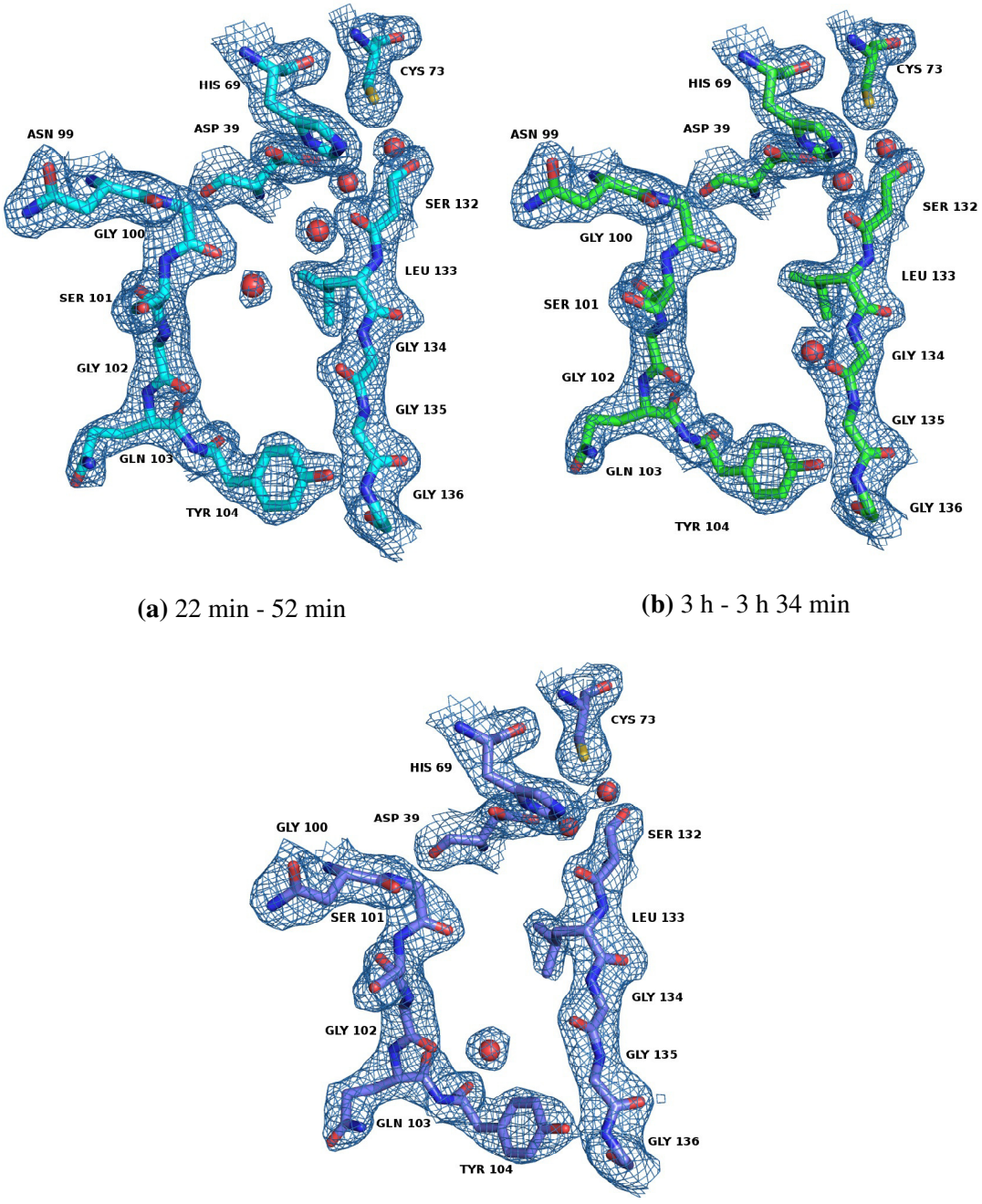
Figure 4.26: Electron Density Map of the Substrate Recognition and Active Site of Proteinase K after Soaking the Crystal with Insulin and Subsequent Cross-linking
2Fo-Fc map (blue) and nF1-mF2 difference density map (purple) of the substrate recognition site (Asn99-Tyr104 and Leu133-Gly136) and active site (Asp39, His69 and Ser224), depicted as sticks, of the proteinase K crystal soaked with insulin and subsequently cross-linked. The electron densities in the standard image (a) as well as in the stereo image (b) are contoured at 1.0σ.

Serial Data Collection using a Tape-Drive Set-up

This newly established method of keeping substrate fragments in the substrate recognition site was now applied in a serial data collection approach. Given the far smaller crystal size that can be applied with this technique and the associated shorter diffusion distances into the crystal, the aim was to investigate whether the substrate fragments fixed to the substrate recognition site are longer compared to the conventional approach. This would be indicative of the ”pass the parcel” theory. As described in Section 3.6.6, a tape-drive data collection set-up at P11, PETRA III, was employed. the time dependant response of sample **C2**, consisting of proteinase k crystals soaked with insulin and subsequent cross-linking with glutaraldehyde as well as sample **NI**, proteinase K crystals soaked with an excess of insulin only, without cross-linking, will be presented. For the former case a summary of the data statistics of the data collected during the time points 22 min - 52 min, 3 h - 3 h 34 min and 21 hours after the addition of the glutaraldehyde is summarised in Table 4.11. Figure 4.27 shows the substrate recognition site for sample **C2** at the time points a) 22 min - 52 min, b) 3 h - 3 h 34 min and c) 21 hours, along with the active site catalytic triad for these time points. From Figure 4.27 it can be seen that the two catalytic water molecules bound in the active site remain bound during the full 24 h study. The remaining two water molecules in the substrate recognition site however are only visible at the start of the study. This is indicative of the fact that these water molecules are displaced due to substrate binding of the insulin. Furthermore, the electron density around these two water molecules in the long term, 21 h, density is reduced compared to the former two datasets. A single further water molecule can be seen after 3 h and 21 h in the substrate recognition site, however these two water molecules are not in the same location. The statistics for the data collected on the proteinase K crystals soaked with an excess of insulin, without subsequent cross-linking, are summarised in Table 4.12. This data was divided into two datasets, 1 min - 1 h 37 min and 1 h 37 min - 2 h 57 min after mixing the crystals with the insulin. The associated electron density maps of these two datasets are shown in Figure 4.28.

Table 4.11: Data Collection and Refinement Statistics for the Serially Collected, Insulin Soaked and Cross-linked, Proteinase K Crystal Data
Summary of the data collection and refinement statistics of the serially collected proteinase K crystal data. The crystals were soaked with insulin solution and subsequently cross-linked. Data points of 22 min - 52 min, 3 h - 3 h 34 min and 20 h 30 min - 21 h 29 min after the cross-linking procedure are presented here.

	22 min - 52 min	3 h - 3 h 34 min	20 h 30 min - 21 h 29 min
Wavelength (Å)	1.033	1.033	1.033
Resolution range (Å)	29.1-2.14 (2.22-2.14)	29.1-2.14 (2.22-2.14)	29.1-2.44 (2.53-2.44)
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$
Unit-cell parameters			
a / b / c (Å)	68.15 / 68.15 / 107.81	68.15 / 68.15 / 107.81	68.15 / 68.15 / 107.81
α / β / γ (°)	90.00 / 90.00 / 90.00	90.00 / 90.00 / 90.00	90.00 / 90.00 / 90.00
Total reflections	4175473 (38547)	2208758 (31013)	610287 (17311)
Unique reflections	28365 (2819)	28365 (2825)	19009 (1915)
Multiplicity	147.2 (13.7)	77.9 (11.0)	32.1 (9.1)
Completeness (%)	99.98 (99.79)	99.98 (99.86)	99.95 (99.63)
Mean I/σ(I)	3.57 (0.88)	2.76 (0.87)	1.91 (0.68)
Rsplit	23.36 (114.82)	29.06 (112.21)	45.67 (145.23)
$CC_{1/2}$	0.989 (0.659)	0.98 (0.64)	0.95 (0.62)
Rwork / Rfree	15.92 / 20.38	16.80 / 20.66	19.78 / 27.56
No. of atoms			
Total	2157	2104	2082
Protein	2077	2031	2054
Het atoms	1	1	1
Water	79	72	27
R.m.s.d., bonds	0.015	0.016	0.010
R.m.s.d., angles	1.61	1.67	1.41
Coordinate error (maximum likelihood) (Å)	0.15	0.16	0.32
Ramachandran			
favored (%)	99.28	99.28	99.28
outliers (%)	0.72	0.72	0.72
Average B factor ($Å^2$)			
Overall	29.9	30.38	34.5
Protein	29.8	25.9	34.6
Solvent	33.1	32.8	26.3



(a) 22 min - 52 min

(b) 3 h - 3 h 34 min

(c) 20 h 30 min-21 h 29 min

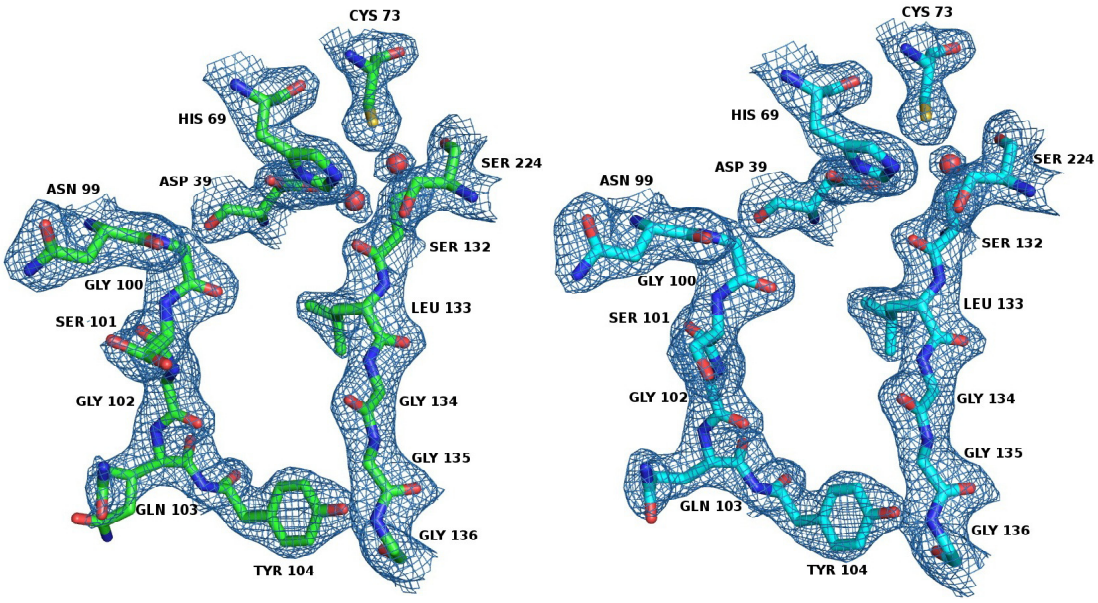
Figure 4.27: Electron Density Maps of the Proteinase K Substrate Recognition and Active Sites for the Serially Collected Insulin Crystal Data with Cross-Linking

2Fo-Fc electron density maps of the substrate recognition site (Asn99-Tyr104 and Leu133-Gly136) and active site (Asp39, His69 and Ser224) of proteinase K for the time intervals (a) 22 min - 52 min, (b) 3 h - 3 h 34 min and (c) 20 h 30 min - 21 h 29 min after the soaking and cross-linking procedure. The red spheres represent water molecules. All maps are contoured at 1.0σ.

Table 4.12: Data Collection and Refinement Statistics for the Serially Collected, Insulin Soaked Proteinase K Crystal Data
Summary of the data collection and refinement statistics of the serially collected proteinase K crystal data. The crystals were soaked with insulin solution. Data points of 1 min - 1 h 37 min and 1 h 37 min - 2 h 57 min after the soaking procedure are presented here.

	1 min - 1 h 37 min	1 h 37 min - 2 h 57 min
Wavelength (Å)	1.033	1.033
Resolution range (Å)	29.1-2.24 (2.32-2.24)	29.1-2.14 (2.32-2.24)
Space group	$P4_32_12$	$P4_32_12$
Unit-cell parameters		
a / b / c (Å)	68.15 / 68.15 / 107.81	68.15 / 68.15 / 107.81
α / β / γ (°)	90.00 / 90.00 / 90.00	90.00 / 90.00 / 90.00
Total reflections	1360464 (23798)	1266079 (19720)
Unique reflections	24640 (2443)	24640 (2443)
Multiplicity	55.2 (9.8)	51.4 (8.2)
Completeness (%)	99.98 (99.88)	99.89 (98.98)
Mean I/σ(I)	2.05 (0.58)	2.00 (0.63)
Rsplit	39.76 (167.37)	42.81 (150.46)
$CC_{1/2}$	0.970 (0.525)	0.962 (0.584)
Rwork / Rfree	19.64 / 26.51	21.00 / 25.23
No. of atoms		
Total	2124	2079
Protein	2096	2040
Het atoms	1	1
Water	27	38
R.m.s.d., bonds	0.012	0.011
R.m.s.d., angles	1.59	1.43
Coordinate error (maximum likelihood) (Å)	0.251	0.282
Ramachandran favored (%)	99.60	99.64
Ramachandran outliers (%)	0.40	0.36
Average B factor (Å^2)		
Overall	35.7	33.3
Protein	35.8	33.8
Solvent	31.1	30.5

Similarly as in Figure 4.27, it can be seen in Figure 4.28 that only two water molecules in the catalytic triad are bound, however with significantly reduced density for the later dataset compared to the first dataset. This corresponds to what was observed for the cross-linked datasets, however in the cross-linked state it took 21 h to reach this state, whereas in the non-cross-linked case it becomes visible after only 2-3 h. Therefore, this could be an indication that cross-linking the proteinase K crystal significantly reduces the enzyme activity in the crystal.



(a) 1 min - 1 h 37 min

(b) 1 h 37 min - 2 h 57 min

Figure 4.28: Electron Density Maps of the Proteinase K Substrate Recognition and Active Sites for the Serially Collected Insulin Soaked Crystal Data 2Fo-Fc electron density maps of the substrate recognition site (Asn99-Tyr104 and Leu133-Gly136) and active site (Asp39, His69, and Ser224) of proteinase K for the time intervals (a) 1 min - 1 h 37 min and (b) 1 h 37 min - 2 h 57 min after mixing with the insulin. The red spheres represent water molecules. All maps are contoured at 1.0σ.

Chapter 5

Discussion

5.1 Sample Preparation for Single Particle Imaging

Both virus samples were classified in terms of size distribution and mono-dispersity. For the mycovirus the DLS results correlate well with the TEM and SAXS results, giving rise to a particle size of 30 nm. The Nanosight results however gave a larger size of 169 nm with a very broad standard deviation of 67 nm. Nanosight instruments are used for particle sizes ranging from a minimum of 10 nm up to 1 µm, therefore the mycovirus particles are right at the edge of the detection range, which could explain the severe overestimation of the size along with the broad standard deviation. Since the Nanosight experiments were conducted at the UKE, it is also conceivable that the particles formed aggregates during transportation. However, this was not observed for any of the other measurements or during the dialysis experiments and is therefore probably not the cause for the deviation in the Nanosight size estimate. In particular, the TEM images showed that the viruses were very mono-disperse with a size of approximately 30 nm.

The adenovirus particles were shown to have a size of approximately 90 nm using DLS and SAXS. Nanosight experiments were not performed for the adenovirus particles due to the conflicting results achieved for the mycovirus sample. Furthermore, in contrast to the mycovirus, the structure of the adenovirus has already been extensively studied [89, 72, 75] and it was therefore mainly interesting for single particle diffractive imaging experiments towards method and algorithm development. To this end, different buffer conditions were tested for both virus samples and it was possible to dialyse both

viruses into buffers that are compatible with aerodynamic lens delivery into vacuum, a pre-requisite for single particle coherent diffractive imaging experiments. SAXS experiments were performed for both virus samples in their SPI compatible solutions to gain structural insights into both virus samples, and a summary is given in Table 5.1.

Table 5.1: DLS and SAXS Results Overview for the Mycovirus and Adenovirus Samples
Summary of the mycovirus and adenovirus molecular weight and size properties obtained from SAXS and DLS experiments.

	DLS	**SAXS**			
	R_h [nm]	R_g [nm]	D_{max} [nm]	Porod volume [nm^3]	MW_{Porod} [MDa]
Mycovirus	22.3 +/- 1.9	16.5	47.5	24 540	14.4
Adenovirus	46.6 +/- 4.7	36.5	88.1	279 395	164.4

From this it can be seen that the size estimation by the various methods are well in agreement, and in particular the sizes are in agreement with what was determined by DLS prior to dialysis. The calculated *ab initio* model for the adenovirus is clearly hexagonal in shape (Figure 4.9), which is compliant with what has been reported previously for this type of virus [72, 75]. The *ab initio* model calculated from the SAXS data for the mycovirus sample (Figure 4.5) reveals a similar ”life-ring” shape as was observed in the TEM images (Figure 4.1). This is the first structural study performed for this particular virus.

The crystallisation trials for both virus samples were unsuccessful. This could be due to the fact that when crystallising virus samples, the solvent channels in the crystal are very large due to the globular shape of the viruses. This often makes the crystals very fragile, making it difficult to find cryo-conditions under which the crystals remain intact and often making it necessary to resort to room temperature diffraction data collection. This was however was not realisable in the case of the adenovirus due to the small size of the crystals. Furthermore small perturbations in the very large unit cells would render the diffraction from these crystals uninterpretable.

In summary, the mycovirus and adenovirus samples were classified and they were both deemed to be appropriate for single particle coherent diffractive imaging experiments. For both samples stable buffer conditions were identified that lend themselves

to aerodynamic injection for SPI experiments, and preliminary structural analysis of the samples in these buffers was performed through the application of SAXS. Unfortunately it was not possible to image the virus particles at FLASH during the course of this work due to extensive technical issues during the SPI beamtime, but the samples have been prepared accordingly and are on stand-by for being measured should more FEL beam time become available.

5.2 Serial Crystallography

5.2.1 SIRAS Phase Determination using LCP Soaking and Crystal Delivery

For the first time SIRAS was successfully applied to SMX data, and can therefore directly be compared to SFX data where SIRAS phasing has proven the most popular approach for phase retrieval. Furthermore, this is also the first time native crystals were derivatized directly in the LCP prior to data collection, and from the results presented this appears to have been a very gentle procedure for introducing heavy atoms into crystals with no significant loss in resolution. Moreover, almost full occupancy of the mercury occupation sites was observed within the first half an hour of data collection, indicating that there is no noticeable detriment to the diffusion time caused by the channel structure of the LCP. Membrane proteins are notoriously difficult to crystallize and often yield very small crystals when grown directly in LCP. However they are of high importance for biological processes and knowledge of the structures would provide essential insights into their function, and in particular aide the development of targeted drugs. In addition, the fact that micro-crystals are highly susceptible to radiation damage often makes the collection of a complete data set on a few single crystals impossible. Furthermore, introducing heavy atoms into crystals often results in the crystals becoming more fragile when handling, less structurally intact and even more sensitive to radiation damage. When taking all of these issues into account, this method of crystal derivatization and data collection can become a powerful tool when direct phasing methods are essential and conventional rotational crystallographic methods have failed. Lastly, the data can be collected using a conventional synchrotron radiation source despite using microcrystals.

In serial crystallography the integrated diffraction intensities are determined by averaging over a large number of still diffraction images using Monte-Carlo integration. Despite the resulting data being accurate enough for structure solution using molecular replacement, until now direct phasing methods have only very rarely been successfully employed [8, 94, 105, 60, 62]. This is due to the fact that despite the averaged intensities often being sufficient for finding the heavy atoms in phased difference Fourier maps, the accuracy is still far too low to enable *de novo* phasing. It has previously been reported

that despite the unit cell parameters varying by less than 0.2% in serially collected data at an FEL for SIRAS phasing with a mercury derivative, phasing was successful using 20 000 images [105]. This corresponds well with what was observed for the proteinase K data presented here, and phasing as well as automatic model building of the structure was successful when using as little as 12 000 images in total.

In a similar study it was reported that 12 000 images, collected at an FEL, were sufficient for SIRAS phasing of bR using a heavy atom derivative [60]. This is in good compliance with what was observed when reducing the number of images used for phasing attempt of the proteinase K synchrotron data. When systematically eliminating known errors in serially collected data at an FEL, Nass et al. (2016) demonstrated SAD phasing of a lysozyme Gd derivative using 7 000 images as well as native sulphur SAD of thaumatin with 125 000 images. Yamashita et. al. (2015) further reported that SAD phasing was not possible when using 80 000 patterns (collected at an FEL), similarly to SAD phasing having failed in this case for proteinase K when using all the 64 665 indexed images (collected at a synchrotron). However, the data collected for the mercury derivative in the work presented here was not collected directly at the absorption edge due to experimental limitations. The theoretical L-III absorption edge for mercury is at a wavelength of 1.009 Å [21] whereas the data presented was collected at 1.033 Å. Therefore the anomalous signal measured is only 40.0% of the theoretical achievable value. It is therefore conceivable that mercury SAD phasing would have been successful if the data had been collected at the ideal wavelength.

Despite SIRAS phasing having found wide application in SFX data [8, 105, 60], employing an FEL X-ray source, the applicability of *de novo* phasing methods for serially collected synchrotron radiation data has not yet been widely explored. The first and only attempt to date was undertaken in 2015, where lysozyme was phased through MIRAS using an iodide as well as a gold derivative [94]. Therefore, this is the first time that serial data collected at a synchrotron were directly phased using SIRAS, resulting in maps of very high quality sufficient for automatically building all but the two terminal residues of the structure. Comparing the FEL studies to the results obtained in this work, it can be seen that serial data, regardless of whether collected at a synchrotron or at an FEL, lends itself well to SIRAS phasing. Sufficient data for *de novo* phasing can be collected

in as little as 6 min at an FEL (LCLS) or 25 min at a synchrotron (assuming a combined hit and indexing rate of 30%). In addition, the soaking method outlined in this thesis for obtaining a heavy atom derivative protein crystal by direct soaking in heavy atom supplemented LCP, is very gentle yet highly effective and applicable to serial data collection at both FEL and synchrotron X-ray sources.

5.2.2 Inhibitory Molecule Diffusion Study

In an attempt to investigate the diffusion time associated with differently sized ligands, varying sizes of proteinase K crystals were soaked for different soaking times with two inhibitory molecules of differing size. The occupancy of the ligand in the structure was subsequently refined, and this was used as a gauge for estimating the diffusion rate into the differently sized crystals. It was expected that three dependences would be observed: Firstly, an increase in the occupancy for the same time point when decreasing the crystal size. Thirdly, a decrease in occupancy associated with shorter soaking times when keeping the crystal size constant. And lastly, faster increases in the occupancies when soaking with boric acid when compared with PMSF, due to comparatively small size of the molecule. For the boric acid case, the results summarised in Table 4.5 show neither of the expected dependences. In fact all of the values determined for the occupancies scatter across the range of 0.6-1.0. This indicates that the metric of refining occupancy is not sensitive enough for detecting the slight variabilities in occupancy that was attempted here. A further explanation could also be that the diffusion times are so short, that data points are sampled where all the crystal sizes are already saturated. However due to experimental constraints, shorter soaking times could not be investigated in the current set-up for the larger crystals. The general range of the occupancies correspond well to other structures deposited in the protein data bank, as shown in Table 5.2. This also supports the fact that the soaking times sampled during the course of these experiments are too long.

Analogously to the boric acid study, the much larger molecule PMSF was also soaked into the same sized crystals and the results are summarised in Table 4.6. Here too, none of the expected trends are observable, and the occupancy values scatter between a range

Table 5.2: Boric Acid Occupancies from PDB Entries
Occupancy and mean B-factor values for this study and three PDB entries.

pdb entry	Occupancy	mean B-factor
Study	0.6-1.0	20.0
1S3T	0.5	16.2
3S42	0.9	19.3
5NJW	1.0	23.4

of 0.1-0.7. The lower occupancies compared to the boric acid case could be due to the large size of the PMS molecule and that when binding to Ser224 of the active site only the sulphur and oxygen atoms are fixed, with the aromatic ring remaining flexible.

Furthermore, it is possible that the time points sampled during this study were too long, meaning that the inhibitor sites were already fully occupied at the shortest time points. Due to the experimental procedure employed, it is however not possible to sample shorter soaking times. Finally, the B-factors of the inhibitory molecules were kept constant at 20.0 to allow for a standardised occupancy refinement procedure. Especially for the PMSF case, where the aromatic ring remains flexible, this is a very artificial constraint. In fact, most PDB entries model the PMS molecule as SO_4.

In summary the results obtained in this study were not conclusive. Extensive knowledge of the diffusion pathways into protein crystals is highly relevant for the feasibility of proposed serial, time-resolved experiments by substrate mixing. The method employed here is however not suitable for investigating timing constraints.

5.2.3 Caged Mercury De-Caging Study

In a further attempt to investigate the effectiveness of diffusing ligands into protein crystals, a mercury cage was synthesized. Upon photo-activation, mercury release is triggered. In particular, many caged compounds are intended to trigger chemical reaction for time-resolved studies, and for larger crystals this could be of benefit towards the timing constraints determined by the required diffusion times of the activation substance (such as ATP) into the crystal. Mercury inhibits proteinase K, and therefore initial experiments were designed and conducted to investigate the timing constraints of diffusing a mercury

cage into the protein crystal and then triggering the release by UV exposure. To this end a photo-activated mercury cage was synthesised and soaked into proteinase K crystals. The subsequent release of mercury was to be triggered using UV light. After synthesizing the mercury cage, the main question that arose was threefold. Firstly, how successfully was the mercury incorporated into the cage, considering that it did not show up in the mass spectrometry analysis. Is this due to the fact that the mercury is too loosely bound inside the cage that it gets lost during the plasmarization in the mass spectrometer or was it never incorporated to begin with. Secondly, could all the excess, non-incorporated mercury be washed out during the cage synthesis so that it does not falsify the subsequent measurement. And thirdly, was the UV light source used sufficient for de-caging the mercury with a high yield in a short period of time, or, contrarily, was the UV proportion of the natural light present during synthesis and handling possibly sufficient for an unintentional de-caging of the mercury. Since no clear indication of the mercury cage acting in the desired way was observed during the initial de-caging experiments, and all of the possibly responsible reasons stated above, no further efforts were invested into these investigations.

5.2.4 Substrate Identification Experiments

During the flowcell experiments diffraction data were collected from proteinase K crystals while these were subjected to a constant flow of fresh substrate. This procedure was tested for three short peptides (4-6 amino acids) and not for a single peptide could any residual electron density be seen in the active or substrate recognition sites. Especially in the case of *Suc-Ala-Ala-Pro-Phe-pNa*, a tetrapeptide that is most commonly used in enzyme assays for testing proteinase K activity, the substrate was clearly being turned over by the crystals, indicated by the colour change from clear to yellow of the substrate solution before and after coming into contact with the crystals. However, nothing was visible in the electron density. This lead to the theory that only the surface molecules of the crystal bind the substrate and cleave it and this process obstructs the solvent channels into the crystals, so that the substrate cannot penetrate deeper into the crystal. A different possibility could be that proteinase K has such a fast turn-over time (turnover numbers as high as 620 s^{-1} have been reported [58]), that substrate is not bound to sufficient crystal molecules at

any one time to give enough signal to show up in the electron density where an average over all molecules in the crystal is shown. Therefore, some longer peptides were tested in addition to some less specific peptides in a simple soaking experiment, where the crystals were left to soak in a high concentration of the various possible substrates for 5 minutes, subsequently flash-cooled and then measured using conventional, rotational methods. It was conceivable that when using a longer peptide, two mechanisms are possible. Firstly, for the scenario where the solvent channels are being blocked when binding the substrate, a longer peptide can be cleaved into fragments, that are themselves cleavable by the enzyme and so on. These fragments would then be released by the outer crystal molecules and thereby diffuse further into the crystal. Or, for the second scenario, where turn-over is so fast that the substrate electron density is too washed out to be visible in the electron density, the long peptide could possibly bind to the substrate recognition site while being successively cleaved into smaller fragments, but during this time it remains bound in the substrate recognition site. As discussed in Section 4.2.5, some small residual differences were seen when applying insulin chain B, oxidized from bovine pancreas, as well as horse heart myoglobin (30 and 154 amino acids respectively). There residual electron densities obtained are however too large to be water molecules and given the very high resolution of the diffraction data (1.15 Å) had to originate from something else. Insulin was chosen due to its still relatively small size (as it is more likely to efficiently diffuse into the crystal) as the substrate of choice and was investigated further. From the SDS-Page and DLS results, it is evident that proteinase K, both in solution as well as in crystal form, cleaves the insulin into fragments small enough that they are no longer detectable by SDS-page or DLS. Therefore, the following cross-linking experiments focused on insulin as the substrate.

5.2.5 Substrate Cross-linking into the Substrate Recognition Site of the Enzyme

In Section 4.2.6 it is shown that conditions could be found where the substrate insulin is cross-linked into the active proteinase K crystal to give substantial, connected electron density in the substrate recognition and active site of the enzyme molecule. The

high connectivity of the density observed is evidence that this is a single molecule and not an agglomerate of solvent molecules. Further, it could be conceivable that this is a glutaraldehyde molecule itself that has bound to the active site of the enzyme, however when building a glutaraldehyde molecule into the observed density the length of the observed density extends to more than twice the size of the molecule. It is therefore the most probable that the density observed is indeed the substrate that has been trapped into the substrate recognition sites of the proteinase K crystal molecules. It is noteworthy to mention that this condition was only reached after the crystal was introduced into the insulin and glutaraldehyde solution and then left to soak for a full 7 days.

Since smaller crystals have smaller diffusion times associated for the substrate, the same procedure was to be performed for micron-sized crystals in a serial data collection approach using a tape drive set-up (Section 3.6.6). Given the severely reduced crystal size from 200 µm to a mere 10 µm, the diffusion pathways through the crystals should be 20 times shorter and it was therefore expected that the whole procedure should happen a lot faster, and possibly larger fragments should be cross-linked into the substrate recognition site for the smaller crystals. This is however not what was observed. Even after 24 h of cross-linking, no significant electron density was visible in the active site. Unfortunately, along with this being the longest soaking time possible given the duration of the synchrotron beam time, it was also observed that the smaller crystals are not resilient enough to undergo the cross-linking treatment for 7 days. The sample prepared 7 days prior to beam time (that still diffracted well a few hours after cross-linking, as tested with powder diffraction (data not shown)) no longer yielded any diffraction at the day of the beam time. However what can clearly be taken from Figure 4.27 is that there are clear changes in the water molecule positions initially bound to the active site. This is evidence of some catalytic activity taking place in the crystal. Analogously to the cross-linked crystals, crystals were also soaked in an abundance of insulin without being cross-linked and measured in the same set-up. From Figure 4.28 it can be seen that the same movement in water molecules was observed in this case, however it happens a lot faster than in the cross-linked case. This gives evidence that the cross-linking process itself is a rather slow process, and that quite probably the incubation time for the serial data collection case simply was not long enough. The difficulty therefore in applying this

method towards capturing the enzyme in a specific state is not yet feasible, but when it is only of interest how and in particular, where, a substrate binds to an enzyme this is a very interesting new method. The model system used for these studies, proteinase K, is known to be highly active and cleaves almost everything with a length of at least 4 amino acids. This makes it a ”worst case” test subject for this kind of analysis, and therefore this method should be applicable to a wide range of enzymes and their respective substrates.

In summary, clear visualisation of bound substrate, in the form of electron density, could be established in an active enzyme crystal using this cross-linking approach. However, currently it is only applicable for conventional, rotational diffraction data collection. Further optimisation of the procedure is required to establish adaptability of this approach to micron- or even nanosized crystals for serial diffraction data collection.

Chapter 6

Conclusion and Outlook

6.1 Sample Preparation for Single Particle Imaging

Small angle X-ray scattering experiments gave first structural insights into the ds-RNA mycovirus. Furthermore, both the mycovirus and the adenovirus were prepared for single particle coherent diffractive imaging experiments but could not be applied due to difficulties during beamtime. Should more beamtime become available, these two viruses could be imaged, as their surface structure and shape are beneficial towards method and, particularly, algorithm development. With the start-up of the European XFEL in Hamburg, with an experimental set-up dedicated to single particle imaging, the possibility of delivering a beam of higher brilliance than any currently operational FEL along with its megahertz repetition rate, it is foreseeable that in a few years single particle coherent diffractive imaging could develop into a standardly applied imaging approach for solving the structures of single biological macromolecules.

6.2 Serial Crystallography

A new method of gently soaking micrometer sized crystals with heavy atoms in lipidic cubic phase for subsequent *de novo* phasing of the structure was developed and presented in this thesis. By introducing the heavy atoms into the lipidic cubic phase and afterwards embedding native crystals, handling the derivatized, fragile crystals is made obsolete, as they can be directly measured in the LCP. Further, the serial diffraction collection approach al-

lows data collection at a synchrotron from crystals as small as a few micrometers, and at an FEL this phasing method could even be applied to nanometer sized crystals. Especially for the structure determination of membrane proteins that are often grown in LCP, as it mimics the native membrane environment, grow in showers of microcrystals, and often do not have a homologous structure for phase retrieval by molecular replacement, this method offers an easily applicable *de novo* phasing approach. Serial Crystallography has grown in popularity over the past decade, and numerous free electron lasers have been and are being built all over the world, and serial femtosecond crystallography has become a standard method for data collection of small crystals. This has lead to many synchrotron beam lines also offering serial data collection capabilities, and it is foreseeable that in the future serial data collection at FELs and synchrotrons is going to grow in popularity. Methods such as the one outlined in this thesis for crystal derivation and subsequent *de novo* phasing are going to become indispensable.

Finally, it was presented that substrate can be cross-linked to the active site of an enzyme in the crystal to an extent that it showed clear electron density. Unfortunately this method could not yet be applied to serial data collection, but it can certainly be applied to find the substrate binding site of specific substrates to an enzyme. This could be a very valuable future technique for targeted drug development, where not only the structure of the targeted enzyme is of importance, but also the catalytic mechanism.

Lastly, different diffusion studies were conducted and it was shown that the mixing and diffusion approach towards time-resolved serial crystallography, especially for synchrotron application, still needs to undergo a lot of timing analysis and method development. Further it is unlikely that this approach will ever be standardised, as all the timing constraints are strongly dependant on the system to be studied.

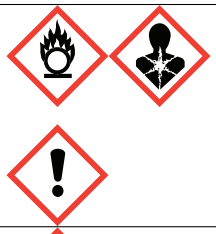
The work done during the course of this thesis has therefore lead to the design of two novel procedures applicable to the wide and ever growing field of protein crystallography. Firstly, a method for heavy-atom derivatization and subsequent *de novo* phasing in a serial diffraction data collection approach applying synchrotron radiation. And secondly, a procedure for fixating a substrate into the active site of an enzyme in a crystal.

Appendix A

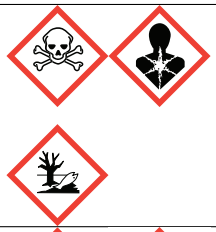
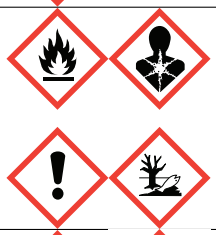
Chemicals and Hazards

A.1 Chemicals

Table A.1: Chemicals with HP-Statements
List of all the chemicals used for this thesis along with their respective hazard (H-) and precautionary (P-) statements (listed alphabetically).

Chemical	GHS Pictograms	H-statements	P-statements
Acetic Acid		226-290-314	280-301+330+331-310-305+351+338
Acrylamide		301-312-315-317-319-332-340-350-361f-372	201-280-302+352-304+340-305+351+338-308+310
Ammonium Acetate	-		
Ammonium Persulfate		272-302-315-317-319-334-335	280-302+352-305+351+338-342+311
Bis-Tris		315-319-335	261-305+351+338
Boric Acid		360fd	201-308+313
Bromophenol Blue	-		

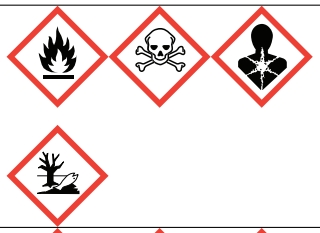
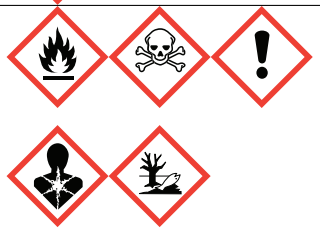
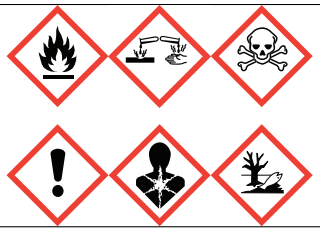
Chemical	GHS Pictograms	H-statements	P-statements
Coomassie Brilliant Blue	-		
DL-α-monoolein	-		
Dimethyl Sulfoxide	-		
Disodium Hydrogen Phosphate	-		
Dithiothreitol		302-315-319	305+351+338
Calcium Chloride		319	305+351+338
Ethanol		225	210
Ethyl Acetate		225-319-336-EUH066	210-233-240-305+351+338-403+235
Glutaraldehyde		301+331+314-317+334+335+400-411	221-273-280-301+310-305+351+338-310
Glycerin	-		
Glycerol	-		
Glycine	-		
Hydrochloric Acid		314-335	260-264-280-301+330+331-303+361+353-304+340-305+351+338-310-321-403+233-405-501
Isopropanol		225-319-336	210-233-240-305+351+338

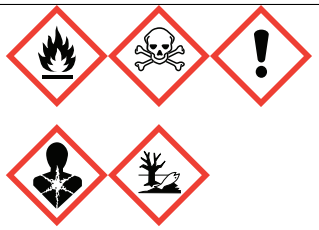
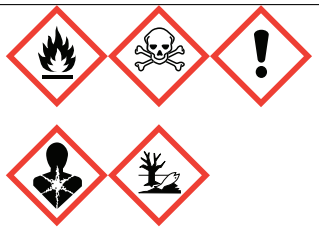
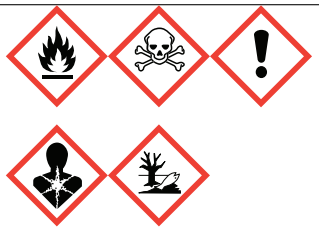
Chemical	GHS Pictograms	H-statements	P-statements
Mercury(II) Acetate		330-310-300-373-410	273-280-302+352-304+340-308+310
Mercury(II) Chloride		341-361f-300-310-372-314-410	201-280-301+310-330-303+361+353+304+340+310-305+351+338
Methanol		225-331-311-301-370	210-233-280-302+352-304+340-308+310-403+235
Nitrogen		280	403
Petroleum Ether		225-304-315-336-361f-373-411	201-210-301+310-331-370+378-501
Phenyl-Methylsulfonyl Fluoride		301-314	260-301+310-303+361+353-305+351+338-405-501
Salicylaldehyde		302-411	273-301+312-330-391-501
Sodium Dodecyl Sulfate		228-302+332-315-318-335-412	210-261-280-301+312+330-305+351+338+310-370+378
Sodium Hydroxide		314-335	280-301+330+331-305+351+338-308+310

Chemical	GHS Pictograms	H-statements	P-statements
Sodium Nitrate		272-319	220-305+351+338
2-Amino-thiophenole		302-314-317-373-400	260-301+330+331-303+361+353-305+351+338-405-501

A.2 Crystallisation Screens

Table A.2: Crystallisation Screens with HP-Statements
List of the crystallisation screens used for this thesis along with their respective hazard (H-) and precautionary (P-) statements.

Crystallisation Screen	GHS Pictograms	H-statements	P-statements
$AmSO_4$-Suite (Quiagen)		225-301-330-350-340-360FD-372-411	101-201-273-280-309+311
ComPAS-Suite (Quiagen)		225-301-302-315-319-331-332-335-340-350-360FD-373-411	101-201-270-273-280-305+351+338-309+311-313
JCSG-plus (Molecular Dimensions)		225-301-312-315-318-331-335-350-411	101-201-270-280-305+351+338-309+311-313

Crystallisation Screen	GHS Pictograms	H-statements	P-statements
Morpheus (Molecular Dimensions)		225-301-302-315-319-331-332-335-340-350-360FD-373-411	101-201-270-273-280-305+351+338-309+311-313
PACT premier (Molecular Dimensions)		301-331-412	101-270-273-280-309+311
Structure (Molecular Dimensions)		225-301-302-315-319-331-332-335-340-350-360FD-373-411	101-201-270-273-280-305+351+338-309+311-313
SturaFootprint& Macrosol (Molecular Dimensions)		225-301-302-315-319-331-332-335-340-350-360FD-373-411	101-201-270-273-280-305+351+338-309+311-313

A.3 GHS and Risk Symbols

Figure A.1: GHS Pictograms
Summary of all GHS pictograms along with the respective nomenclature.

A.4 Hazard Statements

H225	Highly flammable liquid and vapour.
H226	Flammable liquid and vapour.
H228	Flammable solid.
H272	May intensify fire; oxidiser.
H280	Contains gas under pressure; may explode if heated.
H290	May be corrosive to metals.
H300	Fatal if swallowed.
H301	Toxic if swallowed.
H302	Harmful if swallowed.
H302+H332	Harmful if swallowed or if inhaled.
H304	May be fatal if swallowed and enters airways.
H310	Fatal in contact with skin.
H311	Toxic in contact with skin.
H312	Harmful in contact with skin.
H314	Causes severe skin burns and eye damage.
H315	Causes skin irritation.
H317	May cause an allergic skin reaction.
H318	Harmful if swallowed or if inhaled.
H319	Causes serious eye irritation.
H330	Fatal if inhaled.
H331	Toxic if inhaled.
H332	Harmful if inhaled.
H334	May cause allergy or asthma symptoms or breathing difficulties if inhaled.
H335	May cause respiratory irritation.
H336	May cause drowsiness or dizziness.
H340	May cause genetic defects.
H341	Suspected of causing genetic defects.
H350	May cause cancer.
H360fd	May damage fertility or the unborn child.
H361f	Suspected of damaging fertility.
H370	Causes damage to organs.
H372	Causes damage to organs through prolonged or repeated exposure.
H373	May cause damage to organs through prolonged or repeated exposure.
H400	Very toxic to aquatic life.
H410	Very toxic to aquatic life with long lasting effects.
H411	Toxic to aquatic life with long lasting effects.
H412	Harmful to aquatic life with long lasting effects.
EUH066	Repeated exposure may cause skin dryness or cracking.

A.5 Precautionary Statements

P101	If medical advice is needed, have product container or label at hand.
P201	Obtain special instructions before use.
P210	Keep away from heat, hot surfaces, sparks, open flames and other ignition sources. No smoking.
P220	Keep away from clothing and other combustible materials.
P233	Keep container tightly closed.
P240	Ground and bond container and receiving equipment.
P260	Do not breathe dust/fume/gas/mist/vapours/spray.
P261	Avoid breathing dust/fume/gas/mist/vapours/spray.
P264	Wash thoroughly after handling.
P270	Do not eat, drink or smoke when using this product.
P273	Avoid release to the environment.
P280	Wear protective gloves/protective clothing/eye protection/face protection.
P301+310	IF SWALLOWED: Immediately call a POISON CENTRE or doctor/physician.
P301+312	IF SWALLOWED: Call a POISON CENTRE or doctor if you feel unwell.
P301+312+330	IF SWALLOWED: Call a POISON CENTRE or doctor/physician if you feel unwell. Rinse mouth.
P301+330+331	IF SWALLOWED: Rinse mouth. Do NOT induce vomiting.
P302+352	IF ON SKIN: Wash with plenty of soap and water.
P303+361+353	IF ON SKIN (or hair): Remove/Take off immediately all contaminated clothing. Rinse skin with water or shower.
P304+340	IF INHALED: Remove person to fresh air and keep comfortable for breathing.
P304+340+310	IF INHALED: Remove person to fresh air and keep comfortable for breathing. Immediately call a POISON CENTRE or doctor/physician.
P305+351+338	IF IN EYES: Rinse cautiously with water for several minutes. Remove contact lenses, if present and easy to do. Continue rinsing.
P305+351+338+310	IF IN EYES: Rinse cautiously with water for several minutes. Remove contact lenses, if present and easy to do. Continue rinsing. Immediately call a POISON CENTRE or doctor/physician.
P308+310	IF exposed or concerned: Immediately call a POISON CENTRE or doctor.
P308+313	IF exposed or concerned: Get medical advice/attention.
P309+311	IF exposed or you feel unwell: Call a POISON CENTRE or doctor/physician.
P310	Immediately call a POISON CENTRE or doctor.
P321	Specific treatment (see respective MSDS).
P330	Rinse mouth.
P331	Do NOT induce vomiting.

P342+311	If experiencing respiratory symptoms: Call a POISON CENTRE or doctor.
P370+378	In case of fire: Use dry sand, dry extinguishing powder or alcohol-resistant foam for extinction.
P391	Collect spillage.
P403	Store in a well-ventilated place.
P403+233	Store in a well-ventilated place. Keep container tightly closed.
P403+235	Store in a well-ventilated place. Keep cool.
P405	Store locked up.
P501	Dispose of contents / container in accordance with local / regional / national / international regulations.

Appendix B

List of Equipment

Table B.1: Equipment List
List of all the equipment used for this thesis (listed alphabetically) along with the instrument type and manufacturer.

Device	Instrument	Manufacturer
Beamlines	P11	DESY, PETRA III, DESY, Hamburg
	P12	EMBL, PETRA III, DESY, Hamburg
	P13	EMBL, PETRA III, DESY, Hamburg
	P14	EMBL, PETRA III, DESY, Hamburg
Laminar Flow Pump	Biorad Econo Pump	Biorad Laboratories Inc., USA
Centrifuges	5415R/ 5415C/ 5804R/ 5810R MinispinPlus	Eppendorf, Germany
	Thinky ARE-250	Intertronics, UK
Crystallisation Robots	Honeybee 961	Genomic Solutions, USA
	Oryx4	Douglas Instruments, UK
DLS Device	SpectroLight 300	Xtal Concepts GmbH
	SpectroLight 600	Xtal Concepts GmbH
Freezer -20°C	Liebherr premium	Liebherr, Germany
	B35-85	FRYKA-Kältetechnik, Germany
Imaging	CLSM Zeiss	Zeiss, Germany
	CrystalScore	Diversified Scientific Inc., USA
	Microscope SZX12 with camera DP10	Olympus, Japan

Device	Instrument	Manufacturer
Incubator	RUMED 3001	Rubarth, Germany
Magnetic Stirrer	VMS-A	VWR International
Microbalance	Sartorius TE3102S CP2245-OCE	Sartorius, Germany Sartorius
Nanosight	LM10	Malvern Instruments
pH Meter	SevenEasy	Mettler-Toledo, Switzerland
Photospectrometry	Nanodrop 2000c	ThermoScientific, Peqlab, Germany
Pipetts	Eppendorf Research, 2 µL, 10 µL, 200 µL, 1000 µL	Eppendorf, Germany
Pipetting robots	Lissy	Zinsser, Germany
Plasma cleaner	Zepto	Diener electronic, Germany
SDS-PAGE	Four Gel Caster (SE275) EV 231 (Voltage Source) SE260 Mighty Small II Deluxe Mini	Hoefer, USA Peqlab, Germany Hoefer, USA
Thermomixer	Comfort	Eppendorf
Transmission Electron Microscope	FEI Tecnai G20	FEI, Germany
UV-light source	CrystalLIGHT 100	Nabitec, Germany
Vortex Mixer	VF2	Janke & Kunkel, IKA Labortechnik, Germany
X-ray source	IµS Imaging Plate X-Ray Detection System	Incoatec, Germany Mar Research, Germany

Bibliography

[1] Paul D. Adams, Pavel V. Afonine, Gábor Bunkóczi, Vincent B. Chen, Ian W. Davis, Nathaniel Echols, Jeffrey J. Headd, Li-Wei Hung, Gary J. Kapral, Ralf W. Grosse-Kunstleve, Airlie J. McCoy, Nigel W. Moriarty, Robert Oeffner, Randy J. Read, David C. Richardson, Jane S. Richardson, Thomas C. Terwilliger, and Peter H. Zwart. *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D*, 66(2):213–221, Feb 2010.

[2] A. Aquila, A. Barty, C. Bostedt, S. Boutet, G. Carini, D. dePonte, P. Drell, S. Doniach, K. H. Downing, T. Earnest, H. Elmlund, V. Elser, M. Gühr, J. Hajdu, J. Hastings, S. P. Hau-Riege, Z. Huang, E. E. Lattman, F. R. N. C. Maia, S. Marchesini, A. Ourmazd, C. Pellegrini, R. Santra, I. Schlichting, C. Schroer, J. C. H. Spence, I. A. Vartanyants, S. Wakatsuki, W. I. Weis, and G. J. Williams. The linac coherent light source single particle imaging road map. *Structural Dynamics*, 2(4):041701, 2015.

[3] Andrew Aquila, Mark S Hunter, R Bruce Doak, Richard A Kirian, Petra Fromme, Thomas A White, Jakob Andreasson, David Arnlund, Saša Bajt, Thomas R M Barends, Miriam Barthelmess, Michael J Bogan, Christoph Bostedt, Hervé Bottin, John D Bozek, Carl Caleman, Nicola Coppola, Jan Davidsson, Daniel P DePonte, Veit Elser, Sascha W Epp, Benjamin Erk, Holger Fleckenstein, Lutz Foucar, Matthias Frank, Raimund Fromme, Heinz Graafsma, Ingo Grotjohann, Lars Gumprecht, Janos Hajdu, Christina Y Hampton, Andreas Hartmann, Robert Hartmann, Stefan Hau-Riege, Günter Hauser, Helmut Hirsemann, Peter Holl, James M Holton, André Hömke, Linda Johansson, Nils Kimmel, Stephan Kassemeyer, Faton Krasniqi, Kai-Uwe Kühnel, Mengning Liang, Lukas Lomb, Erik Malmerberg, Stefano Marchesini, Andrew V Martin, Filipe RNC Maia, Marc Messerschmidt,

Karol Nass, Christian Reich, Richard Neutze, Daniel Rolles, Benedikt Rudek, Artem Rudenko, Ilme Schlichting, Carlo Schmidt, Kevin E Schmidt, Joachim Schulz, M Marvin Seibert, Robert L Shoeman, Raymond Sierra, Heike Soltau, Dmitri Starodub, Francesco Stellato, Stephan Stern, Lothar Strüder, Nicusor Timneanu, Joachim Ullrich, Xiaoyu Wang, Garth J Williams, Georg Weidenspointner, Uwe Weierstall, Cornelia Wunderer, Anton Barty, John C H Spence, and Henry N Chapman. Time-resolved protein nanocrystallography using an X-ray free-electron laser. *Optics Express*, 20:2706–2716, 2012.

[4] Peter Atkins and Julio de Paula. *Atkins' Physical Chemistry*. Oxford University Press, Great Britain, eighth edition, 2006.

[5] Thomas Barends, Thomas A. White, Anton Barty, Lutz Foucar, Marc Messerschmidt, Roberto Alonso-Mori, **Sabine Botha**, Henry Chapman, R. Bruce Doak, Lorenzo Galli, Cornelius Gati, Matthias Gutmann, Jason Koglin, Anders Markvardsen, Karol Nass, Dominik Oberthur, Robert L. Shoeman, Ilme Schlichting, and Sébastien Boutet. Effects of self-seeding and crystal post-selection on the quality of Monte Carlo-integrated SFX data. *Journal of Synchrotron Radiation*, 22(3):644–652, May 2015.

[6] Thomas Barends, Thomas A. White, Anton Barty, Lutz Foucar, Marc Messerschmidt, Roberto Alonso-Mori, **Sabine Botha**, Henry Chapman, R. Bruce Doak, Lorenzo Galli, Cornelius Gati, Matthias Gutmann, Jason Koglin, Anders Markvardsen, Karol Nass, Dominik Oberthur, Robert L. Shoeman, Ilme Schlichting, and Sébastien Boutet. Effects of self-seeding and crystal post-selection on the quality of Monte Carlo-integrated SFX data. *Journal of Synchrotron Radiation*, 22(3):644–652, May 2015.

[7] Thomas R. M. Barends, Lutz Foucar, Albert Ardevol, Karol Nass, Andrew Aquila, **Sabine Botha**, R. Bruce Doak, Konstantin Falahati, Elisabeth Hartmann, Mario Hilpert, Marcel Heinz, Matthias C. Hoffmann, Jürgen Köfinger, Jason E. Koglin, Gabriela Kovacsova, Mengning Liang, Despina Milathianaki, Henrik T. Lemke, Jochen Reinstein, Christopher M. Roome, Robert L. Shoeman, Garth J. Williams,

Irene Burghardt, Gerhard Hummer, Sébastien Boutet, and Ilme Schlichting. Direct observation of ultrafast collective motions in co myoglobin upon ligand dissociation. *Science*, 350(6259):445–450, 2015.

[8] Thomas R. M. Barends, Lutz Foucar, **Sabine Botha**, R. Bruce Doak, Robert L. Shoeman, Karol Nass, Jason E. Koglin, Garth J. Williams, Sebastien Boutet, Marc Messerschmidt, and Ilme Schlichting. De novo protein crystal structure determination from X-ray free-electron laser data. *Nature*, 505:244–247, Jan 2014.

[9] Anton Barty, Richard A. Kirian, Filipe R. N. C. Maia, Max Hantke, Chun Hong Yoon, Thomas A. White, and Henry Chapman. *Cheetah*: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *Journal of Applied Crystallography*, 47(3):1118–1131, Jun 2014.

[10] Christian Betzel, S. Gourinath, Pravindra Kumar, Punit Kaur, Markus Perbandt, Susanne Eschenburg, and Tej P. Singh. Structure of a Serine Protease Proteinase K from Tritirachium album limber at 0.98 Å Resolution. *Biochemistry*, 40(10):3080–3088, 2001.

[11] Kenneth R. Beyerlein, Dennis Dierksmeyer, Valerio Mariani, Manuela Kuhn, Iosifina Sarrou, Angelica Ottaviano, Salah Awel, Juraj Knoska, Silje Fuglerud, Olof Jönsson, Stephan Stern, Max O. Wiedorn, Oleksandr Yefanov, Luigi Adriano, Richard Bean, Anja Burkhardt, Pontus Fischer, Michael Heymann, Daniel A. Horke, Katharina E. J. Jungnickel, Elena Kovaleva, Olga Lorbeer, Markus Metz, Jan Meyer, Andrew Morgan, Kanupriya Pande, Saravanan Panneerselvam, Carolin Seuring, Aleksandra Tolstikova, Julia Lieske, Steve Aplin, Manfred Roessle, Thomas A. White, Henry N. Chapman, Alke Meents, and Dominik Oberthuer. Mix-and-diffuse serial synchrotron crystallography. *IUCrJ*, 4(6), Nov 2017.

[12] Sébastien Boutet, Lutz Foucar, Thomas R. M. Barends, **Sabine Botha**, R. Bruce Doak, Jason E. Koglin, Marc Messerschmidt, Karol Nass, Ilme Schlichting, M. Marvin Seibert, Robert L. Shoeman, and Garth J. Williams. Characterization and use of the spent beam for serial operation of LCLS. *Journal of Synchrotron Radiation*, 22(3):634–643, May 2015.

[13] Sébastien Boutet, Lukas Lomb, Garth J. Williams, Thomas R. M. Barends, Andrew Aquila, R. Bruce Doak, Uwe Weierstall, Daniel P. DePonte, Jan Steinbrener, Robert L. Shoeman, Marc Messerschmidt, Anton Barty, Thomas A. White, Stephan Kassemeyer, Richard A. Kirian, M. Marvin Seibert, Paul A. Montanez, Chris Kenney, Ryan Herbst, Philip Hart, Jack Pines, Gunther Haller, Sol M. Gruner, Hugh T. Philipp, Mark W. Tate, Marianne Hromalik, Lucas J. Koerner, Niels van Bakel, John Morse, Wilfred Ghonsalves, David Arnlund, Michael J. Bogan, Carl Caleman, Raimund Fromme, Christina Y. Hampton, Mark S. Hunter, Linda C. Johansson, Gergely Katona, Christopher Kupitz, Mengning Liang, Andrew V. Martin, Karol Nass, Lars Redecke, Francesco Stellato, Nicusor Timneanu, Dingjie Wang, Nadia A. Zatsepin, Donald Schafer, James Defever, Richard Neutze, Petra Fromme, John C. H. Spence, Henry N. Chapman, and Ilme Schlichting. High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. *Science*, 337(6092):362–364, 2012.

[14] Anja Burkhardt, Tim Pakendorf, Bernd Reime, Jan Meyer, Pontus Fischer, Nicolas Stübe, Saravanan Panneerselvam, Olga Lorbeer, Karolina Stachnik, Martin Warmer, Philip Rödig, Dennis Göries, and Alke Meents. Status of the crystallography beamlines at PETRA III. *The European physical journal / Plus*, 131(3):56, 2016.

[15] Henry N. Chapman, Anton Barty, Michael J. Bogan, Sébastien Boutet, Matthias Frank, Stefan P. Hau-Riege, Stefano Marchesini, Bruce W. Woods, Saša Bajt, W. Henry Benner, Richard A. London, Elke Plönjes, Marion Kuhlmann, Rolf Treusch, Stefan Düsterer, Thomas Tschentscher, Jochen R. Schneider, Eberhard Spiller, Thomas Möller, Christoph Bostedt, Matthias Hoener, David A. Shapiro, Keith O. Hodgson, David van der Spoel, Florian Burmeister, Magnus Bergh, Carl Caleman, Gösta Huldt, M. Marvin Seibert, Filipe R. N. C. Maia, Richard W. Lee, Abraham Szöke, Nicusor Timneanu, and Janos Hajdu. Femtosecond diffractive imaging with a soft-x-ray free-electron laser. *Nature Physics*, 2:839, 2006.

[16] Henry N. Chapman, Petra Fromme, Anton Barty, Thomas A. White, Richard A. Kirian, Andrew Aquila, Mark S. Hunter, Joachim Schulz, Daniel P. DePonte,

Uwe Weierstall, R. Bruce Doak, Filipe R. N. C. Maia, Andrew V. Martin, Ilme Schlichting, Lukas Lomb, Nicola Coppola, Robert L. Shoeman, Sascha W. Epp, Robert Hartmann, Daniel Rolles, Artem Rudenko, Lutz Foucar, Nils Kimmel, Georg Weidenspointner, Mengning Liang, Miriam Barthelmess, Carl Caleman, Sebastien Boutet, Michael J. Bogan, Jacek Krzywinski, Christoph Bostedt, Sasa Bajt, Lars Gumprecht, Benedikt Rudek, Benjamin Erk, Carlo Schmidt, Andre Homke, Christian Reich, Daniel Pietschner, Lothar Struder, Gunter Hauser, Hubert Gorke, Joachim Ullrich, Sven Herrmann, Gerhard Schaller, Florian Schopper, Heike Soltau, Kai-Uwe Kuhnel, Marc Messerschmidt, John D. Bozek, Stefan P. Hau-Riege, Matthias Frank, Christina Y. Hampton, Raymond G. Sierra, Dmitri Starodub, Garth J. Williams, Janos Hajdu, Nicusor Timneanu, M. Marvin Seibert, Jakob Andreasson, Andrea Rocker, Olof Jonsson, Martin Svenda, Stephan Stern, Karol Nass, Robert Andritschke, Claus-Dieter Schroter, Faton Krasniqi, Mario Bott, Kevin E. Schmidt, Xiaoyu Wang, Ingo Grotjohann, James M. Holton, Thomas R. M. Barends, Richard Neutze, Stefano Marchesini, Raimund Fromme, Sebastian Schorb, Daniela Rupp, Marcus Adolph, Tais Gorkhover, Inger Andersson, Helmut Hirsemann, Guillaume Potdevin, Heinz Graafsma, Bjorn Nilsson, and John C. H. Spence. Femtosecond X-ray protein nanocrystallography. *Nature*, 470:73–77, Feb 2011.

[17] Anchi Cheng, Bernard Hummel, Hong Qiu, and Martin Caffrey. A simple mechanical mixer for small viscous lipid-containing samples. *Chemistry and Physics of Lipids*, 95(1):11 – 21, 1998.

[18] Robert K.Y. Cheng, Rafael Abela, and Michael Hennig. X-ray free electron laser: opportunities for drug discovery. *Essays In Biochemistry*, 61(5):529–542, 2017.

[19] Chelsie E Conrad, Shibom Basu, Daniel James, Dingjie Wang, Alexander Schaffer, Shatabdi Roy-Chowdhury, Nadia A Zatsepin, Andrew Aquila, Jesse Coe, Cornelius Gati, Mark S Hunter, Jason E Koglin, Christopher Kupitz, Garrett Nelson, Ganesh Subramanian, Thomas A White, Yun Zhao, James Zook, Sébastien Boutet, Vadim Cherezov, John C H Spence, Raimund Fromme, Uwe Weierstall, and Petra

Fromme. A novel inert crystal delivery medium for serial femtosecond crystallography. *IUCrJ*, 2:421–430, 2015.

[20] K. Cowtan. 'DM': An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, 31:34–38, 1994.

[21] D. T. Cromer. Anomalous dispersion corrections computed from self-consistent field relativistic Dirac–Slater wave functions. *Acta Crystallographica*, 18(1):17–23, Jan 1965.

[22] Hasan Demirci, Raymond G. Sierra, Hartawan Laksmono, Robert L. Shoeman, **Sabine Botha**, Thomas R. M. Barends, Karol Nass, Ilme Schlichting, R. Bruce Doak, Cornelius Gati, Garth J. Williams, Sébastien Boutet, Marc Messerschmidt, Gerwald Jogl, Albert E. Dahlberg, Steven T. Gregory, and Michael J. Bogan. Serial femtosecond X-ray diffraction of 30S ribosomal subunit microcrystals in liquid suspension at ambient temperature using an X-ray free-electron laser. *Acta Crystallographica Section F*, 69(9):1066–1069, Sep 2013.

[23] D P DePonte, U Weierstall, K Schmidt, J Warner, D Starodub, J C H Spence, and R B Doak. Gas dynamic virtual nozzle for generation of microscopic droplet streams. *Journal of Physics D: Applied Physics*, 41(19):195505, 2008.

[24] A. J. M. Duisenberg. Indexing in single-crystal diffractometry with an obstinate list of reflections. *Journal of Applied Crystallography*, 25(2):92–96, Apr 1992.

[25] Tomas Ekeberg, Martin Svenda, Chantal Abergel, Filipe R. N. C. Maia, Virginie Seltzer, Jean-Michel Claverie, Max Hantke, Olof Jönsson, Carl Nettelblad, Gijs van der Schot, Mengning Liang, Daniel P. DePonte, Anton Barty, M. Marvin Seibert, Bianca Iwan, Inger Andersson, N. Duane Loh, Andrew V. Martin, Henry Chapman, Christoph Bostedt, John D. Bozek, Ken R. Ferguson, Jacek Krzywinski, Sascha W. Epp, Daniel Rolles, Artem Rudenko, Robert Hartmann, Nils Kimmel, and Janos Hajdu. Three-dimensional reconstruction of the giant mimivirus particle with an x-ray free-electron laser. *Phys. Rev. Lett.*, 114:098102, Mar 2015.

[26] Paul Emsley, Bernhard Lohkamp, William G. Scott, and Kevin Cowtan. Features and Development of Coot. *Acta Crystallographica Section D - Biological Crystallography*, 66:486–501, 2010.

[27] Yiping Feng, Roberto Alonso-Mori, Thomas. R. M. Barends, Vladimir D. Blank, **Sabine Botha**, Matthieu Chollet, D. S. Damiani, R. B. Doak, J. M. Glownia, Jason M. Koglin, Hilmar T. Lemke, Marc Messerschmidt, Karol Nass, S. Nelson, Ilme Schlichting, Robert L. Shoeman, Yu. V. Shvyd'ko, M. Sikorski, Sanghoon Song, Stanislav Stoupin, S. Terentyev, Garth J. Williams, Diwen Zhu, A. Robert, and Sebastien Boutet. Demonstration of simultaneous experiments using thin crystal multiplexing at the Linac Coherent Light Source. *Journal of Synchrotron Radiation*, 22(3):626–633, May 2015.

[28] Lutz Foucar, Anton Barty, Nicola Coppola, Robert Hartmann, Peter Holl, Uwe Hoppe, Stephan Kassemeyer, Nils Kimmel, Jochen Küpper, Mirko Scholz, Simone Techert, Thomas A. White, Lothar Strüder, and Joachim Ullrich. Cass—cfel-asg software suite. *Computer Physics Communications*, 183(10):2207 – 2213, 2012.

[29] D. Franke, M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H. D. T. Mertens, A. G. Kikhney, N. R. Hajizadeh, J. M. Franklin, C. M. Jeffries, and D. I. Svergun. *ATSAS 2.8*: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography*, 50(4):1212–1225, Aug 2017.

[30] Daniel Franke and Dmitri I. Svergun. *DAMMIF*, a program for rapid *ab-initio* shape determination in small-angle scattering. *Journal of Applied Crystallography*, 42(2):342–346, Apr 2009.

[31] W. Friedrich, Knipping P., and M. Laue. Interferenzerscheinungen bei röntgenstrahlen. *Ann. Phys.*, 346:971–988, 1913.

[32] Lorenzo Galli, Sang-Kil Son, Thomas R. M. Barends, Thomas A. White, Anton Barty, **Sabine Botha**, Sébastien Boutet, Carl Caleman, R. Bruce Doak, Max H. Nanao, Karol Nass, Robert L. Shoeman, Nicusor Timneanu, Robin Santra, Ilme

Schlichting, and Henry N. Chapman. Towards phasing using high X-ray intensity. *IUCrJ*, 2(6):627–634, Nov 2015.

[33] Cornelius Gati, Gleb Bourenkov, Marco Klinge, Dirk Rehders, Francesco Stellato, Dominik Oberthür, Oleksandr Yefanov, Benjamin P. Sommer, Stefan Mogk, Michael Duszenko, Christian Betzel, Thomas R. Schneider, Henry N. Chapman, and Lars Redecke. Serial crystallography on *in vivo* grown microcrystals using synchrotron radiation. *IUCrJ*, 1(2):87–94, Mar 2014.

[34] Cornelius Gati, Dominik Oberthuer, Oleksandr Yefanov, Richard D. Bunker, Francesco Stellato, Elaine Chiu, Shin-Mei Yeh, Andrew Aquila, Shibom Basu, Richard Bean, Kenneth R. Beyerlein, **Sabine Botha**, Sébastien Boutet, Daniel P. DePonte, R. Bruce Doak, Raimund Fromme, Lorenzo Galli, Ingo Grotjohann, Daniel R. James, Christopher Kupitz, Lukas Lomb, Marc Messerschmidt, Karol Nass, Kimberly Rendek, Robert L. Shoeman, Dingjie Wang, Uwe Weierstall, Thomas A. White, Garth J. Williams, Nadia A. Zatsepin, Petra Fromme, John C. H. Spence, Kenneth N. Goldie, Johannes A. Jehle, Peter Metcalf, Anton Barty, and Henry N. Chapman. Atomic structure of granulin determined from native nanocrystalline granulovirus using an x-ray free-electron laser. *Proceedings of the National Academy of Sciences*, 114(9):2247–2252, 2017.

[35] Alexander Gorel, Koji Motomura, Hironobu Fukuzawa, R. Bruce Doak, Marie Luise Grünbein, Mario Hilpert, Ichiro Inoue, Marco Kloos, Gabriela Kovácsová, Eriko Nango, Karol Nass, Christopher M. Roome, Robert L. Shoeman, Rie Tanaka, Kensuke Tono, Yasumasa Joti, Makina Yabashi, So Iwata, Lutz Foucar, Kiyoshi Ueda, Thomas R. M. Barends, and Ilme Schlichting. Multi-wavelength anomalous diffraction de novo phasing using a two-colour X-ray free-electron laser with wide tunability. *Nature Communications*, 8, 2017.

[36] Max F. Hantke, Dirk Hasse, Filipe R. N. C. Maia, Tomas Ekeberg, Katja John, Martin Svenda, N. Duane Loh, Andrew V. Martin, Nicusor Timneanu, Daniel S. D. Larsson, Gijs van der Schot, Gunilla H. Carlsson, Margareta Ingelman, Jakob Andreasson, Daniel Westphal, Mengning Liang, Francesco Stellato, Daniel P. De-

Ponte, Robert Hartmann, Nils Kimmel, Richard A. Kirian, M. Marvin Seibert, Kerstin Mühlig, Sebastian Schorb, Ken Ferguson, Christoph Bostedt, Sebastian Carron, John D. Bozek, Daniel Rolles, Artem Rudenko, Sascha Epp, Henry N. Chapman, Anton Barty, Janos Hajdu, and Inger Andersson. High-throughput imaging of heterogeneous cell organelles with an X-ray laser. *Nature Photonics*, 8:943, 2014.

[37] Johan Hattne, Nathaniel Echols, Rosalie Tran, Jan Kern, Richard J Gildea, Aaron S Brewster, Roberto Alonso-Mori, Carina Glöckner, Julia Hellmich, Hartawan Laksmono, Raymond G Sierra, Benedikt Lassalle-Kaiser, Alyssa Lampe, Guangye Han, Sheraz Gul, Dörte DiFiore, Despina Milathianaki, Alan R Fry, Alan Miahnahri, William E White, Donald W Schafer, M Marvin Seibert, Jason E Koglin, Dimosthenis Sokaras, Tsu-Chien Weng, Jonas Sellberg, Matthew J Latimer, Pieter Glatzel, Petrus H Zwart, Ralf W Grosse-Kunstleve, Michael J Bogan, Marc Messerschmidt, Garth J Williams, Sébastien Boutet, Johannes Messinger, Athina Zouni, Junko Yano, Uwe Bergmann, Vittal K Yachandra, Paul D Adams, and Nicholas K Sauter. Accurate macromolecular structures using minimal measurements from x-ray free-electron lasers. *Nature Methods*, 11:545, 2014.

[38] https://www.sigmaaldrich.com/. Chemical formula for glutaraldehyde by sigma aldrich.

[39] Wolfgang Kabsch. *XDS*. *Acta Crystallographica Section D*, 66(2):125–132, Feb 2010.

[40] Wolfgang Kabsch. Processing of X-ray snapshots from crystals in random orientations. *Acta Crystallographica Section D*, 70(8):2204–2216, Aug 2014.

[41] Stephan Kassemeyer, Jan Steinbrener, Lukas Lomb, Elisabeth Hartmann, Andrew Aquila, Anton Barty, Andrew V. Martin, Christina Y. Hampton, Saša Bajt, Miriam Barthelmess, Thomas R.M. Barends, Christoph Bostedt, Mario Bott, John D. Bozek, Nicola Coppola, Max Cryle, Daniel P. DePonte, R. Bruce Doak, Sascha W. Epp, Benjamin Erk, Holger Fleckenstein, Lutz Foucar, Heinz Graafsma, Lars Gumprecht, Andreas Hartmann, Robert Hartmann, Günter Hauser, Helmut Hirse-

mann, André Hömke, Peter Holl, Olof Jönsson, Nils Kimmel, Faton Krasniqi, Mengning Liang, Filipe R.N.C. Maia, Stefano Marchesini, Karol Nass, Christian Reich, Daniel Rolles, Benedikt Rudek, Artem Rudenko, Carlo Schmidt, Joachim Schulz, Robert L. Shoeman, Raymond G. Sierra, Heike Soltau, John C. H. Spence, Dmitri Starodub, Francesco Stellato, Stephan Stern, Gunter Stier, Martin Svenda, Georg Weidenspointner, Uwe Weierstall, Thomas A. White, Cornelia Wunderer, Matthias Frank, Henry N. Chapman, Joachim Ullrich, Lothar Strüder, Michael J. Bogan, and Ilme Schlichting. Femtosecond free-electron laser x-ray diffraction data sets for algorithm development. *Opt. Express*, 20(4):4149–4158, Feb 2012.

[42] Jan Kern, Roberto Alonso-Mori, Rosalie Tran, Johan Hattne, Richard J. Gildea, Nathaniel Echols, Carina Glöckner, Julia Hellmich, Hartawan Laksmono, Raymond G. Sierra, Benedikt Lassalle-Kaiser, Sergey Koroidov, Alyssa Lampe, Guangye Han, Sheraz Gul, Dörte DiFiore, Despina Milathianaki, Alan R. Fry, Alan Miahnahri, Donald W. Schafer, Marc Messerschmidt, M. Marvin Seibert, Jason E. Koglin, Dimosthenis Sokaras, Tsu-Chien Weng, Jonas Sellberg, Matthew J. Latimer, Ralf W. Grosse-Kunstleve, Petrus H. Zwart, William E. White, Pieter Glatzel, Paul D. Adams, Michael J. Bogan, Garth J. Williams, Sébastien Boutet, Johannes Messinger, Athina Zouni, Nicholas K. Sauter, Vittal K. Yachandra, Uwe Bergmann, and Junko Yano. Simultaneous femtosecond x-ray spectroscopy and diffraction of photosystem ii at room temperature. *Science*, 340(6131):491–495, 2013.

[43] Jan Kern, Rosalie Tran, Roberto Alonso-Mori, Sergey Koroidov, Nathaniel Echols, Johan Hattne, Mohamed Ibrahim, Sheraz Gul, Hartawan Laksmono, Raymond G. Sierra, Richard J. Gildea, Guangye Han, Julia Hellmich, Benedikt Lassalle-Kaiser, Ruchira Chatterjee, Aaron S. Brewster, Claudiu A. Stan, Carina Glöckner, Alyssa Lampe, Dörte DiFiore, Despina Milathianaki, Alan R. Fry, M. Marvin Seibert, Jason E. Koglin, Erik Gallo, Jens Uhlig, Dimosthenis Sokaras, Tsu-Chien Weng, Petrus H. Zwart, David E. Skinner, Michael J. Bogan, Marc Messerschmidt, Pieter Glatzel, Garth J. Williams, Sébastien Boutet, Paul D. Adams, Athina Zouni, Johannes Messinger, Nicholas K. Sauter, Uwe Bergmann, Junko Yano, and Vittal K.

Yachandra. Taking snapshots of photosynthetic water oxidation using femtosecond x-ray diffraction and spectroscopy. *Nature Communications*, 5:4371, 2014.

[44] Richard A. Kirian, Xiaoyu Wang, Uwe Weierstall, Kevin E. Schmidt, John C. H. Spence, Mark Hunter, Petra Fromme, Thomas White, Henry N. Chapman, and James Holton. Femtosecond protein nanocrystallography—data analysis methods. *Opt. Express*, 18(6):5713–5723, Mar 2010.

[45] Richard A. Kirian, Thomas A. White, James M. Holton, Henry N. Chapman, Petra Fromme, Anton Barty, Lukas Lomb, Andrew Aquila, Filipe R. N. C. Maia, Andrew V. Martin, Raimund Fromme, Xiaoyu Wang, Mark S. Hunter, Kevin E. Schmidt, and John C. H. Spence. Structure-factor analysis of femtosecond microdiffraction patterns from protein nanocrystals. *Acta Crystallographica Section A*, 67(2):131–140, Mar 2011.

[46] Petr V. Konarev, Vladimir V. Volkov, Anna V. Sokolova, Michel H. J. Koch, and Dmitri I. Svergun. *PRIMUS*: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography*, 36(5):1277–1282, Oct 2003.

[47] Christopher Kupitz, Shibom Basu, Ingo Grotjohann, Raimund Fromme, Nadia A. Zatsepin, Kimberly N. Rendek, Mark S. Hunter, Robert L. Shoeman, Thomas A. White, Dingjie Wang, Daniel James, Jay-How Yang, Danielle E. Cobb, Brenda Reeder, Raymond G. Sierra, Haiguang Liu, Anton Barty, Andrew L. Aquila, Daniel Deponte, Richard A. Kirian, Sadia Bari, Jesse J. Bergkamp, Kenneth R. Beyerlein, Michael J. Bogan, Carl Caleman, Tzu-Chiao Chao, Chelsie E. Conrad, Katherine M. Davis, Holger Fleckenstein, Lorenzo Galli, Stefan P. Hau-Riege, Stephan Kassemeyer, Hartawan Laksmono, Mengning Liang, Lukas Lomb, Stefano Marchesini, Andrew V. Martin, Marc Messerschmidt, Despina Milathianaki, Karol Nass, Alexandra Ros, Shatabdi Roy-Chowdhury, Kevin Schmidt, Marvin Seibert, Jan Steinbrener, Francesco Stellato, Lifen Yan, Chunhong Yoon, Thomas A. Moore, Ana L. Moore, Yulia Pushkar, Garth J. Williams, Sébastien Boutet, R. Bruce Doak, Uwe Weierstall, Matthias Frank, Henry N. Chapman, John C. H. Spence, and Petra

Fromme. Serial time-resolved crystallography of photosystem ii using a femtosecond x-ray laser. *Nature*, 513:261, 2014.

[48] Jochen Küpper, Stephan Stern, Lotte Holmegaard, Frank Filsinger, Arnaud Rouzée, Artem Rudenko, Per Johnsson, Andrew V. Martin, Marcus Adolph, Andrew Aquila, Sa ša Bajt, Anton Barty, Christoph Bostedt, John Bozek, Carl Caleman, Ryan Coffee, Nicola Coppola, Tjark Delmas, Sascha Epp, Benjamin Erk, Lutz Foucar, Tais Gorkhover, Lars Gumprecht, Andreas Hartmann, Robert Hartmann, Günter Hauser, Peter Holl, Andre Hömke, Nils Kimmel, Faton Krasniqi, Kai-Uwe Kühnel, Jochen Maurer, Marc Messerschmidt, Robert Moshammer, Christian Reich, Benedikt Rudek, Robin Santra, Ilme Schlichting, Carlo Schmidt, Sebastian Schorb, Joachim Schulz, Heike Soltau, John C. H. Spence, Dmitri Starodub, Lothar Strüder, Jan Thøgersen, Marc J. J. Vrakking, Georg Weidenspointner, Thomas A. White, Cornelia Wunderer, Gerard Meijer, Joachim Ullrich, Henrik Stapelfeldt, Daniel Rolles, and Henry N. Chapman. X-ray diffraction from isolated and strongly aligned gas-phase molecules with a free-electron laser. *Phys. Rev. Lett.*, 112:083002, Feb 2014.

[49] Gerrit Langer, Serge X Cohen, Victor S Lamzin, and Anastassis Perrakis. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protocols*, 3:1171–1179, Jun 2008.

[50] Xuanxuan Li, Chun-Ya Chiu, Hsiang-Ju Wang, Stephan Kassemeyer, **Sabine Botha**, Robert L. Shoeman, Robert M. Lawrence, Christopher Kupitz, Richard Kirian, Daniel James, Dingjie Wang, Garrett Nelson, Marc Messerschmidt, Sébastien Boutet, Garth J. Williams, Elisabeth Hartmann, Aliakbar Jafarpour, Lutz M. Foucar, Anton Barty, Henry Chapman, Mengning Liang, Andreas Menzel, Fenglin Wang, Shibom Basu, Raimund Fromme, R. Bruce Doak, Petra Fromme, Uwe Weierstall, Michael H. Huang, John C. H. Spence, Ilme Schlichting, Brenda G. Hogue, and Haiguang Liu. Diffraction data of core-shell nanoparticles from an x-ray free electron laser. *Scientific Data*, 4:170048, 2017.

[51] Ne-Te Duane Loh and Veit Elser. Reconstruction algorithm for single-particle diffraction imaging experiments. *Phys. Rev. E*, 80:026705, Aug 2009.

[52] Lukas Lomb, Thomas R. M. Barends, Stephan Kassemeyer, Andrew Aquila, Sascha W. Epp, Benjamin Erk, Lutz Foucar, Robert Hartmann, Benedikt Rudek, Daniel Rolles, Artem Rudenko, Robert L. Shoeman, Jakob Andreasson, Sasa Bajt, Miriam Barthelmess, Anton Barty, Michael J. Bogan, Christoph Bostedt, John D. Bozek, Carl Caleman, Ryan Coffee, Nicola Coppola, Daniel P. DePonte, R. Bruce Doak, Tomas Ekeberg, Holger Fleckenstein, Petra Fromme, Maike Gebhardt, Heinz Graafsma, Lars Gumprecht, Christina Y. Hampton, Andreas Hartmann, Günter Hauser, Helmut Hirsemann, Peter Holl, James M. Holton, Mark S. Hunter, Wolfgang Kabsch, Nils Kimmel, Richard A. Kirian, Mengning Liang, Filipe R. N. C. Maia, Anton Meinhart, Stefano Marchesini, Andrew V. Martin, Karol Nass, Christian Reich, Joachim Schulz, M. Marvin Seibert, Raymond Sierra, Heike Soltau, John C. H. Spence, Jan Steinbrener, Francesco Stellato, Stephan Stern, Nicusor Timneanu, Xiaoyu Wang, Georg Weidenspointner, Uwe Weierstall, Thomas A. White, Cornelia Wunderer, Henry N. Chapman, Joachim Ullrich, Lothar Strüder, and Ilme Schlichting. Radiation damage in protein serial femtosecond crystallography using an x-ray free-electron laser. *Phys. Rev. B*, 84:214111, Dec 2011.

[53] U. Lorenz, N. M. Kabachnik, E. Weckert, and I. A. Vartanyants. Impact of ultrafast electronic damage in single-particle x-ray imaging experiments. *Phys. Rev. E*, 86:051911, Nov 2012.

[54] Filipe R. N. C. Maia, Tomas Ekeberg, Nicu şor Tîmneanu, David van der Spoel, and Janos Hajdu. Structural variability and the incoherent addition of scattered intensities in single-particle diffraction. *Phys. Rev. E*, 80:031905, Sep 2009.

[55] Filipe R. N. C. Maia, Tomas Ekeberg, David van der Spoel, and Janos Hajdu. *Hawk*: the image reconstruction package for coherent X-ray diffractive imaging. *Journal of Applied Crystallography*, 43(6):1535–1539, Dec 2010.

[56] S. Marchesini. Invited article: A unified evaluation of iterative projection algorithms for phase retrieval. *Review of Scientific Instruments*, 78(1):011301, 2007.

[57] Valerio Mariani, Andrew Morgan, Chun Hong Yoon, Thomas J. Lane, Thomas A. White, Christopher O'Grady, Manuela Kuhn, Steve Aplin, Jason Koglin, Anton Barty, and Henry N. Chapman. *OnDA*: online data analysis and feedback for serial X-ray imaging. *Journal of Applied Crystallography*, 49(3):1073–1080, Jun 2016.

[58] Kazuyuki Morihara and Hiroshige Tsuzuki. Specificity of proteinase k from tritirachium album limber for synthetic peptides. *Agricultural and Biological Chemistry*, 39(7):1489–1492, 1975.

[59] Garib N. Murshudov, Pavol Skubák, Andrey A. Lebedev, Navraj S. Pannu, Roberto A. Steiner, Robert A. Nicholls, Martyn D. Winn, Fei Long, and Alexei A. Vagin. *REFMAC*5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D*, 67(4):355–367, Apr 2011.

[60] Takanori Nakane, Shinya Hanashima, Mamoru Suzuki, Haruka Saiki, Taichi Hayashi, Keisuke Kakinouchi, Shigeru Sugiyama, Satoshi Kawatake, Shigeru Matsuoka, Nobuaki Matsumori, Eriko Nango, Jun Kobayashi, Tatsuro Shimamura, Kanako Kimura, Chihiro Mori, Naoki Kunishima, Michihiro Sugahara, Yoko Takakyu, Shigeyuki Inoue, Tetsuya Masuda, Toshiaki Hosaka, Kensuke Tono, Yasumasa Joti, Takashi Kameshima, Takaki Hatsui, Makina Yabashi, Tsuyoshi Inoue, Osamu Nureki, So Iwata, Michio Murata, and Eiichi Mizohata. Membrane protein structure determination by SAD, SIR, or SIRAS phasing in serial femtosecond crystallography using an iododetergent. *Proceedings of the National Academy of Sciences*, 113(46):13039–13044, 2016.

[61] Karol Nass, Lutz Foucar, Thomas R. M. Barends, Elisabeth Hartmann, **Sabine Botha**, Robert L. Shoeman, R. Bruce Doak, Roberto Alonso-Mori, Andrew Aquila, Saša Bajt, Anton Barty, Richard Bean, Kenneth R. Beyerlein, Maike Bublitz, Nikolaj Drachmann, Jonas Gregersen, H. Olof Jönsson, Wolfgang Kabsch, Stephan Kassemeyer, Jason E. Koglin, Michael Krumrey, Daniel Mattle, Marc Messerschmidt, Poul Nissen, Linda Reinhard, Oleg Sitsel, Dimosthenis Sokaras, Garth J.

Williams, Stefan Hau-Riege, Nicusor Timneanu, Carl Caleman, Henry N. Chapman, Sébastien Boutet, and Ilme Schlichting. Indications of radiation damage in ferredoxin microcrystals using high-intensity X-FEL beams. *Journal of Synchrotron Radiation*, 22(2):225–238, Mar 2015.

[62] Karol Nass, Anton Meinhart, Thomas R. M. Barends, Lutz Foucar, Alexander Gorel, Andrew Aquila, **Sabine Botha**, R. Bruce Doak, Jason Koglin, Mengning Liang, Robert L. Shoeman, Garth Williams, Sebastien Boutet, and Ilme Schlichting. Protein structure determination by single-wavelength anomalous diffraction phasing of X-ray free-electron laser data. *IUCrJ*, 3(3):180–191, May 2016.

[63] Colin Nave and Elspeth F. Garman. Towards an understanding of radiation damage in cryocooled macromolecular crystals. *Journal of Synchrotron Radiation*, 12(3):257–260, May 2005.

[64] Richard Neutze, Remco Wouts, David van der Spoel, Edgar Weckert, and Janos Hajdu. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature*, pages 752–757, 2000.

[65] Santosh Panjikar, Venkataraman Parthasarathy, Victor S. Lamzin, Manfred S. Weiss, and Paul A. Tucker. *Auto-Rickshaw*: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallographica Section D*, 61(4):449–457, Apr 2005.

[66] Santosh Panjikar, Venkataraman Parthasarathy, Victor S. Lamzin, Manfred S. Weiss, and Paul A. Tucker. On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. *Acta Crystallographica Section D*, 65(10):1089–1097, Oct 2009.

[67] Harold R. Powell, Owen Johnson, and Andrew G. W. Leslie. Autoindexing diffraction images with *iMosflm*. *Acta Crystallographica Section D*, 69(7):1195–1203, Jul 2013.

[68] Lars Redecke, Karol Nass, Daniel P. DePonte, Thomas A. White, Dirk Rehders, Anton Barty, Francesco Stellato, Mengning Liang, Thomas R.M. Barends,

Sébastien Boutet, Garth J. Williams, Marc Messerschmidt, M. Marvin Seibert, Andrew Aquila, David Arnlund, Sasa Bajt, Torsten Barth, Michael J. Bogan, Carl Caleman, Tzu-Chiao Chao, R. Bruce Doak, Holger Fleckenstein, Matthias Frank, Raimund Fromme, Lorenzo Galli, Ingo Grotjohann, Mark S. Hunter, Linda C. Johansson, Stephan Kassemeyer, Gergely Katona, Richard A. Kirian, Rudolf Koopmann, Chris Kupitz, Lukas Lomb, Andrew V. Martin, Stefan Mogk, Richard Neutze, Robert L. Shoeman, Jan Steinbrener, Nicusor Timneanu, Dingjie Wang, Uwe Weierstall, Nadia A. Zatsepin, John C. H. Spence, Petra Fromme, Ilme Schlichting, Michael Duszenko, Christian Betzel, and Henry N. Chapman. Natively Inhibited Trypanosoma brucei Cathepsin B Structure Determined by Using an X-ray Laser. *Science*, 339(6116):227–230, 2013.

[69] Philip Roedig, Ramona Duman, Juan Sanchez-Weatherby, Ismo Vartiainen, Anja Burkhardt, Martin Warmer, Christian David, Armin Wagner, and Alke Meents. Room-temperature macromolecular crystallography using a micro-patterned silicon chip with minimal background scattering. *Journal of Applied Crystallography*, 49(3):968–975, 2016.

[70] Christian G. Roessler, Anthony Kuczewski, Richard Stearns, Richard Ellson, Joseph Olechno, Allen M. Orville, Marc Allaire, Alexei S. Soares, and Annie Héroux. Acoustic methods for high-throughput protein crystal mounting at next-generation macromolecular crystallographic beamlines. *Journal of Synchrotron Radiation*, 20(5):805–808, Sep 2013.

[71] Bernhard Rupp. *Biomolecular Crystallograhy*. Garland Science, New York, USA, first edition, 2010.

[72] John J. Rux and Roger M. Burnett. Adenovirus structure. *Human Gene Therapy*, 15:1167–1176, 2004.

[73] W. C. Röntgen. Ueber eine neue art von strahlen. *Ann. Phys.*, 300:1–11, 1898.

[74] Gourinath S, Degenhardt M, Eschenburg S, Moore K, Delucas LJ, Betzel C, and Singh T P. Mercury induced modifications in the stereochemistry of the active site

through Cys-73 in a serine protease–crystal structure of the complex of a partially modified proteinase K with mercury at 1.8 A resolution. *Indian J Biochem Biophys*, 38:298–302, 2001.

[75] Carmen San Martín. Latest insights on adenovirus structure and assembly. *Viruses*, 4(5):847–877, 2012.

[76] Nicholas K. Sauter, Johan Hattne, Ralf W. Grosse-Kunstleve, and Nathaniel Echols. New Python-based methods for data processing. *Acta Crystallographica Section D*, 69(7):1274–1282, Jul 2013.

[77] Michael R. Sawaya, Duilio Cascio, Mari Gingery, Jose Rodriguez, Lukasz Goldschmidt, Jacques-Philippe Colletier, Marc M. Messerschmidt, Sébastien Boutet, Jason E. Koglin, Garth J. Williams, Aaron S. Brewster, Karol Nass, Johan Hattne, **Sabine Botha**, R. Bruce Doak, Robert L. Shoeman, Daniel P. DePonte, Hyun-Woo Park, Brian A. Federici, Nicholas K. Sauter, Ilme Schlichting, and David S. Eisenberg. Protein crystal structure obtained at 2.9 Å resolution from injecting bacterial cells into an x-ray free-electron laser beam. *Proceedings of the National Academy of Sciences*, 111(35):12769–12774, 2014.

[78] Ilme Schlichting. Serial femtosecond crystallography: the first five years. *IUCrJ*, 2(2):246–255, Mar 2015.

[79] Marius Schmidt. Mix and inject: Reaction initiation by diffusion for time-resolved macromolecular crystallography. *Advances in Condensed Matter Physics*, 2013(167276), 2013.

[80] Kenneth S Schmitz. *An Introduction to Dynamic Light Scattering by Macromolecules*. Academic Press Inc., San Diego, California, 1990.

[81] M. Marvin Seibert, Tomas Ekeberg, Filipe R. N. C. Maia, Martin Svenda, Jakob Andreasson, Olof Jönsson, Duško Odić, Bianca Iwan, Andrea Rocker, Daniel Westphal, Max Hantke, Daniel P. DePonte, Anton Barty, Joachim Schulz, Lars Gumprecht, Nicola Coppola, Andrew Aquila, Mengning Liang, Thomas A. White, Andrew Martin, Carl Caleman, Stephan Stern, Chantal Abergel, Virginie Seltzer,

Jean-Michel Claverie, Christoph Bostedt, John D. Bozek, Sébastien Boutet, A. Alan Miahnahri, Marc Messerschmidt, Jacek Krzywinski, Garth Williams, Keith O. Hodgson, Michael J. Bogan, Christina Y. Hampton, Raymond G. Sierra, Dmitri Starodub, Inger Andersson, Saša Bajt, Miriam Barthelmess, John C. H. Spence, Petra Fromme, Uwe Weierstall, Richard Kirian, Mark Hunter, R. Bruce Doak, Stefano Marchesini, Stefan P. Hau-Riege, Matthias Frank, Robert L. Shoeman, Lukas Lomb, Sascha W. Epp, Robert Hartmann, Daniel Rolles, Artem Rudenko, Carlo Schmidt, Lutz Foucar, Nils Kimmel, Peter Holl, Benedikt Rudek, Benjamin Erk, André Hömke, Christian Reich, Daniel Pietschner, Georg Weidenspointner, Lothar Strüder, Günter Hauser, Hubert Gorke, Joachim Ullrich, Ilme Schlichting, Sven Herrmann, Gerhard Schaller, Florian Schopper, Heike Soltau, Kai-Uwe Kühnel, Robert Andritschke, Claus-Dieter Schröter, Faton Krasniqi, Mario Bott, Sebastian Schorb, Daniela Rupp, Marcus Adolph, Tais Gorkhover, Helmut Hirsemann, Guillaume Potdevin, Heinz Graafsma, Björn Nilsson, Henry N. Chapman, and Janos Hajdu. Single mimivirus particles intercepted and imaged with an x-ray laser. *Nature*, 470:78, 2011.

[82] George M. Sheldrick. Macromolecular phasing with SHELXE(Conference Paper). *Zeitschrift fur Kristallographie*, 217:644–650, 2002.

[83] George M. Sheldrick. Experimental phasing with *SHELXC/D/E*: combining chain tracing with density modification. *Acta Crystallographica Section D*, 66(4):479–485, Apr 2010.

[84] Raymond G. Sierra, Hartawan Laksmono, Jan Kern, Rosalie Tran, Johan Hattne, Roberto Alonso-Mori, Benedikt Lassalle-Kaiser, Carina Glöckner, Julia Hellmich, Donald W. Schafer, Nathaniel Echols, Richard J. Gildea, Ralf W. Grosse-Kunstleve, Jonas Sellberg, Trevor A. McQueen, Alan R. Fry, Marc M. Messerschmidt, Alan Miahnahri, M. Marvin Seibert, Christina Y. Hampton, Dmitri Starodub, N. Duane Loh, Dimosthenis Sokaras, Tsu-Chien Weng, Petrus H. Zwart, Pieter Glatzel, Despina Milathianaki, William E. White, Paul D. Adams, Garth J. Williams, Sébastien Boutet, Athina Zouni, Johannes Messinger, Nicholas K. Sauter, Uwe Bergmann, Junko Yano, Vittal K. Yachandra, and Michael J. Bo-

gan. Nanoflow electrospinning serial femtosecond crystallography. *Acta Crystallographica Section D*, 68(11):1584–1587, Nov 2012.

[85] Alexei S Soares, Matthew A Engel, Richard Stearns, Sammy Datwani, Joe Olechno, Richard Ellson, John M Skinner, Marc Allaire, and Allen M Orville. Acoustically mounted microcrystals yield high resolution x-ray structures. *Biochemistry*, 50:4399–4401, 2011.

[86] Claudiu A. Stan, Despina Milathianaki, Hartawan Laksmono, Raymond G. Sierra, Trevor A. McQueen, Marc Messerschmidt, Garth J. Williams, Jason E. Koglin, Thomas J. Lane, Matt J. Hayes, Serge A. H. Guillet, Mengning Liang, Andrew L. Aquila, Philip R. Willmott, Joseph S. Robinson, Karl L. Gumerlock, **Sabine Botha**, Karol Nass, Ilme Schlichting, Robert L. Shoeman, Howard A. Stone, and Sébastien Boutet. Liquid explosions induced by x-ray laser pulses. *Nature Physics*, 12:966, 2016.

[87] David P. Steensma, Marc A. Shampo, and Robert A. Kyle. Max perutz and the structure of hemoglobin. *Mayo Clinic Proceedings*, 90:0025–6196, 2015.

[88] Francesco Stellato, Dominik Oberthür, Mengning Liang, Richard Bean, Cornelius Gati, Oleksandr Yefanov, Anton Barty, Anja Burkhardt, Pontus Fischer, Lorenzo Galli, Richard A. Kirian, Jan Meyer, Saravanan Panneerselvam, Chun Hong Yoon, Fedor Chervinskii, Emily Speller, Thomas A. White, Christian Betzel, Alke Meents, and Henry N. Chapman. Room-temperature macromolecular serial crystallography using synchrotron radiation. *IUCrJ*, 1(4):204–212, Jul 2014.

[89] P. L. Stewart, R. M. Burnett, Walter Doerfler, and Petra Böhm. *Adenovirus Structure by X-ray Crystallography and Electron Microscopy*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1995.

[90] Michihiro Sugahara, Eiichi Mizohata, Eriko Nango, Mamoru Suzuki, Tomoyuki Tanaka, Tetsuya Masuda, Rie Tanaka, Tatsuro Shimamura, Yoshiki Tanaka, Chiyo Suno, Kentaro Ihara, Dongqing Pan, Keisuke Kakinouchi, Shigeru Sugiyama, Michio Murata, Tsuyoshi Inoue, Kensuke Tono, Changyong Song, Jaehyun Park,

Takashi Kameshima, Takaki Hatsui, Yasumasa Joti, Makina Yabashi, and So Iwata. Grease matrix as a versatile carrier of proteins for serial crystallography. *Nature Methods*, 12:61, 2014.

[91] D. I. Svergun. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography*, 25(4):495–503, Aug 1992.

[92] D I Svergun. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J*, 76:2879–2886, Jun 1999.

[93] Jason Tenboer, Shibom Basu, Nadia Zatsepin, Kanupriya Pande, Despina Milathianaki, Matthias Frank, Mark Hunter, Sébastien Boutet, Garth J. Williams, Jason E. Koglin, Dominik Oberthuer, Michael Heymann, Christopher Kupitz, Chelsie Conrad, Jesse Coe, Shatabdi Roy-Chowdhury, Uwe Weierstall, Daniel James, Dingjie Wang, Thomas Grant, Anton Barty, Oleksandr Yefanov, Jennifer Scales, Cornelius Gati, Carolin Seuring, Vukica Srajer, Robert Henning, Peter Schwander, Raimund Fromme, Abbas Ourmazd, Keith Moffat, Jasper J. Van Thor, John C. H. Spence, Petra Fromme, Henry N. Chapman, and Marius Schmidt. Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science*, 346(6214):1242–1246, 2014.

[94] **Sabine Botha**, Karol Nass, Thomas R. M. Barends, Wolfgang Kabsch, Beatrice Latz, Florian Dworkowski, Lutz Foucar, Ezequiel Panepucci, Meitian Wang, Robert L. Shoeman, Ilme Schlichting, and R. Bruce Doak. Room-temperature serial crystallography at synchrotron X-ray sources using slowly flowing free-standing high-viscosity microstreams. *Acta Crystallographica Section D*, 71(2):387–397, Feb 2015.

[95] Isabel Usón and George M Sheldrick. Advances in direct methods for protein crystallography. *Current Opinion in Structural Biology*, 9(5):643 – 648, 1999.

[96] A. Vagin and A. Teplyakov. *MOLREP*: an Automated Program for Molecular Replacement. *Journal of Applied Crystallography*, 30(6):1022–1025, Dec 1997.

[97] Gijs van der Schot, Martin Svenda, Filipe R. N. C. Maia, Max Hantke, Daniel P. DePonte, M. Marvin Seibert, Andrew Aquila, Joachim Schulz, Richard Kirian, Mengning Liang, Francesco Stellato, Bianca Iwan, Jakob Andreasson, Nicusor Timneanu, Daniel Westphal, F. Nunes Almeida, Dusko Odic, Dirk Hasse, Gunilla H. Carlsson, Daniel S. D. Larsson, Anton Barty, Andrew V. Martin, Sebastian Schorb, Christoph Bostedt, John D. Bozek, Daniel Rolles, Artem Rudenko, Sascha Epp, Lutz Foucar, Benedikt Rudek, Robert Hartmann, Nils Kimmel, Peter Holl, Lars Englert, Ne-Te Duane Loh, Henry N. Chapman, Inger Andersson, Janos Hajdu, and Tomas Ekeberg. Imaging single cells in a beam of live cyanobacteria with an X-ray laser. *Nature Communications*, 6:5704, 2015.

[98] Andrew C. Wallace, Roman A. Laskowski, and Janet M. Thornton. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering, Design and Selection*, 8(2):127–134, 1995.

[99] U. Weierstall, J. C. H. Spence, and R. B. Doak. Injector for scattering measurements on fully solvated biospecies. *Review of Scientific Instruments*, 83(3):035108, 2012.

[100] Uwe Weierstall, Daniel James, Chong Wang, Thomas A. White, Dingjie Wang, Wei Liu, John C. H. Spence, R. Bruce Doak, Garrett Nelson, Petra Fromme, Raimund Fromme, Ingo Grotjohann, Christopher Kupitz, Nadia A. Zatsepin, Haiguang Liu, Shibom Basu, Daniel Wacker, Gye Won Han, Vsevolod Katritch, Sébastien Boutet, Marc Messerschmidt, Garth J. Williams, Jason E. Koglin, M. Marvin Seibert, Markus Klinker, Cornelius Gati, Robert L. Shoeman, Anton Barty, Henry N. Chapman, Richard A. Kirian, Kenneth R. Beyerlein, Raymond C. Stevens, Dianfan Li, Syed T. A. Shah, Nicole Howe, Martin Caffrey, and Vadim Cherezov. Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nature Communications*, 5(3309), February 2014.

[101] Thomas A. White, Anton Barty, Francesco Stellato, James M. Holton, Richard A. Kirian, Nadia A. Zatsepin, and Henry N. Chapman. Crystallographic data process-

ing for free-electron laser sources. *Acta Crystallographica Section D*, 69(7):1231–1240, Jul 2013.

[102] Thomas A. White, Richard A. Kirian, Andrew V. Martin, Andrew Aquila, Karol Nass, Anton Barty, and Henry N. Chapman. *CrystFEL*: a software suite for snapshot serial crystallography. *Journal of Applied Crystallography*, 45(2):335–341, Apr 2012.

[103] Martyn D. Winn, Charles C. Ballard, Kevin D. Cowtan, Eleanor J. Dodson, Paul Emsley, Phil R. Evans, Ronan M. Keegan, Eugene B. Krissinel, Andrew G. W. Leslie, Airlie McCoy, Stuart J. McNicholas, Garib N. Murshudov, Navraj S. Pannu, Elizabeth A. Potterton, Harold R. Powell, Randy J. Read, Alexei Vagin, and Keith S. Wilson. Overview of the *CCP*4 suite and current developments. *Acta Crystallographica Section D*, 67(4):235–242, Apr 2011.

[104] Keitaro Yamashita, Naoyuki Kuwabara, Takanori Nakane, Tomohiro Murai, Eiichi Mizohata, Michihiro Sugahara, Dongqing Pan, Tetsuya Masuda, Mamoru Suzuki, Tomomi Sato, Atsushi Kodan, Tomohiro Yamaguchi, Eriko Nango, Tomoyuki Tanaka, Kensuke Tono, Yasumasa Joti, Takashi Kameshima, Takaki Hatsui, Makina Yabashi, Hiroshi Manya, Tamao Endo, Ryuichi Kato, Toshiya Senda, Hiroaki Kato, So Iwata, Hideo Ago, Masaki Yamamoto, Fumiaki Yumoto, and Toru Nakatsu. Experimental phase determination with selenomethionine or mercury-derivatization in serial femtosecond crystallography. *IUCrJ*, 4(5):639–647, Sep 2017.

[105] Keitaro Yamashita, Dongqing Pan, Tomohiko Okuda, Michihiro Sugahara, Atsushi Kodan, Tomohiro Yamaguchi, Tomohiro Murai, Keiko Gomi, Naoki Kajiyama, Eiichi Mizohata, Mamoru Suzuki, Eriko Nango, Kensuke Tono, Yasumasa Joti, Takashi Kameshima, Jaehyun Park, Changyong Song, Takaki Hatsui, Makina Yabashi, So Iwata, Hiroaki Kato, Hideo Ago, Masaki Yamamoto, and Toru Nakatsu. An isomorphous replacement method for efficient de novo phasing for serial femtosecond crystallography. *Scientific Reports*, 5(14017), 2015.

[106] Xu Zhang and Yi Chen. Photo-controlled metal-ion (zn2+ and cd2+) release in aqueous tween-20 micelle solution. *Phys. Chem. Chem. Phys.*, 14:2312–2316, 2012.

Publications

The following publications contain work and results obtained during the course of this thesis:

Sabine Botha, Daniela Baitan, Katharina J. E. Jungnickel, Dominik Oberthür, Christina Schmidt, Stephan Stern, Max O. Wiedorn, Markus Perbandt, Henry N. Chapman and Christian Betzel; *De novo protein structure determination by heavy-atom soaking in lipidic cubic phase and SIRAS phasing using Serial Synchrotron Crystallography*; *IUCrJ*; (submitted).

Sabine Botha, Christian Betzel; *Electron density reveals successful cross-linking of the substrate into the active site of the enzyme after crystallisation* ; *Acta Crystallographica Section F*; (Manuscript in Progress).

Full publication record

Sabine Botha, Karol Nass, Thomas R. M. Barends, Wolfgang Kabsch, Beatrice Latz, Florian Dworkowski, Lutz Foucar, Ezequiel Panepucci, Meitian Wang, Robert L. Shoeman, Ilme Schlichting, and R. Bruce Doak. Room-temperature serial crystallography at synchrotron X-ray sources using slowly flowing free-standing high-viscosity microstreams. *Acta Crystallographica Section D*, 71(2):387–397, Feb 2015.

Co-authoured Papers (in chronological order):

Xuanxuan Li, Chun-Ya Chiu, Hsiang-Ju Wang, Stephan Kassemeyer, **Sabine Botha**, Robert L. Shoeman, Robert M. Lawrence, Christopher Kupitz, Richard Kirian, Daniel

James, Dingjie Wang, Garrett Nelson, Marc Messerschmidt, Sébastien Boutet, Garth J. Williams, Elisabeth Hartmann, Aliakbar Jafarpour, Lutz M. Foucar, Anton Barty, Henry Chapman, Mengning Liang, Andreas Menzel, Fenglin Wang, Shibom Basu, Raimund Fromme, R. Bruce Doak, Petra Fromme, Uwe Weierstall, Michael H. Huang, John C. H. Spence, Ilme Schlichting, Brenda G. Hogue, and Haiguang Liu. Diffraction data of core-shell nanoparticles from an x-ray free electron laser. *Scientific Data*, 4:170048, 2017.

Cornelius Gati, Dominik Oberthuer, Oleksandr Yefanov, Richard D. Bunker, Francesco Stellato, Elaine Chiu, Shin-Mei Yeh, Andrew Aquila, Shibom Basu, Richard Bean, Kenneth R. Beyerlein, **Sabine Botha**, Sébastien Boutet, Daniel P. DePonte, R. Bruce Doak, Raimund Fromme, Lorenzo Galli, Ingo Grotjohann, Daniel R. James, Christopher Kupitz, Lukas Lomb, Marc Messerschmidt, Karol Nass, Kimberly Rendek, Robert L. Shoeman, Dingjie Wang, Uwe Weierstall, Thomas A. White, Garth J. Williams, Nadia A. Zatsepin, Petra Fromme, John C. H. Spence, Kenneth N. Goldie, Johannes A. Jehle, Peter Metcalf, Anton Barty, and Henry N. Chapman. Atomic structure of granulin determined from native nanocrystalline granulovirus using an x-ray free-electron laser. *Proceedings of the National Academy of Sciences*, 114(9):2247–2252, 2017.

Claudiu A. Stan, Despina Milathianaki, Hartawan Laksmono, Raymond G. Sierra, Trevor A. McQueen, Marc Messerschmidt, Garth J. Williams, Jason E. Koglin, Thomas J. Lane, Matt J. Hayes, Serge A. H. Guillet, Mengning Liang, Andrew L. Aquila, Philip R. Willmott, Joseph S. Robinson, Karl L. Gumerlock, **Sabine Botha**, Karol Nass, Ilme Schlichting, Robert L. Shoeman, Howard A. Stone, and Sébastien Boutet. Liquid explosions induced by x-ray laser pulses. *Nature Physics*, 12:966, 2016.

Karol Nass, Anton Meinhart, Thomas R. M. Barends, Lutz Foucar, Alexander Gorel, Andrew Aquila, **Sabine Botha**, R. Bruce Doak, Jason Koglin, Mengning Liang, Robert L. Shoeman, Garth Williams, Sebastien Boutet, and Ilme Schlichting. Protein structure determination by single-wavelength anomalous diffraction phasing of X-ray free-electron laser data. *IUCrJ*, 3(3):180–191, May 2016.

Lorenzo Galli, Sang-Kil Son, Thomas R. M. Barends, Thomas A. White, Anton Barty, **Sabine Botha**, Sébastien Boutet, Carl Caleman, R. Bruce Doak, Max H. Nanao, Karol

Nass, Robert L. Shoeman, Nicusor Timneanu, Robin Santra, Ilme Schlichting, and Henry N. Chapman. Towards phasing using high X-ray intensity. *IUCrJ*, 2(6):627–634, Nov 2015.

Thomas R. M. Barends, Lutz Foucar, Albert Ardevol, Karol Nass, Andrew Aquila, **Sabine Botha**, R. Bruce Doak, Konstantin Falahati, Elisabeth Hartmann, Mario Hilpert, Marcel Heinz, Matthias C. Hoffmann, Jürgen Köfinger, Jason E. Koglin, Gabriela Kovacsova, Mengning Liang, Despina Milathianaki, Henrik T. Lemke, Jochen Reinstein, Christopher M. Roome, Robert L. Shoeman, Garth J. Williams, Irene Burghardt, Gerhard Hummer, Sébastien Boutet, and Ilme Schlichting. Direct observation of ultrafast collective motions in co myoglobin upon ligand dissociation. *Science*, 350(6259):445–450, 2015.

Thomas Barends, Thomas A. White, Anton Barty, Lutz Foucar, Marc Messerschmidt, Roberto Alonso-Mori, **Sabine Botha**, Henry Chapman, R. Bruce Doak, Lorenzo Galli, Cornelius Gati, Matthias Gutmann, Jason Koglin, Anders Markvardsen, Karol Nass, Dominik Oberthur, Robert L. Shoeman, Ilme Schlichting, and Sébastien Boutet. Effects of self-seeding and crystal post-selection on the quality of Monte Carlo-integrated SFX data. *Journal of Synchrotron Radiation*, 22(3):644–652, May 2015.

Yiping Feng, Roberto Alonso-Mori, Thomas. R. M. Barends, Vladimir D. Blank, **Sabine Botha**, Matthieu Chollet, D. S. Damiani, R. B. Doak, J. M. Glownia, Jason M. Koglin, Hilmar T. Lemke, Marc Messerschmidt, Karol Nass, S. Nelson, Ilme Schlichting, Robert L. Shoeman, Yu. V. Shvyd'ko, M. Sikorski, Sanghoon Song, Stanislav Stoupin, S. Terentyev, Garth J. Williams, Diwen Zhu, A. Robert, and Sebastien Boutet. Demonstration of simultaneous experiments using thin crystal multiplexing at the Linac Coherent Light Source. *Journal of Synchrotron Radiation*, 22(3):626–633, May 2015.

Sébastien Boutet, Lutz Foucar, Thomas R. M. Barends, **Sabine Botha**, R. Bruce Doak, Jason E. Koglin, Marc Messerschmidt, Karol Nass, Ilme Schlichting, M. Marvin Seibert, Robert L. Shoeman, and Garth J. Williams. Characterization and use of the spent beam for serial operation of LCLS. *Journal of Synchrotron Radiation*, 22(3):634–643, May 2015.

Karol Nass, Lutz Foucar, Thomas R. M. Barends, Elisabeth Hartmann, **Sabine Botha**, Robert L. Shoeman, R. Bruce Doak, Roberto Alonso-Mori, Andrew Aquila, Saša Bajt, Anton Barty, Richard Bean, Kenneth R. Beyerlein, Maike Bublitz, Nikolaj Drachmann, Jonas Gregersen, H. Olof Jönsson, Wolfgang Kabsch, Stephan Kassemeyer, Jason E. Koglin, Michael Krumrey, Daniel Mattle, Marc Messerschmidt, Poul Nissen, Linda Reinhard, Oleg Sitsel, Dimosthenis Sokaras, Garth J. Williams, Stefan Hau-Riege, Nicusor Timneanu, Carl Caleman, Henry N. Chapman, Sébastien Boutet, and Ilme Schlichting. Indications of radiation damage in ferredoxin microcrystals using high-intensity X-FEL beams. *Journal of Synchrotron Radiation*, 22(2):225–238, Mar 2015.

Michael R. Sawaya, Duilio Cascio, Mari Gingery, Jose Rodriguez, Lukasz Goldschmidt, Jacques-Philippe Colletier, Marc M. Messerschmidt, Sébastien Boutet, Jason E. Koglin, Garth J. Williams, Aaron S. Brewster, Karol Nass, Johan Hattne, **Sabine Botha**, R. Bruce Doak, Robert L. Shoeman, Daniel P. DePonte, Hyun-Woo Park, Brian A. Federici, Nicholas K. Sauter, Ilme Schlichting, and David S. Eisenberg. Protein crystal structure obtained at 2.9 Å resolution from injecting bacterial cells into an x-ray free-electron laser beam. *Proceedings of the National Academy of Sciences*, 111(35):12769–12774, 2014.

Thomas R. M. Barends, Lutz Foucar, **Sabine Botha**, R. Bruce Doak, Robert L. Shoeman, Karol Nass, Jason E. Koglin, Garth J. Williams, Sebastien Boutet, Marc Messerschmidt, and Ilme Schlichting. De novo protein crystal structure determination from X-ray free-electron laser data. *Nature*, 505:244–247, Jan 2014.

Hasan Demirci, Raymond G. Sierra, Hartawan Laksmono, Robert L. Shoeman, **Sabine Botha**, Thomas R. M. Barends, Karol Nass, Ilme Schlichting, R. Bruce Doak, Cornelius Gati, Garth J. Williams, Sébastien Boutet, Marc Messerschmidt, Gerwald Jogl, Albert E. Dahlberg, Steven T. Gregory, and Michael J. Bogan. Serial femtosecond X-ray diffraction of 30S ribosomal subunit microcrystals in liquid suspension at ambient temperature using an X-ray free-electron laser. *Acta Crystallographica Section F*, 69(9):1066–1069, Sep 2013.

Acknowledgements

First of all, I would like to thank my supervisor, Prof. Dr. Christian Betzel, for giving me my project and the opportunity to work in his laboratory during the course of my thesis. Furthermore I am grateful for his assessment of my work during the past three years as well as the final thesis. I would also like to thank PD Dr. Markus Perbandt for the advice and guidance he contributed to my work during my time in Hamburg. Further I would like to express my gratitude towards Prof. Dr. Andrew Torda for his thesis appraisal and to the examiners of my disputation.

I would like to thank the members of my lab, in particular a humongous thank you to Tina, thank you for everything. And Tessie, thank for everything else. Thank you also to Svetlana and Dana, Aileen and Aline, I have no idea how I would have finished this thesis without such special friends supporting me and my work. Robin, thank you also for your friendship, and for all the advice and data discussions. And Karlos, thank you for showing up when and where you did ;)

Thank you Petra Belda all of your hard work. I'm sorry for the unnecessary grievances I managed to cause you during the past three years. We do know that this lab wouldn't run without you.

A very special thank you to Pof. Dr. Dr. Henry Chapman, not only for always putting so much of your equipment at my disposal, but also for your kindness and generosity in enabling my chance here in Hamburg. Dominik Oberthür, thank you for all your advice and assistance along the way. I am also truly in your dept after the generous tape-drive beamtime gift last November! Thank you also to Stephan Stern, Max Wiedorn, and all the others at CFEL that assisted some sort of experiment somewhere along the road.

Further I would like to thank all of the P11 staff for providing such a fabulous beamline where the majority of my data was collected. And thank you Anja Burkhardt and Olga Lorbeer for all your support during all my non-standard setup experiments and doing your best to help me realize my beamtime plans. Sorry for the midnight calls.

My gratitude also to Steve Aplin and the DESY IT, for helping me with the huge amounts of data and giving me access to the Maxwell cluster so that I could finish processing my data in this century.

A special thank you to PD Dr. Cornelia Heinze and Dr. Thomas Speiseder for providing me with the virus samples. Lena Carstensen, thank you for being so patient with me in my attempt to do organic chemistry in your lab. Thank you Alex for assisting me with your Nanosight set-up, Thomas and Carolin thank you for helping me take TEM images of my viruses. I also cannot conclude without thanking my former colleague Dr. Thomas Barends for teaching me most of what I know about protein crystallography.

Beate, Jenny, Maren, Fernando, Kike, Anja. Danke, ihr wisst wofür.

Finally, I would like to thank Hannah and Aida for always encouraging me to carry on and persevering through all of my university studies (and before). Thank you Marc, I couldn't have done it without you. Both privately and professionally. Mina, Nyala, Bessie and Fluffy, thank you for always being the highlight of my days. And last but certainly not least, I would like to express a special thank you to my parents and sisters for always supporting me and believing in me. *Dankie vir alles, altyd.*

Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hamburg, May 2, 2018

List of Abbreviations

ALA	Alanine
ARG	Arginine
ASN	Asparagine
ASP	Aspartic acid
BA	Boric Acid
bR	Bacteriorhodopsin
CYS	Cysteine
DESY	Deutsches Elektronen Synchrotron
DLS	Dynamic Light Scattering
DMSO	Dimethyl Sulfoxide
CYS	Cysteine
EMBL	European Molecular Biology Laboratory
ESI-MS	Electrospray Ionization Mass Spectrometry
FEL	Free Electron Laser
FLASH	Free electron LASer Hamburg
FOM	Figure of Merit
GDVN	Gas Dynamic Virtual Nozzle
GHS	Globally Harmonised System
GLN	Glutamine
GLU	Glutamic acid
GLY	Glycine
h	Hours
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic Acid
HIS	Histidine
HPLC	High Performance Liquid Chromatography
ID	Inner Diameter
ILE	Isoleucine
Insulin	Insulin, Chain B oxidised
K	Kelvin
LCLS	Linac Coherent Light Source
LCP	Lipidic Cubic Phase
LEU	Leucine
LYS	Lysine
MET	Methionine
min	Minutes

MIRAS	Multiple Isomorphous Replacement using Anomalous Scattering
MW	Molecular Weight
MWCO	Molecular Weight Cut-off
OD	Outer Diameter
PDB	Protein Data Bank
PETRA	Positron Elektron Tandem Ring Anlage
PHE	Phenylalanine
PMS	Phenylmethane Sulfonyl
PMSF	Phenylmethane Sulfonyl Fluoride
pnCCD	Charge Coupled Device with pn-junctions as registers
PRO	Proline
s	Seconds
SAD	Single-wavelength Anomalous Diffraction
SER	Serine
SIR	Single Isomorphous Replacement
SIRAS	Single Isomorphous Replacement with Anomalous Scattering
SAXS	Small Angle X-ray Scattering
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SFX	Serial Femtosecond Crystallography
SMX	Serial Millisecond Crystallography
SX	Serial Crystallography
TCEP	Tris(2-carboxyethyl)phosphine
TEM	Transmission Electron Microscopy
THR	Threonine
TLC	Thin-layer Chromatography
TRIS	Tris-(hydroxymethyl)-aminomethane
TRP	Tryptophan
TYR	Tyrosine
UKE	Universitätsklinikum Hamburg-Eppendorff
UV	Ultraviolet
VAL	Valine
XFEL	X-ray Free Electron Laser

List of Figures

List of Tables