Analyzing Crystal Growth Phenomena and Mechanisms for the Production and Optimization of Protein Crystals for Serial Crystallography

Dissertation

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

vorgelegt von

Daniela – Paraschiva Baitan

August, 2018

Die vorliegende Arbeit wurde im Zeitraum von April 2014 bis August 2018 in der Arbeitsgruppe von Prof. Christian 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. Dr. Christian Betzel
2. Gutachter: Prof. Dr. Henning Tidow

Datum der Disputation: 21.09.2018

Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, die vorliegende Dissertation selbst verfasst und keine anderen als die angegebenen Hilfsmittel benutzt zu haben.

August 22nd, Daniela – Paraschiva Baitan

Table of Contents

Table of Contents

Abstract

Protein crystallization represents nowadays one of the main limiting factors in X-ray crystallography for structure retrieval. The technical advancements of high intensity X-ray radiation sources, such as third generation synchrotrons and free-electron lasers (FELs) extended the possibilities for achieving structural information of biological targets by using serial crystallographic methods and protein crystals that are in the nano- and micrometer size range. The high potential of serial crystallography and successful results obtained so far in this field have posed a great interest and demand on sample preparation for generating sub-microscopic crystals. As for the preparation of protein crystals for conventional crystallography, the samples needed for serial femtosecond and/or millisecond crystallography (SFX and SMX) require in-depth knowledge of the mechanisms of nucleation and crystal growth in order to optimize the crystallization methods for obtaining nano- and microcrystals.

The present work describes a recent developed method of automated crystallization (XtalController900), coupled with dynamic light scattering (DLS) for preparation and scoring of protein microcrystals prior to data collection. The analysis of three soluble proteins (thaumatin from *thaumatococcus daniellii*, Plasmodium falciparum glutathione S-transferase - PfGST and SP – target sample) with DLS during the crystallization trials, showed a clear difference between successful crystallization and sample precipitation by evaluating the assembly mechanism of protein molecules by means of the hydrodynamic radius distribution (R_h) map interpretation. It was found that the ability to measure the change in particle radii of a crystallization droplet can help to understand and to modify the necessary crystallization conditions in order to obtain protein crystals and avoid protein precipitation.

A detailed study of the nucleation and crystal growth phases of three proteins was conducted with the use of DLS and showed different particle assembly mechanisms during the formation of crystals, which were later visualized by electron microscopy. The microscopic images suggest that the R_h distribution attributed to nucleation

might show a two-step mechanism of nucleation involving the transition of clusters (approximately 200-300 nm) with high protein concentration into crystal nuclei (approx. 400 – 600 nm). Furthermore, the crystal growth stage indicated by a R_h distribution between approx. 800 and 2000 nm was identified as crystalline material. The *in situ* DLS maps and the parameter plots collected during automated protein crystallization were used to plot experimental phase diagrams, explaining the main differences of the final crystallization experiment results.

The PfGST crystals and thaumatin microcrystals produced with the XtalController900 were quantified by classic X-ray crystallography and powder diffraction, respectively. The information given by the experimental phase diagrams was utilized to optimize the production of larger amounts of protein microcrystals to be used for serial crystallography applications.

Further, serial millisecond crystallography for experimental phasing was successfully carried out for the thaumatin microcrystals using the tape-drive data collection approach. The serially collected data could be used for native Sulphur phasing and when supplementing the correct substructure, the calculated electron density sufficed for automatic building of the entire protein structure.

Zusammenfassung

Analyse von Kristallwachstums Phänomene, und der Mechanismes zur Produktion und Optimierung von Proteinkristallen für serielle Kristallografie

Die Kristallisation von Proteinen ist noch immer einer der limitierenden Faktoren in Röntgenkristallographie-Untersuchungen zur Strukturaufklärung von biologischen Makromoleküle. Der technische Fortschritt hochintensiver Röntgenstrahlungsquellen wie Synchrotrons der dritten Generation und Freie-Elektronen-Laser (FELs) bieten neue Möglichkeiten strukturelle Informationen biologischer Makromoleküle zu erhalten, dies mittels seriellen kristallographischen Methoden an Kristallen im Nano- und Mikrometerbereich. Das hohe Potential der seriellen Kristallographie und die bemerkenswerten Ergebnisse, die bisher auf diesem Gebiet erzielt wurden, haben ein großes Interesse und eine große Nachfrage nach der Probenvorbereitung zur Produktion von submikroskopischen Kristallen geschaffen. Im Vergleich zur Herstellung von Proteinkristallen für die klassische Kristallographie, erfordern die für die serielle Femtosekunden- und/oder Millisekunden-Kristallographie (SFX und SMX) benötigten Proben ein präzises Verständnis der Mechanismen der Nukleation und des Kristallwachstums, um die Kristallisationsmethoden für Mikro- und Nanokristallen zu optimieren.

Die vorliegende Arbeit beschreibt ein kürzlich entwickeltes Verfahren zur automatisierten Kristallisation (XtalController900), gekoppelt mit dynamischer Lichtstreuung (DLS) zur Produktion und Bewertung von Proteinmikrokristallen vor der Datensammlung. Die Analyse der Kristallisationsbedingungen von drei löslichen Proteinen (Thaumatin aus *thaumatococcus daniellii*, Plasmodium *falciparum* Glutathione S-Transferase (PfGST) un SP – zielprobe) mit DLS zeigte eine klare Unterscheidung zwischen erfolgreicher Kristallisation und Probenfällung durch Auswertung des Anordnungsmechanismus von Proteinmolekülen mittels hydrodynamischer Radiusverteilung (R_h). Die Ergebnisse weisen darauf hin, dass die Bewertung der Änderungen der Teilchenradien eines Kristallisationstropfens dazu

beitrag en kann, die notwendigen Bedingungen eines Kristallisationsansatzes zu untersuchen und dahingehend zu modifizieren Proteinkristalle zu erhalten und Proteinfällung zu vermeiden.

Eine detaillierte Untersuchung der Nukleations- und Kristallwachstumsphasen der drei Proteine wurde mittels DLS durchgeführt und zeigte verschiedene Mechanismen von Phasenübergängen während der Bildung von Kristallen, die durch Elektronenmikroskopie analysiert warden konnten. Die mikroskopischen Aufnahmen legen nahe, dass die der Nukleation zugeschriebene R_h-Verteilung einen zweistufigen Nukleationsmechanismus zeigen könnte, bei dem ein Übergang von Clustern mit hoher Proteinkonzentration (ca. 200-300 nm) in stabile Kristallkeime (ca. 400-600 nm) stattfindet. Darüber hinaus wurde Probenmaterial aus der Kristallwachstumsphase, die sich durch eine R_h-Verteilung zwischen ca. 800 und 3000 nm auszeichnet, erfolgreich als Proteinkristalle identifiziert. Komplementäre *In-situ*-DLS-Messungen die während der automatisierten Proteinkristallisation gesammelt wurden, wurden verwendet, um experimentelle Phasendiagramme aufzuzeichnen, die die Hauptunterschiede im endgültigen Kristallisationsergebnis erklären konnten.

Die mit dem XtalController900 hergestellten PfGST-Kristalle und Thaumatin-Mikrokristalle wurden durch klassische Röntgenkristallographie und Pulverdiffraktometrie quantifiziert. Die Informationen, die durch die experimentellen Phasendiagramme geliefert wurden, lieferten die Grundlage für die Produktion größerer Mengen von Proteinmikrokristallen für weitere serielle Kristallographieanwendungen.

Die serielle Millisekunden-Kristallographie zur experimentellen Phasenbestimmung von Thaumatin wurde erfolgreich an Thaumatin-Mikrokristallen mit Synchrotronstrahlung durchgeführt. Die seriell gesammelten Daten wurden für die native Schwefel-Phasierung verwendet, und mit Angabe der Substruktur konnte eine Elektronendichte generiert werden, die erfolgreich zur automatischen Strukturlösung des Proteins führte.

Publications associated with this thesis

Baitan, Daniela, Robin Schubert, Arne Meyer, Karsten Dierks, Markus Perbandt, and Christian Betzel. 2018. "Growing Protein Crystals with Distinct Dimensions Using Automated Crystallization Coupled with In Situ Dynamic Light Scattering." *JoVE*, no. 138: e57070. https://doi.org/doi:10.3791/57070.

Botha, Sabine, Daniela Baitan, Katharina E J Jungnickel, Dominik Oberthür, C Schmidt, Stephan Stern, Max Wiedorn, Markus Perbandt, Henry N Chapman, and Christian Betzel. 2018. "De Novo Protein Structure Determination by Heavy-Atom Soaking in Lipidic Cubic Phase and SIRAS Phasing Using Serial Synchrotron Crystallography." *IUCrJ* in press: 1–7. https://doi.org/10.1107/S2052252518009223.

Schubert, Robin, Arne Meyer, Daniela Baitan, Karsten Dierks, Markus Perbandt, and Christian Betzel. 2017. "Real-Time Observation of Protein Dense Liquid Cluster Evolution during Nucleation in Protein Crystallization." *Crystal Growth and Design* 17 (3): 954–58. https://doi.org/10.1021/acs.cgd.6b01826.

Chapter I

Outline of research, background and introduction

I.1. Outline and aim of research

The production of biomolecular nano- and microcrystals has become of substantial relevance for protein structure determination, due to the revolutionary development of serial crystallography methods. Serial crystallography is based on collecting diffraction data on nano- and microcrystals using ultrashort X-ray pulses of high brilliance provided by free electron lasers (FELs) or highly brilliant synchrotron radiation (SR). Because of the small size of the crystals and due to the transferred electromagnetic energy from an X-ray pulse causing a full ionization of all atoms inside a crystal, the data collection is achieved by exposing crystals one by one in random orientations to the beam and collecting one diffraction image per crystal. The method is based on collecting diffraction patterns from several thousands of crystals, and the diffraction information is collected prior to crystal destruction (Neutze et al. 2000). The striking results obtained (Boutet et al. 2012; Kupitz, Basu, et al. 2014; Neutze and Moffat 2013) have posed a strong demand on crystalline sample preparation where the well-known conventional methods of protein crystallization are not feasible anymore. Since FELs are yet not widely spread and the available time for data collection is limited, serial crystallography at synchrotron radiation has been under constant development where data collection on microcrystals can be carried out in a similar fashion as at FELs (Levantino et al. 2015; Stellato et al. 2014).

The main goal of this thesis research work was to identify and develop new methods and routes for the systematic production of nano- and microcrystals, as well as subjecting the obtained crystals to structure determination using serial crystallography at a synchrotron radiation source. The XtalController technology was first reported in 2008 for the optimization of a protein crystallization experiment in real time by means of *in situ* dynamic light scattering (Meyer et al. 2012). Within the work of this thesis, one important goal was the development of a more advanced

XtalController device that can accommodate a crystallization experiment at varying environmental conditions as well as fully controlling its crystallization path in real time in order to achieve the necessary conditions for a precise crystallization outcome. Due to the size range of particles that can be investigated with dynamic light scattering (DLS), the coupling of the DLS optics with the automated crystallization device can provide valuable information by means of the hydrodynamic radii distribution of particle at different crystallization stages. Hence, the use of the ultra-sensitive balance onto which the crystallization droplet is placed, provides online feedback about the protein and precipitant concentrations within the droplet throughout the entire experiment. This feedback is very important, as experimental phase diagrams can be derived in order to understand the process of the sample undergoing crystallization. In chapter II, the designing and testing of the newly developed XtalController900 device is presented, as well as its potential for monitoring the development of the radius distribution of the protein molecules during protein crystallization and protein aggregation. The results discussed are from a well-characterized protein (thaumatin from *Thaumatococcus daniellii*) and two target proteins: gluthatione s-transferase from the malarial parasite *Plasmodium falciparum* (PfGST) and a target protein, for which the acronym SP will be used in this thesis due to the ongoing research studies in the group of prof. Rolf Hilgenfeld at the University of Lübeck.

A major goal of the current work is to investigate the process of nucleation during protein crystallization with the use of the XtalController900. In chapter III, different, selected proteins were utilized for analysing the distinct mechanisms of nucleation and crystal growth phenomena by *in situ* DLS. The results obtained are further analysed by cryo-electron microscopy in order to investigate the theory of a two-step nucleation of protein molecules (Vekilov 2010). The aim is to evaluate the hydrodynamic radii distribution within the protein droplets at different crystallization stages in the light of liquid-dense clusters as nucleation precursors. The influence of droplet stirring during crystallization, as well as crystallization studies altering the chemical potential, are shown for thaumatin and PfGST, followed by experimental phase diagrams. The quality of the obtained crystal suspensions is

cross-checked and analysed by X-ray powder diffraction and electron microscopy, to pinpoint and understand the specific dimensions of the particle radii throughout the crystallization experiments.

Using the information provided by the experimental phase diagrams, further optimization of the samples for serial crystallography is presented in chapter IV. The goal was to produce a large amount of stable crystals that can be used for serial studies at a synchrotron radiation source. Furthermore, a special focus is addressed to evaluate the quality of the crystals. In chapter V, conventional X-ray data collection, as well as serial data collection at a synchrotron, are presented for the previously produced PfGST crystals and the thaumatin microcrystals respectively. The goal was to cross check the quality of the crystals produced with the XtalController900, as well as to evaluate the possibility of serial data collection on microcrystals at a synchrotron facility using a new data collection approach. The tape drive data collection method – developed for mix and diffuse experiments (Beyerlein et al. 2017) was used for serial data collection of the thaumatin microcrystals. The diffraction images obtained were then used for structure determination using native sulphur phasing.

I.2. Protein crystallization, nucleation and crystal growth mechanisms

Crystallization of biological molecules such as enzymes, viruses, proteins and nucleic acids has revolutionized the methods employed in medicine for the treatment and prevention of diseases (Sumner, 1948; Perutz, 1969; Kam, Shore, and Feher, 1978; Michel, 1983). Macromolecular crystals are defined as an ordered periodic assembly of macromolecules, held together by weak intermolecular interactions (Rupp, 2010). Protein crystals are often small (from nano- to micro- sizes), and fragile to mechanical stress, with a high sensitivity to environmental changes. The main reason for arranging protein molecules into an ordered array (protein crystals), is to gain structural information by applying X-ray crystallography, that can further elucidate

biological functions and mechanisms trivial for advances in the medicinal and pharmaceutical fields (Caldwell et al. 2001; Ricci and Brems 2004; Fowler et al. 2005).

In order to obtain structural information from molecules, several techniques have been developed: Nuclear Magnetic Resonance – NMR (Kremer and Kalbitzer 2001), Atomic Force Microscopy – AFM (Rajendran, Endo, and Sugiyama 2012), X-ray scattering methods (Svergun and Koch 2003; Sichun 2014) as well as electron and neutron diffraction (Henderson et al. 1990; Hughes et al. 2007; Blakeley et al. 2008). During the past years, the rapid advances in X-ray crystallography and molecular biology have posed a great interest and attention to protein crystallization, being the main method towards obtaining the three dimensional structure of macromolecules (William W. Wilson and Delucas 2014). Crystallizing biological molecules remains a crucial step in sample preparation since there are no means of directly predicting the conditions for an optimal crystal formation environment. In the present, protein crystallization is considered to be the bottleneck in the course of structure determination, being the central problem in biological crystallography.

For a protein in solution to crystallize, at first it must move from its stable solution state towards supersaturation. This is usually achieved by changing the protein solution conditions with the help of adding polar chemicals (precipitant agents) or by altering the physical parameters of the protein solution, such as temperature and pH (McPherson 2009). Once supersaturation is achieved, the protein solution is far from equilibrium in a metastable environment. If the new conditions are favorable for crystallization, the macromolecules will associate with one another giving rise to nucleation by forming new entities consisting of internal order showing a regular lattice pattern in their structure. These entities are known as nucleation sites that will promote the growth of crystals while the solution equilibrates in the new phase. However, although supersaturation towards crystal formation can be achieved by many means (variation of protein and precipitant concentration, time and method of precipitant addition, removal of water from the crystallization solution, exchange of solvent by dialysis, time for precipitant addition, etc.), the crystallization is not always achieved and for many trials the outcomes are protein aggregates and precipitates, or the sample suspension remains homogeneous without any optical

changes. An important tool in visualizing the change in protein solubility are the phase diagrams, showing the state of the protein molecules as a function of all relevant variables in the system. The most common forms are the two-dimensional phase diagrams displaying the protein concentration as a function of another parameter (Saridakis et al. 1994). More detailed three-dimensional diagrams as well as more complex representations, are sometimes used for achieving a higher understanding about the main drivers in protein crystallization (C. Sauter et al. 1999; Ewing, Forsythe, and Pusey 1994).

A convenient way to represent the different phases involved in protein crystallization is the 2D solubility diagram, where the protein concentration is plotted on the vertical axis and the precipitant concentration is displayed on the horizontal axis as shown in Figure I-1. Usually, the diagram is separated into three distinct regions showing undersaturated and saturated protein in solution, a region identified as nucleation and a third region for mapping aggregates and protein precipitation. However, in the first part of the diagram the protein solution can be visualized in a more detailed way: at low precipitant concentrations, the protein solution is usually attributed to an undersaturated region, and as the precipitant concentration rises in solution, the protein sample enters the supersaturation region. Once the protein solution advances towards high supersaturation, the equilibrium is lost, and a metastable phase is created, where the kinetics are hampered and do not allow the system to relax into equilibrium. For protein molecules to self-assemble into crystals, a net drop in free energy must take place during the process of crystallization, resulting in a negative value for the free energy (ΔG^*) of crystallization. The simplest description for the net drop in free energy is given by the Gibb's free energy shown in equation I-1, where ΔH is the enthalpy of the system, T is the absolute temperature, $\Delta S_{protein}$ is the entropy of the protein solution and $\Delta S_{solvent}$ the entropy of the solvent.

$$\Delta G^* = \Delta H - T(\Delta S_{protein} + \Delta S_{solvent})$$

Equation I-1
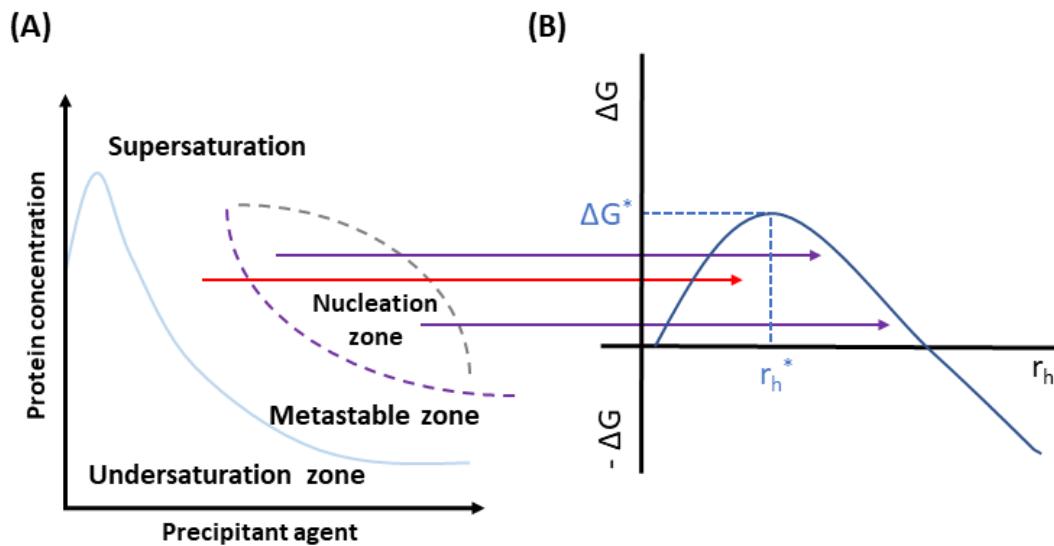
Figure I-1: Solubility diagram and energetics of nucleation
(A) Oswald-Miers diagram divided into different zones: undersaturation, supersaturation, metastability, and nucleation; (B) the energetics of the system expressing the kinetics of crystallization allowing the prediction of nuclei formation and growth into a protein crystal. *Image modified from Wlodawer, Dauter and Jaskolski, 2017.*

For a favorable crystallization outcome an activation barrier towards equilibrium must be overcome for a nucleus (the smallest critical ordered entity r_h^* from which a crystal will further grow) to form (Figure I-1 B). In equation I-1, the energy released by crystallization represents the difference between the changes in enthalpic energy (ΔH) and entropic energy (ΔS) (Gibbs 1978). As the solution undergoes the metastable region, the entropy of the system is represented by two parts: the entropy of protein molecules (ΔS protein) and the precipitant solution molecules (ΔS solvent). Once the activation barrier is overcome by the spontaneous formation of nucleation sites, the process becomes fully driven by the system`s thermodynamics. At this stage, the nuclei are equally likely to fall apart or to proceed with crystal formation and can be localized in the solubility diagrams in the nucleation zone. The energetic requirement (negative ΔS) attributed to crystal formation represents an increase in order, where the entropy inside the nuclei decreases with the order of the crystal lattice. Usually, the conditions necessary to

exploit and control these energetic differences are optimization of temperature, pH and precipitant composition. The description of protein crystallization as discussed in this part has its fundamentals in the classical theory of nucleation stipulating that protein crystals are formed from nuclei arising in supersaturated solutions. The nucleation is therefore governed by the balance between the bulk and surface energy of a one-step phase transition from protein molecules in supersaturation to crystal nuclei and crystal growth driven by kinetics.

Studies on protein crystallization have revealed self-assembly mechanisms for protein crystal formation that cannot be rationalized by the theory of nucleation outlined above (Vekilov 2010; Pan, Vekilov, and Lubchenko 2010). The findings describe a new theory of crystal nucleation where the initiation is based on a two-step mechanism of molecules assembling. At first, the supersaturated molecules arrange in the form of liquid dense clusters where the protein concentration is higher than in solution. In a second step, an internal order of the clusters is achieved, resulting in crystal nuclei. These results have established new fundamentals of the crystallization theory for a great number of proteins (Kashchiev, Vekilov, and Kolomeisky 2005; Vekilov 2010; Gliko et al. 2005; Pan et al. 2007). The most recent results suggest that for some proteins, nucleation is driven by oriented attachments between subcritical cluster polymorphs showing an already existing degree of crystallinity. With the use of electron diffraction, it could be shown how the polymorph selection is based on specific building blocks at an early stage of structure formation for each space group (Van Driessche et al. 2018).

The two-step mechanism involving the activity of liquid dense clusters is believed to be vital in the process of protein nucleation and crystallization (Maes et al. 2015; Kashchiev, Vekilov, and Kolomeisky 2005). Furthermore, the different mechanisms of clusters self-assembly into crystal nuclei show how protein crystallization cannot be viewed from a single perspective anymore. In chapter III, the understanding and new findings in this particular field are discussed, along with new experimental evidence which is crucially needed as a part of the rationalization of protein crystallization, particularly when targets are in focus of structure determination.

I.3. Dynamic light scattering for macromolecular crystallization

A successful crystallization of macromolecules requires at first a suitable sample environment with several conditions, such as: sufficient protein solubility allowing a reasonable high concentration in solution, a high purity of sample preparation along with a single protein species in solution. Such information can usually be assessed by native polyacrylamide gel electrophoresis (PAGE), size-exclusion chromatography (SEC) or analytical ultracentrifugation (AUC). Among these methods, dynamic light scattering (DLS) is a much faster and sensitive technique, providing valuable information about protein particles in solution, such as determining the size of particles, the sample homogeneity as well as the oligomeric states of a protein. The DLS method is particularly well suited for studying protein particles undergoing crystallization because it can detect and measure the homogeneity of a sample, giving feedback about the different particle size distribution as well as the average particle size in solution. The information achieved from performing DLS measurements in a protein solution have proved to be highly important in protein crystallization (Zulauf and D'Arcy 1992; Ferré-D'Amaré and Burley 1994). It has been well established that proteins with a monodisperse hydrodynamic radius distribution in solution have a higher probability of crystallizing, whereas samples with significant polydispersity result in no crystals or crystals with poor diffraction quality (Niesen et al. 2008; Ii et al. 2010).

A dynamic light scattering experiment measures the intensity fluctuations of light scattered by molecules in solution undergoing Brownian motion. When a sample is illuminated by a monochromatic light source (typically a laser), the light scattered by the molecules has two qualities – a certain intensity and it shows fluctuations. The analysis of the averaged intensity is called static light scattering (SLS) and can be used along with the concentration and angle-dependence, for calculating the molecular weight of particles in solution and the second virial coefficient B_{22} (indicator showing weather the probed solution has repulsive or attractive interactions between molecules (Baldwin, Crumley, and Carter Jr 1986; Santos and Castanho 1996). The temporal fluctuations in the scattered light are thus the domain of dynamic light

scattering (DLS), whereas the term dynamic describes the fluctuations of the detected signal, which in turn yield information about the diffusion coefficient of particles and indirectly about their size. While SLS employs measurements of the averaged intensity over approximately one second or more, a DLS measurement uses a long time series of mean intensity values of the scattered light where the averaging is done over short time intervals, usually 100 ns. Averaging over such short time frames ensures that fluctuations can be precisely tracked.

In a typical DLS measurement the monochromatic beam (laser diode) is guided into the sample holder where particles in solution are following Brownian motion (Figure I-2). As the particles are moving freely in the solution when passing through the incident beam, light is scattered in all directions by the particles. The scattered light forms a permanently changing interference pattern, due to the movements of molecules. The scattered light is collected at an angle θ from the incident beam by a detector, usually a photomultiplier tube (PMT). This detector registers intensity fluctuations caused by the ever-changing interference pattern. In order to extract further information from these fluctuations, an autocorrelator device calculates the time-dependent autocorrelation function (ACF) from the detector output signal (equation I-2); in this equation, $g(\tau)$ represents the autocorrelation function of the light intensity *I* (comparing the intensity of the signal at time *t* to the intensity at a very small time later *t+ τ)*. When the solution contains only one particle size the ACF can be described as an exponentially decaying function, as seen in equation I-3. Here, *D* is the diffusion coefficient, *q* the scattering vector and *a* is a factor between 0 and 1, depending on coherence properties of the light source. Particles with large physical dimension (radius) diffuse slowly in solution whereas smaller particles diffuse more rapidly, meaning that the intensity fluctuations seen through time are slower for larger particles. Therefore, the correlation of the intensity signal will take a longer time to decay for larger particles, whereas for small particles the correlation of the signal will decrease more quickly. The rapidness or the average frequency of the fluctuations thus becomes a measure of how fast the particles move, meaning a measure of the translational diffusion coefficient *D* (Harding and Jumel 1998). The program Contin (Provenche 1982) is used to convert the measured ACF into a decay

time constant distribution and this distribution is converted into a radius distribution according to equation I-3.
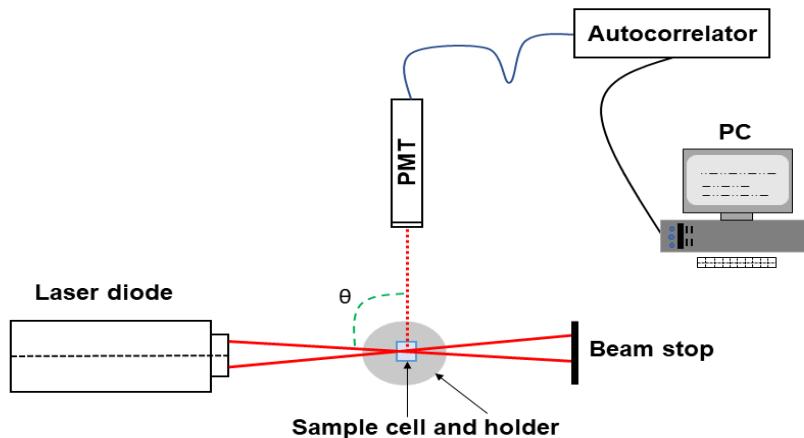
Figure I-2: Scheme of a DLS setup for measurements
The laser beam is sent into the sample cuvette and the scattered light is detected by a photomultiplier tube (PMT) at a certain angle in this case defined by θ. The signal is further transferred to an autocorrelator device and the data is then evaluated with the use of software.

According to the Stokes-Einstein equation, the diffusion coefficient D for a small sphere diffusing through a viscous media can be calculated as shown in equation I-4, where T is the absolute temperature, k_B is the universal Boltzmann constant, η the viscosity and R_h the hydrodynamic radius of the sphere. Assuming that a protein is roughly spherical, the apparent hydrodynamic radius of the protein can be obtained from the Stokes-Einstein calculations, as shown in equation I-5. Finally, the equation can be written using different terms (equation I-6) based on the number of molecules contained in one mole of substance (the Avogadro number – N_A), in order to obtain an estimation of the particle's molecular weight - M_w (Bergfors 2009; Proteau, Shi and Cygler 2010).

$$g(\tau) = \frac{< I(t)I(t+\tau) >_t}{I(t)^2}$$

Equation I-2

$$g(\tau) = 1 + ae^{-2Dq^2r}$$

Equation I-3

$$D = \frac{k_bT}{6\pi\eta R_h}$$

Equation I-4

$$R_h = \frac{k_bT}{6\pi\eta D}$$

Equation I-5

$$R_h M_w = N_a\rho\frac{4}{3}\pi R_h^3$$

Equation I-6

Dynamic light scattering has become a routine method for characterizing a protein sample in solution before crystallization. The utility of DLS has proven to be very useful for the optimization of protein buffer composition, enhancing the chances for successful crystallization (Goon et al. 2003; Shi et al. 2009). For membrane proteins, the critical micelle concentration (CMC) as well as protein-detergent complexes could be investigated by means of DLS to obtain stable solutions for crystallization trials (Meyer et al. 2015). The method is nowadays also widely used for rapid screening of protein crystallization conditions (Baldwin, Crumley, and Carter Jr 1986; Ferré-D'Amaré and Burley 1994).

New developments have enabled the monitoring of counter-diffusion crystallization in glass capillaries by *in situ* DLS, as well as investigating protein crystallization in cuvettes or directly in commercial protein crystallization plates (Mikol, Hirsch, and Giege 1989; W. William Wilson 1990; Dierks et al. 2008; Oberthür 2012). The crystallography community found the method to also be appropriate for

investigating the stages of protein crystallization, such as nucleation and crystal growth, gaining information from DLS measurements in cuvettes (Juarez-Martines et al. 2001; W William Wilson 2003; Skouri et al. 1995).

In recent years, new methods using *in situ* DLS coupled with crystallization in smaller volumes, have shown to provide essential information to improve a protein crystallization experiment throughout the different stages of crystallization. This could be achieved by online monitoring of the hydrodynamic radii over time, which can indicate clear predictions for obtaining different sizes of protein crystals (Meyer et al. 2012; Baitan et al. 2018) . Furthermore, by monitoring a crystallization experiment with *in situ* DLS, a new radii population was found and attributed to early processes of nucleation (Schubert et al. 2017). The successful and promising results obtained throughout the years with this technique open the possibility for new studies using *in situ* DLS for tracking and gaining new insights into phase transitions and the mechanisms involved in nucleation and crystal growth stages during protein crystallization.

I.4. X-ray Crystallography of biological samples

Before the discovery of X-rays by Wilhelm Conrad Röntgen in the late 19th century, determination of crystal structures was only possible by precise measurements of the angles between well-developed crystal faces followed by vector plots and plane symmetry calculations (Röntgen 1898). The revolutionary discovery of a new type of invisible radiation sourced from a cathode-ray tube opened a new era of novel discoveries. In the year 1912 Max von Laue opened the field of X-ray crystallography by testing the wave nature of X-rays with the use of a copper sulfate crystal, work awarded with a Nobel Prize in physics for the discovery of crystal diffraction (Friedrich, P., and Laue 1913). However, the diffraction results were only later resolved and interpreted as reflections by W.L. Bragg and W. H. Bragg, who established the fundamental law of diffraction theory (Bragg and Bragg 1913). Ever since, the theory of diffraction has been known as Bragg's Law with applications and breakthroughs in various research fields such as physics, chemistry and biology.

The advancements in the field of crystallography have enabled the development of biomolecular X-ray crystallography in the 1930s, which had an enormous impact on the field of molecular biology. The first protein structure was determined in 1958 from sperm whale myoglobin (Kendrew et al. 1958), and to date over 140,000 structures of macromolecules have been determined and deposited in the Protein Data Bank (PDB) (Berman et al. 2000). Nowadays, the availability of light sources and their technical advancements has enabled many research institutes in structural biology to own and use in-house X-ray sources for standardized operations and routine diffraction measurements of biological samples. Conventional biomolecular crystallographic approaches require protein crystals with sizes of at least a few tens of micrometers and with good diffraction quality in order to allow protein structure determination.

Synchrotron facilities are the most commonly used radiation sources for high resolution data collection. The most powerful radiation sources at the moment are known to be PETRA III (Hamburg, Germany), ESRF (Grenoble, France), DIAMOND (Oxfordshire, England), APS (Chicago, USA) and Spring-8 (Harima Science Park City, Japan). The main characteristics of synchrotron radiation can be described in terms of defined quantities of high flux (large number of photons), high brilliance (small source size and well collimated), and tunable wavelength. For data collection, typically one target crystal with dimensions usually above 50 μm is mounted on a goniometer stage and rotated through the X-ray beam at room or cryogenic temperature (approximately 100 K), preventing the destruction of the crystal due to the high flux power. The use of cryogenic temperatures slows down the free radical production which are created by the ionizing synchrotron radiation, providing more time for data collection. The routine use of cryogenic temperatures (100 K) has led also to new technologies, where cryo-preserved samples are handled by a robot-operated machine and the data collection is conducted remotely (Burkhardt et al. 2016; Cianci et al. 2017). Due to the small beam size, and with the use of ultra-fast pixel area detectors, crystals in small dimensions of only a few microns can nowadays be used for diffraction measurements in a single or a multi-crystal data collection approach (Zander et al. 2015). However, due to the high

photon flux, the radiation damage caused in the sample is high, leading to significantly lower amounts of useful data for some crystals, even for crystals where data was collected at cryogenic temperatures (Riekel, Burghammer, and Schertler 2005; Garman 2013; Nass et al. 2015). Nevertheless, progress and alternatives have been overcoming this issue where several techniques, such as using various micro-coated chips and crystallization plates alternatives with minimal background scattering proved to be of good use for lowering the impacts of radiation damage (Kisselman et al. 2011; Roedig et al. 2016).

The most novel, fourth generation X-ray sources are the free electron lasers (FELs), which offer high brilliance (more than 9 orders of magnitude brighter than at synchrotrons) intense ultrashort X-ray pulses in millisecond time intervals (Su et al. 2015; Schlichting 2015). Inspired by the X-ray characteristics, novel methods of data collection – single particle coherent diffractive imaging (SPI) and serial femtosecond crystallography (SFX) were developed by following the principle of "diffraction before destruction" (Neutze et al. 2000; Chapman et al. 2011; Boutet, Lomb, Williams, Barends, Aquila, Hunter, et al. 2012; Aquila et al. 2012). In comparison to conventional X-ray data collection, for SFX studies as referenced before, a large number of small crystals (smaller than 100 nm) are exposed to the X-ray beam, one after the other in random orientations. Prior to complete crystal destruction, diffraction images (one diffraction image per crystal) are collected and afterwards merged together to reconstruct a complete data set. The main advantage of the method relies on the use of nano- and microcrystals in their native buffer, too small to have ever been used before for X-ray data collection. One major goal of using X-ray FELs is performing time-resolved investigations on proteins and especially enzymes at room-temperature. Time-resolved studies can be performed applying serial data collection by soaking a caged substrate into the crystals that will diffuse extremely fast and directly bind to the active center. The photo-activation of the caged substrate (typically from nano-seconds to milliseconds) and the progress of the reaction within the sample is then done with the use of a pump laser with the characteristic wavelength for de-caging (Schmidt 2013). Furthermore, photo-active proteins can be probed in a similar way (Kupitz, Basu, et al. 2014b;

Barends et al. 2015; Pande et al. 2016). The information that can be gained from such studies can pave the way towards understanding enzymatic reactions with relevance in drug discovery.

Inspired by the X-ray free electron lasers, the synchrotron facilities have extended the experimental beamlines towards serial crystallography, where many protein crystals of small sizes are exposed serially in order to obtain data sets. This new data collection approach is called serial millisecond crystallography (SMX) (Nogly et al. 2015; Martin-Garcia et al. 2017). Although the diffraction principles are the same as for conventional crystallography methods, the sample preparation and delivery to the beam, the data collection method and processing are rather different, yet similar to SFX. Serial millisecond crystallography data of similar resolution to those obtained by conventional crystallography and SFX have been successfully reported (Beyerlein et al. 2017; Weinert et al. 2017; Botha et al. 2018). The already existing advancements in SMX and the further development will most probably support the acquisition of better data from smaller crystals which was until recently either impossible or tremendously tedious and time consuming.

Chapter II

Protein crystallization coupled with *in situ* dynamic light scattering (DLS)

II.1. General considerations and development of the XtalController technique

The experimental studies presented in this chapter were carried out in order to investigate the behavior of protein molecules in solution during crystallization experiments. The final outcome of the experiments such as protein crystals, protein aggregation and/or precipitation were visualized by using the built-in microscope of the XtalController900. However, prior to achieving visual information about the final state of the droplets, the *in situ* DLS maps of the R_h distribution were used to understand specific assemblies of protein molecules that could provide early-stage information about the final outcome for the ongoing crystallization experiments.

The XtalController technique is an automated crystallization device that uses *in situ* Dynamic Light Scattering (DLS) for sample quality investigation during crystallization (Meyer et al. 2012; Schubert et al. 2017). The patented version (Patent EP 2 588 649) of the XtalController was upgraded in order to facilitate the production of protein crystals while achieving information about nucleation and crystal growth phenomena.

The upgraded crystallization technique – The XtalController900 (Baitan et al. 2018) allows a full sample investigation such as monitoring the initial protein quality, protein stability during changes in temperature or sample dilution. The device offers the possibility to evaluate experiments during precipitant addition while directly monitoring and tracking the development of the protein molecules during crystallization. Optimization and scale up of screening results is also possible by reaching the necessary conditions for crystallization in a step-wise approach where the precipitant solution is added gradually to the crystallization droplet.

The main goal of this work applying the XtalController900 is to precisely produce high quality protein crystals that are suitable to a broad range of sample analysis,

such as conventional and serial crystallography, neutron diffraction or cryo-electron microscopy.

The device is designed to measure and manipulate crucial parameters based on a modified vapor-diffusion method. It allows monitoring of a protein crystallization experiment at all stages, while having precise knowledge and control of the protein solution throughout the entire crystallization experiment, based on *in situ* DLS and parameters monitoring of the sample.

The crystallization device shown in this thesis has an experimental chamber connected to a CCD camera that allows real-time monitoring of the crystallization drop. The device has a built-in microscope connected to the CCD camera, enabling a maximum resolution of 2.5 µm per pixel. The core of the experimental chamber is an ultrasensitive microbalance with a 1 µg accuracy for tracking the evolution of the sample weight over time. The crystallization is performed as a sitting-drop vapor-diffusion experiment, where the protein drop is placed on a siliconized coverslip that rests strategically on the micro-balance (Figure II-1). Based on weight changes of the droplet (caused by precipitant addition or droplet evaporation), the microbalance gives a precise input to an algorithm for immediate calculation of protein and precipitant concentrations over time. In addition, important crystallization parameters such as temperature and relative humidity are precisely tracked and regulated.

For the addition of chemical substances to the protein droplet, the device is equipped with two micro-dosage systems (contact free piezoelectric dispensers) that work at a picolitre scale. Such a small addition of substances minimizes concentration gradients and convection within the crystallization droplet. The main role of the piezoelectric dispensers is the addition of precipitant and water that can compensate for the natural evaporation of the protein droplet. Moreover, the pumps can be manually adjusted, allowing a precise addition of precipitant into the droplet, while minimizing the possibility of precipitant addition gradients outside of the desired area.

The unique feedback-controlled manipulation is achieved by *in situ* DLS investigation, which can show possible changes in protein morphology throughout the entire experimental procedure. The DLS setup consists of a standard laser that operates at 100 mW laser output, and a wavelength of 660 nm. The photomultiplier tube detector registers scattered light at 142° and the correlator covers a wide sample time range, from 400 ns to 30 s.
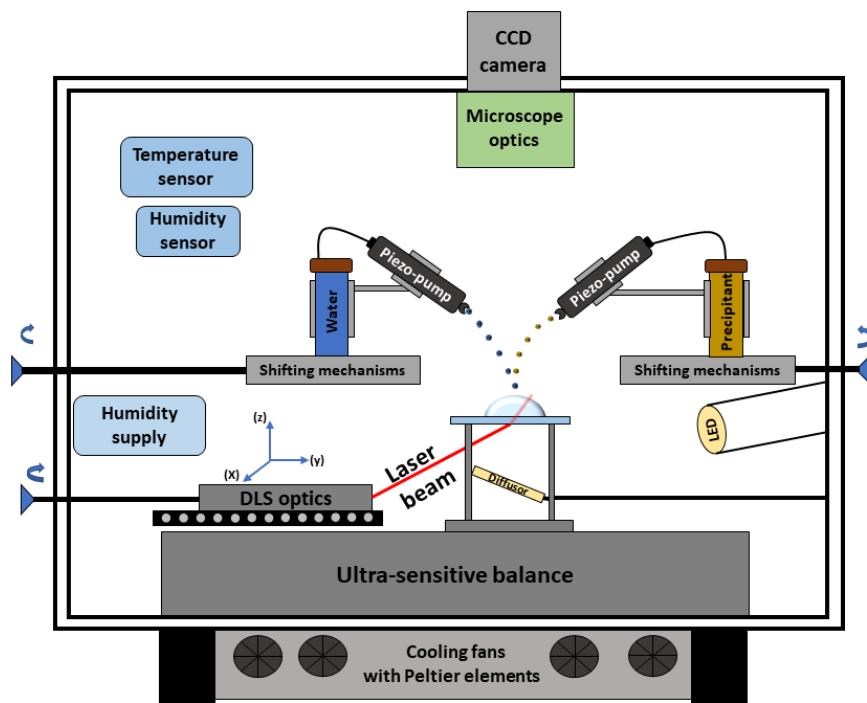

Figure II-1: Schematic representation of the XtalController900

The drawing shows an overview of the experimental chamber with all technical parts required for conducting an automated crystallization experiment: the ultra-sensitive microbalance, the positioning of the DLS optics and the micro-dosage systems used for the addition of solutions and the localization of the environmental sensors along with the microscope optics and the CCD camera. The experimental chamber is temperature isolated with the exception of the bottom-side where cooling fans and Peltier elements are placed in order to deliver the desired temperature.

The DLS optics are placed bellow the coverslip area, where the laser and the detector optics are confocally aligned with the crossing point inside the sample droplet. The scattered light from a small volume (several picolitres) of the sample droplet is recorded and evaluated. The particle distribution derived from this small volume is considered to be representative for the entire droplet volume. The device has two

openings: a front door allowing an easy positioning of the coverslip and a top lid which can be removed, to adjust the shooting position of the micro-dosage dispensers and to accurately place a new protein drop on the coverslip.

II.2. Materials and methods

II.2.1. Sample preparation

The proteins used for the automated crystallization studies were as follows: PfGST (*Plasmodium falciparum*), SP (sample of interest) and thaumatin (*Thaumatococcus daniellii*).

The protein thaumatin from *Thaumatococcus daniellii* was purchased from Sigma-Aldrich Chemie GmbH (Taufkirchen, Germany). The protein solution was prepared using a stock Bis-Tris buffer solution of 50 mM at pH 6.5. As impurities are detrimental to protein crystallization, the thaumatin solution was purified by size exclusion chromatography using an Äkta purifier with a Superdex75 prepacked column (GE Healthcare, USA). The protein solution was then concentrated from the elution fractions using a centrifugal filter concentrator with a MWCO of 3 kDa (Merck Millipore, Schwalbach, Germany). The final thaumatin concentration was determined using a Nanodrop ND-2000 from Thermo-Scientific (Erlangen, Germany). The precipitant solution was sodium tartrate dissolved in ultra-pure water to a final stock concentration of 1.2 M.

The protein SP is a sample of interest in the research group of Prof. Dr. Rolf Hilgenfeld in the field of structural virology, at the biochemistry institute, University of Lübeck (U2L) in Germany. The expression and purification were carried out at U2L and the sample was kindly provided for crystallization studies by Dr. Jian Lei.

The protocol for the expression and purification of the protein glutathione S-transferase PfGST from *Plasmodium falciparum* was first established by Dr. Liebau Eva (Liebau et al. 2002). For these studies, the expression and purification were

followed as described in the literature with minor modifications. After transformation into *E. coli* BL21(DE3) pLysS (Novagen, Schwalbach, Germany), the expression was carried out in a 1 L high density flask and allowed to contact overnight. A volume of 200 mL media was afterwards taken and supplemented with 100 mg · mL^{-1} ampicillin for plasmid selection and was further allowed to grow at 31°C. The cells were induced with 1mM Isopropyl-β-D-thiogalactopyranosid (IPTG) after the optical density of the *E.Coli* bacterial cell proliferation reached a value of 0.7 OD. The harvesting of the cells was done by centrifugation and the resulting pellet was resuspended in phosphate-buffered saline (PBS) and afterwards sonicated. The lysate was separated from the cell debris by centrifugation at 17,000 x g at 4° C for 50 minutes. The supernatant containing the soluble protein was then passed through an affinity column in order to separate and purify the soluble protein. The fast flow column was prepared with 5 mL pure Gluthathione Agarose 4B (Sigma-Aldrich Chemie GbmH) and then washed with 10 column volumes of PBS buffer. The supernatant containing the protein was then placed onto the matrix and allowed to contact for 3h. After collecting the flow-through and washing steps, the protein elution was carried out using Lysis buffer previously prepared by adding 100 mM of L-Gluthathion reduced (Sigma-Aldrich Chemie GmbH) to the PBS buffer solution. Before collecting the eluted protein, the resin and the lysis buffer were allowed to incubate for a period of 10 minutes. The concentration of the protein solution was carried out using an Amicon centrifugal filter concentrator with an MWCO of 5 kDa (Merck Millipore, Schwalbach, Germany). The final protein concentration was 10 mg · mL^{-1} and determined using an extinction coefficient of 1.184 from the absorbance at 280 nm.

Before running experiments all protein, buffer and precipitant solutions were centrifuged for 20 minutes at 16100 x g in order to remove all possible aggregates and impurities. The buffer and precipitant solutions used in the following experiments are summarized in Table 1.

Table 1: Buffer and precipitant stock solutions used for XTC experiments

Protein	Buffer composition	Precipitant composition
Thaumatin	0.5 M Bis – tris pH 6.5	1.2 – 1.5 M Sodium Tartrate
SP	20 mM Tris-HCl, pH=8.5, 150 mM NaCl, 2mM DTT	0.1 M Ammonium Sulphate, 0.1 M HEPES, pH=7.3, 10% PEG 350;
PfGST	0.1 M Phosphate buffer pH 6.5	1.6 – 2.1 M Ammonium Sulphate

II.2.2. Experimental procedure using the XtalController900

Before starting a crystallization experiment, a siliconized coverslip was first cleaned with the use of a soft tissue and placed on the microbalance, inside the experimental chamber of the XtalController900. The temperature inside the device was set to 20°C, a temperature value that was used for all crystallization experiments described in this study. Since the crystallization approach follows a modified vapour diffusion crystallization experiment, the dew point for the relative humidity in the experimental chamber was set just below the temperature value. This has the attribute of providing a relative humidity of 80 to 90%, ensuring a minimal and moderate evaporation of the protein crystallization droplet.

At the beginning, a sample volume of 6 to 8 μL was pipetted onto the coverslip inside the experimental chamber. After the initial weight of the pure protein droplet was registered, the protein weight and concentration were automatically kept constant, in order to compensate for the natural evaporation of the sample. This was achieved with the use of a micro-dosage dispenser that would add picolitre droplets of deionized water to the droplet, stabilizing its initial weight and concentration.

After the protein droplet had been secured in the experimental chamber, *in situ* DLS measurements were performed in two series. The first series comprised of measurements at the beginning of the experiment (initial protein investigation), throughout the crystallization experiment (precipitant addition) and after precipitant addition. The measurements had a duration of 60 seconds each and a time interval of 10 seconds between two measurements. The second series of

measurements were done in order to monitor the sample evolution as it follows crystal growth or precipitation. The measurements were initiated immediately after the first series finished, with a duration of 60 seconds per measurement and a time interval of 5 to 10 minutes between two measurements.

The automated crystallization is acquired by implementing a sequence of three main steps in the graphical user interface (GUI): constant mode for stabilizing the weight of the droplet and for measuring the initial *in situ* DLS, precipitant addition phase inserting the desired concentration of precipitant solution and the time for substance addition and a third step – keeping the crystallization droplet at its final concentration conditions, preventing any further evaporation. After launching the automated crystallization, the path follows three main steps as follows:

1. The droplet is kept constant for a period of 100 seconds. This time frame allows the software to recalculate the speed of water evaporation during precipitant addition and to counter-balance it via the water micro-dosage dispenser;

2. The precipitant addition is initiated by the second micro-dosage dispenser, adding precipitant solution to the protein droplet. The time interval used for this step varied from 20 to 60 minutes, depending on the protein sample, precipitant stock solution and the final precipitant concentration in the crystallization droplet;

3. The crystallization droplet is secured from natural evaporation until the end of the experiment, or if otherwise planned, until a new step of precipitant addition is required or needed.

The microscope with the CCD camera was used from the beginning of the experiment onwards for online inspection of the droplet during the entire crystallization time. Camera images were recorded throughout all experiments. For each protein that was studied, the initial protein concentration, the experimental conditions and the precipitant concentration used for automated protein crystallization are summarized in Tables 2 to 4.

The hydrodynamic radius distribution of the particles in solution measured by *in situ* DLS, the parameter information derived from the balance values, as well as the microscopic pictures of the droplets are shown and discussed in the results section.

Table 2: Acronyms and sample conditions used for thaumatin XTC900 experiments

Protein sample	Thaumatin from *Thaumatococcus daniellii*			
Experiment Acronym	THM_1X	THM_2X	THM_3	THM_4
Initial Protein Conc. (mg · mL^{-1})	14.25	20.50	11.30	11.39
Precipitant stock solution (mol · L^{-1})	1.20	1.50	1.20	1.20
Precipitant addition Conc. (mol · L^{-1})	0.54	0.90	1.17	1.13
Precipitant addition time (minutes)	45	10	30	40
Induced evaporation time (minutes)	-	-	-	-
Final Protein Conc. (mg · mL^{-1})	8.05	3.10	2.65	2.35
Final Precipitant Conc. (mol · L^{-1})	0.54	0.90	1.17	1.13

Table 3: Acronyms and sample conditions used for SP XTC900 experiments

Protein sample	SP			
Experiment Acronym	SP_1X	SP_2X	SP_3	SP_4
Initial Protein Conc. (mg · mL^{-1})	9.85	9.80	10	10
Precipitant stock solution (mol · L^{-1})	10	10	10	10
Precipitant addition Conc. (mol · L^{-1})	7.6	8.0	4.8	4.0
Precipitant addition time (minutes)	35	30	16	42
Induced evaporation time (minutes)	78	87	120	180
Final Protein Conc. (mg · mL^{-1})	6.85	6.10	20	9.60
Final Precipitant Conc. (mol · L^{-1})	12.70	12.13	13.74	6.10

Table 4: Sample conditions used for PfGST XTC900 experiments

Protein sample	PfGST from *Plasmodium falciparum*			
Experiment Acronym	PfGST_1X	PfGST_2X	PfGST_6	PfGST_12
Initial Protein Conc. (mg $\cdot$mL^{-1})	0.50	10.20	13.04	10.50
Precipitant stock solution (mol$\cdot$L^{-1})	1.60	2.10	2.10	1.60
Precipitant addition Conc. (mol$\cdot$L^{-1})	0.80	1.46	2.20	1.0
Precipitant addition time (minutes)	60	85	60	60
Induced evaporation time (minutes)	80	-	-	180
Final Protein Conc. (mg $\cdot$mL^{-1})	6.83	2.70	2.30	13.50
Final Precipitant Conc. (mol$\cdot$L^{-1})	2.17	1.46	2.2	2.50

II.3. Results and discussions

II.3.1. Automated crystallization monitored by *in situ* Dynamic Light Scattering

Automated crystallization was performed for different proteins in order to identify the similarities in crystallization that they may have in common. For each protein one set consisting of two crystallization experiments is presented. The results from the thaumatin experiments are shown as THM_1X and THM_2X in Figure II-2 and Figure II-3 and the SP protein experiments are referred to as SP_1X and SP_2X and displayed in Figure II-4 and Figure II-5. For the PfGST crystallization experiments, the outcomes are described as PfGST_1X (Figure II-7) and PfGST_2X (Figure II-8). In these figures, the radius distribution evolution, as well as the weight of the crystallization droplet together with the derived protein and precipitant concentration are shown. A summary of the experimental conditions used for each experiment is shown in Table 2 to Table 4 and the outcomes are presented in Figure II-2.

The experiments were started with a set of initial *in situ* DLS measurements in order to pre-check the stability and solubility state of the proteins. In all cases a monomodal size distribution was observed, corresponding to the protein in the buffer solution. Immediately after the automated precipitant addition was activated, changes in the hydrodynamic radius of the protein fraction could be recorded. For

all proteins, the initial particle size increased slightly, showing a broader distribution between 2.5 and 10 nm.   This behavior is due and strongly related to the supersaturation of the protein in the presence of precipitant. Upon interaction with salt solutions (such as sodium tartrate or ammonium sulphate) the chemical potential of the protein solution changes, causing a supersaturation of protein molecules in solution (Annunziata et al. 2006).
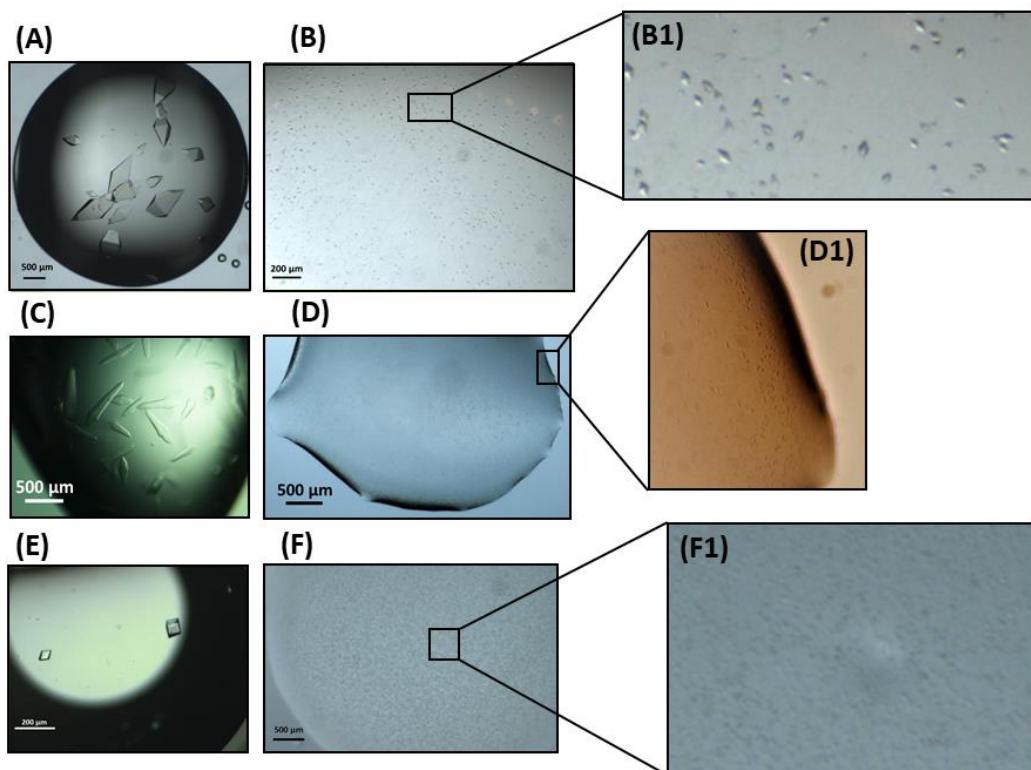

Figure II-2. Experimental outcome from XTC900 crystallization
The pictures show thaumatin results for (A) **THM_1** and (B) **THM_2**; SP results for (C) **SP_1X** and (D) **SP_2X**; and PfGST results for (F) **PfGST_1** and (G) **PfGST_2**. The results show crystals with approximate dimensions of 700 μm to 1 mm, as well as crystals with small dimensions of approximately 50 – 100 μm.

Considering the crystallization is vapor-diffusion based and over time the precipitant concentration increased in all of the shown experiments, the behavior is therefore noticed in the motion of particles and hence recorded by DLS measurements where the diffusion coefficient *D* of the particles is considered (Equation I-1). Hence, in the case of the SP experiments (Figure II-5 and Figure II-6) where a PEG cocktail was used

as the precipitant, although the change in viscosity was accounted for, the increase in the protein particle radius is stronger than for thaumatin or PfGST.

Among protein supersaturation, a second fraction of particles varying from approximately 100 to 500 nm is detected in all experiments as the precipitant concentration rises in the crystallization droplet. Previously, the R_h of similar dimensions was scored as nucleation precursors for protein crystals (Schubert et al. 2017). In another study, protein particles of sizes similar to those reported here represent metastable liquid clusters that are vital in the initial stage of protein nucleation (Gliko et al. 2007). A detailed investigation and evaluation of this particular radii fraction will be discussed in chapter III.

Within the time range of precipitant addition (Figure II-4 to Figure II-8 highlighted in grey) all protein droplets encountered a further increase in particle size towards 1000 nm. As precipitant addition is finished and the droplets are monitored at a constant rate, the growth in particle size continues following the principles of crystal growth (Boistelle and Astier 1988; M. Ataka 1995; Gadomski and Siódmiak 2003). Hence, as crystallization is favored the protein concentration slowly decreases, as a result of crystal formation. Since the growth kinetics of protein crystals is dependent on a series of factors, such as time of precipitant addition, ionic strength and concentration, the rate for the protein solution uptake is different in all cases. For thaumatin, the protein fraction was observed to have a slow decrease over time as shown in Figure II-3 and Figure II-4, whereas for the SP crystallization the protein fraction was rapidly consumed during crystal growth as seen in Figure II-5 and Figure II-6. Nevertheless, the kinetics in thaumatin crystallization were governed by two main factors – high protein concentration and fast precipitant addition, both playing a major role in the final crystalline concentration of the protein droplets. A more detailed study regarding the formation of droplets with a high crystal concentration is discussed in chapter III.3.3 and chapter IV. The final results of the experiments (Figure II-2) show protein crystals in two different morphologies that can be further used for X-ray data collection using classic or serial crystallography.
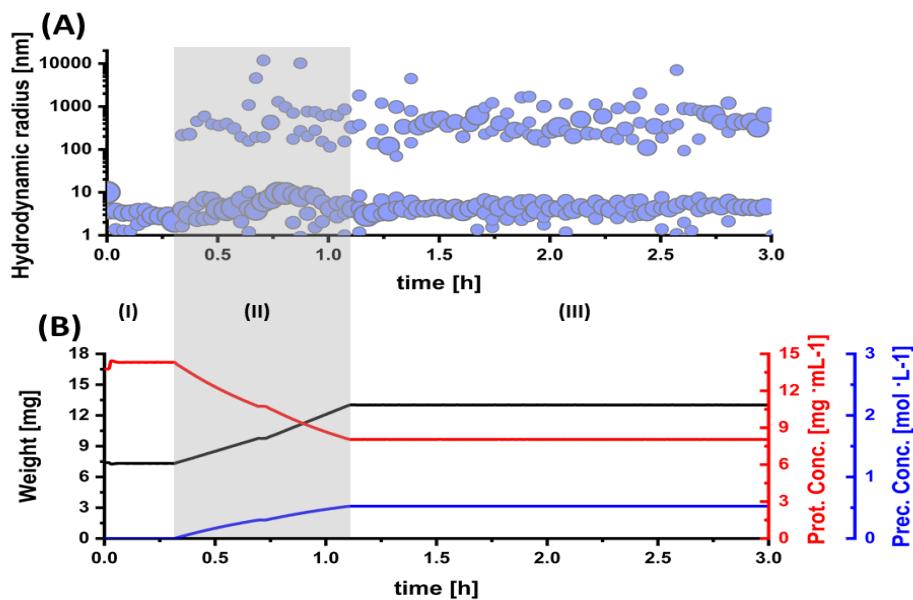
Figure II-3: Thaumatin THM_1X (A) Radius Distribution and (B) Parameter plot derived from the XTC900

The plots show the 3 steps of automated crystallization: (I) Initial DLS measurements for evaluating the state of the protein solution prior to crystallization; (II) Main precipitant addition phase highlighted in grey showing the development of particles in solution (A) as the precipitant concentration in the crystallization droplet increases over time (blue line); (III) The droplet conditions are kept constant while monitoring the evolution of the R_h distribution over time.
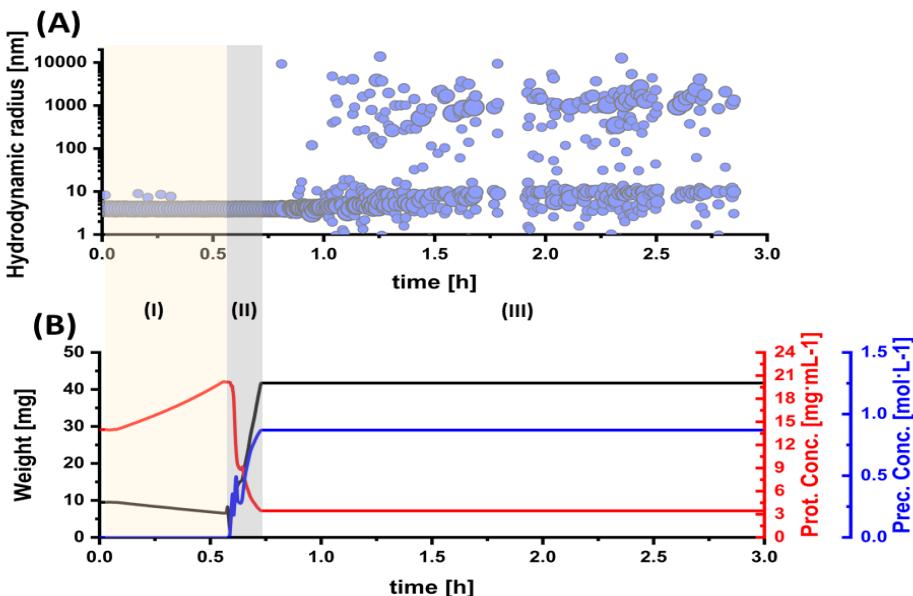

Figure II-4: Thaumatin THM_2X (A) Radius Distribution and (B) Parameter plot derived from the XTC900

The plots show the 3 steps of automated crystallization: (I) Initial DLS measurements while the droplet is undergoing controlled evaporation to increase protein concentration (red line); (II) Main precipitant addition phase highlighted in grey showing the state of protein particles in solution (A) as the precipitant concentration in the crystallization droplet increases over time (blue plot); (III) The evolution of the R_h distribution over time after precipitant addition is finished. The droplet conditions are kept constant as the nucleation and crystal growth of the droplet are tracked with *in situ* DLS.
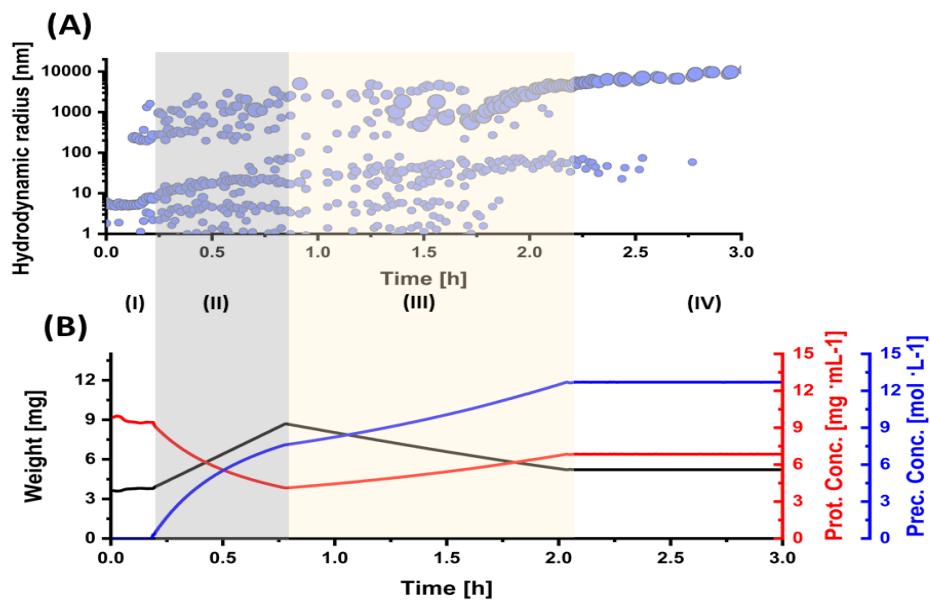
Figure II-5: SP_1X (A) Radius Distribution and (B) Parameter plot derived from the XTC900
The plots show the 4 steps involved in automated crystallization: (I) Initial DLS measurements for evaluating the state of the protein solution prior to crystallization; (II) Main precipitant addition phase highlighted in grey showing the development of particles in solution (A) as the precipitant concentration in the crystallization droplet increases over time (blue line); (III) Droplet induced evaporation to increase the precipitant (blue line) and protein (red line) concentration in order to promote nucleation; (IV) The evolution of the R_h distribution over time after precipitant addition is finished and the droplet conditions are kept constant.
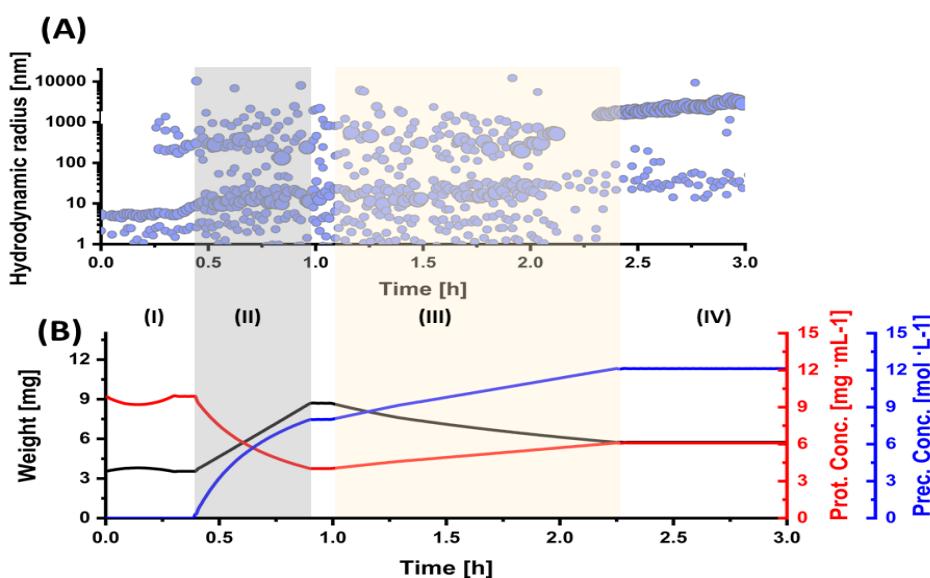

Figure II-6: SP_2X (A) Radius Distribution and (B) Parameter plot derived from the XTC900
The plots show the 4 steps involved in automated crystallization: (I) Initial DLS measurements for evaluating the state of the protein solution prior to crystallization; (II) Main precipitant addition phase highlighted in grey showing the development of protein molecules in solution (A) as the precipitant concentration in the crystallization droplet increases over time (blue line); (III) Induced evaporation for promoting the nucleation of the sample; (IV) Evolution of the R_h distribution over time while the droplet parameters are kept constant.
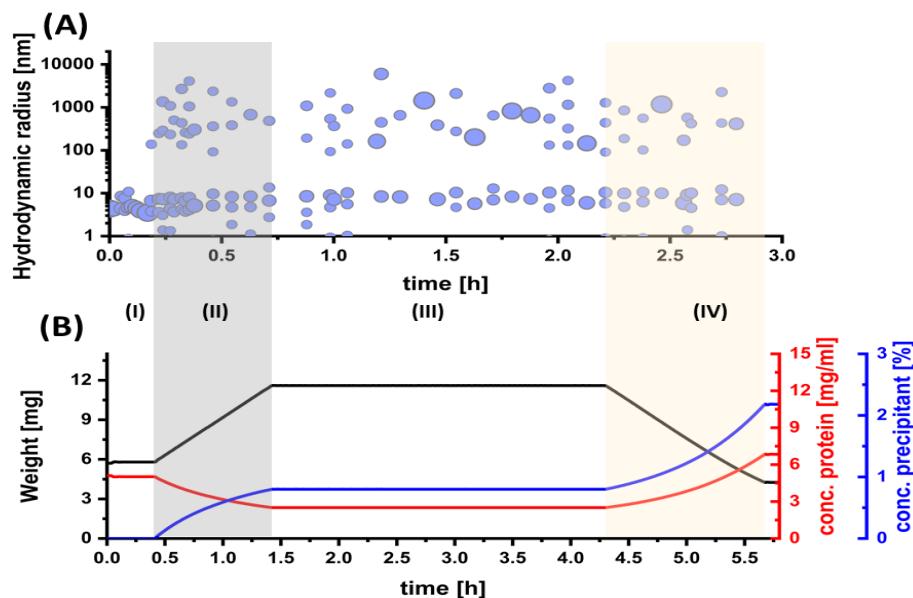
Figure II-7: PfGST_1X (A) Radius Distribution and (B) Parameter plot derived from the XTC900

The plots show the 4 steps involved in automated crystallization: (I) Initial DLS measurements for evaluating the state of the protein solution prior to crystallization; (II) Main precipitant addition phase showing the development of protein particles in solution (A) as the precipitant concentration in the crystallization droplet increases over time (blue line); (III) The evolution of the R_h distribution over time after precipitant addition is finished while the droplet conditions are kept constant; (IV) Induced evaporation to increase the precipitant (blue line) and protein (red line) concentration in order to induce nucleation.
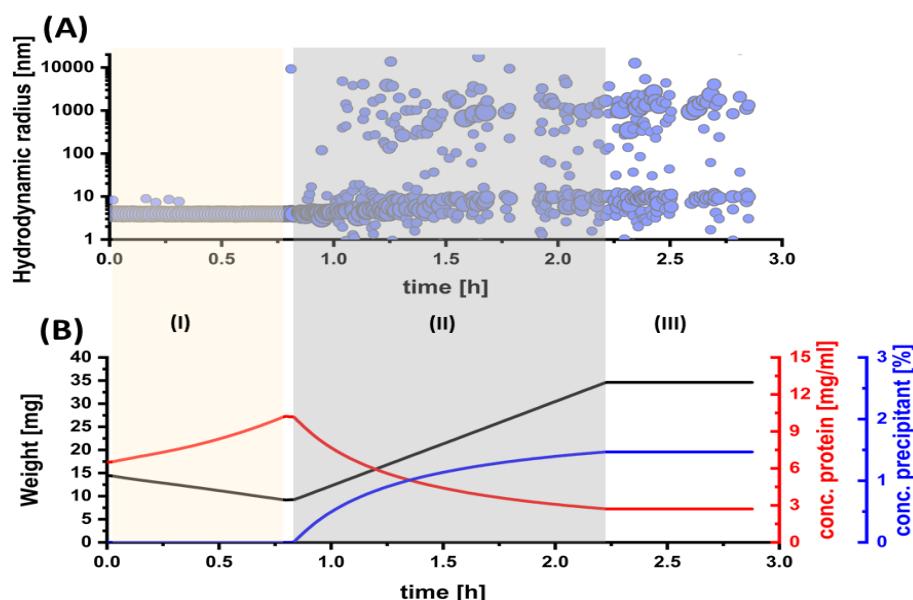

Figure II-8: PfGST_2X (A) Radius Distribution and (B) Parameter plot derived from the XTC900

The plots show the 3 steps involved in automated crystallization: (I) Initial DLS measurements while the droplet is undergoing controlled evaporation for an increase in protein concentration (red line); (II) Main precipitant addition phase showing the development of protein particles in solution (A) as the precipitant concentration in the crystallization droplet increases over time (blue plot); (III) The evolution of the R_h distribution over time after precipitant addition is finished, while the droplet conditions are kept constant.

II.3.2. *In situ* DLS maps distinguishing protein aggregation from protein crystallization

The radius distribution of particles displayed previously were plotted with the use of the XtalController900 data. Although the experiments discussed in the previous section exhibit a different size of particles during nucleation and crystal growth, a typical crystallization pattern can be observed. For all figures showing the radius distribution plots a prominent gap between the protein fraction and the second radii fraction (attributed to nucleation and crystal growth) can be observed. This can be explained considering the classical theory of nucleation and crystal growth. When nucleation is favored (exceeding the boundary of nuclei formation energy), protein molecules would come together following a similar principle of assembly. This constitutes the initiation of the nucleation phase where protein particles assemble into nuclei events within the same period of time. Nevertheless, if a second step of nucleation would proceed, this should again be noticed as a pattern, where the nuclei can be detected at a similar R_h size. In the case of SP crystallization, a second nucleation step is noticed for SP_2X (Figure II-6) where induced evaporation enhanced the development of a second nucleation phase favoring the formation of many small crystals.

For each protein, two additional experiments were carried out. The outcome was found to be precipitation as shown in Figure II-9. In contrast to the results achieved for crystallization, the experimental pathway that lead to precipitation of the studied proteins showed major differences for the radius distribution of the particles in solution.

The difference in the radius distribution between successful crystallization and precipitation for thaumatin is shown in Figure II-10. As observed previously, when the droplet undergoes crystallization, a second fraction of particles develops at approximately $100 - 1000$ nm (Figure II-10 A − B). As crystal growth proceeds, the protein fraction diminishes with the growth of protein crystals.
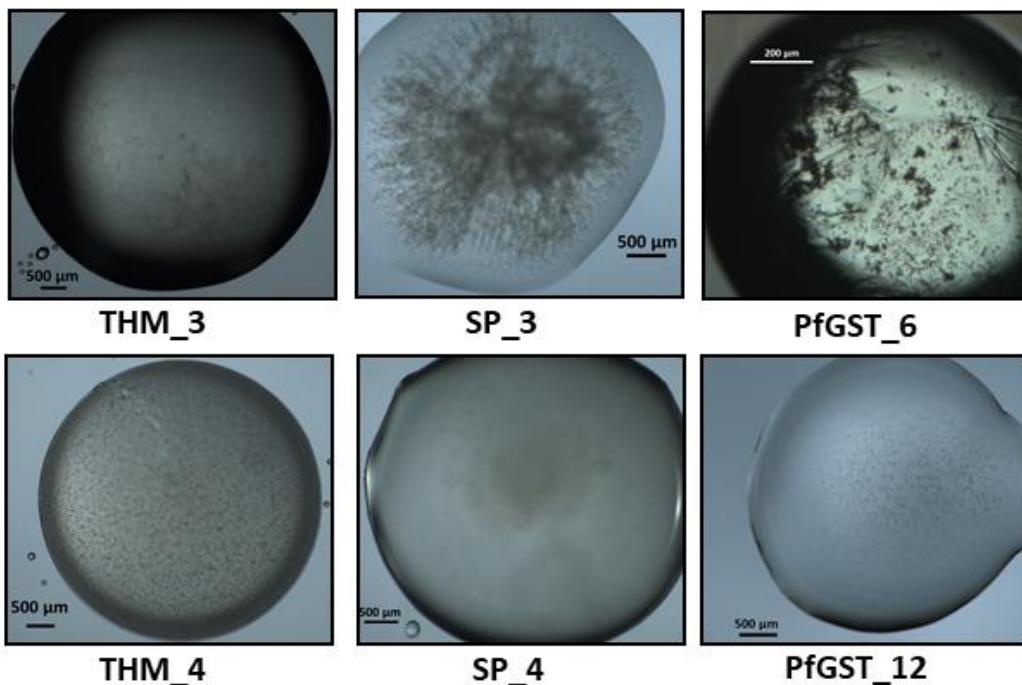
Figure II-9: Experimental outcome from the XTC900 experiments
The pictures show different types of protein precipitation for thaumatin (**THM_3X** and **THM_4X**),
SP protein (**SP_3X** and **SP_4X**) and PfGST (**PfGST_6** and **PfGST_12**).

During the growth of crystals, the protein molecules available in solution will be consumed as crystals become larger. Depending on when and how a crystallization droplet reaches equilibrium, the protein solution can be entirely or partially consumed during this stage (Asherie 2004; Bergfors 2003). Equally, for experiments where the droplets resulted in precipitation, the development of the protein molecules follows a different mechanism. During precipitation addition the protein molecules respond to the presence of precipitant agent reaching supersaturation and possible nucleation. However, the formation of the second fraction (100 – 1000 nm) is disrupted as precipitation has started (Figure II-10 C, D). The visual inspection of the droplets is shown in Figure II-9. The radius distribution maps are clearly showing how upon precipitation the protein fraction vanishes as a broad distribution of particles takes place within the droplet; a broad R_h fraction attributed to aggregation of soluble protein.

(A) Final protein concentration: 8.05 mg $\cdot$ mL^{-1}

Final precipitant concentration: 0.54 mol $\cdot$ L^{-1}

(B) Final protein concentration: 3.10 mg $\cdot$ mL^{-1}

Final precipitant concentration: 0.90 mol $\cdot$ L^{-1}

(C) Final protein concentration: 2.65 mg $\cdot$ mL^{-1}

Final precipitant concentration: 1.17 mol $\cdot$ L^{-1}

(D) Final protein concentration: 2.35 mg $\cdot$ mL^{-1}

Final precipitant concentration: 1.13 mol $\cdot$ L^{-1}
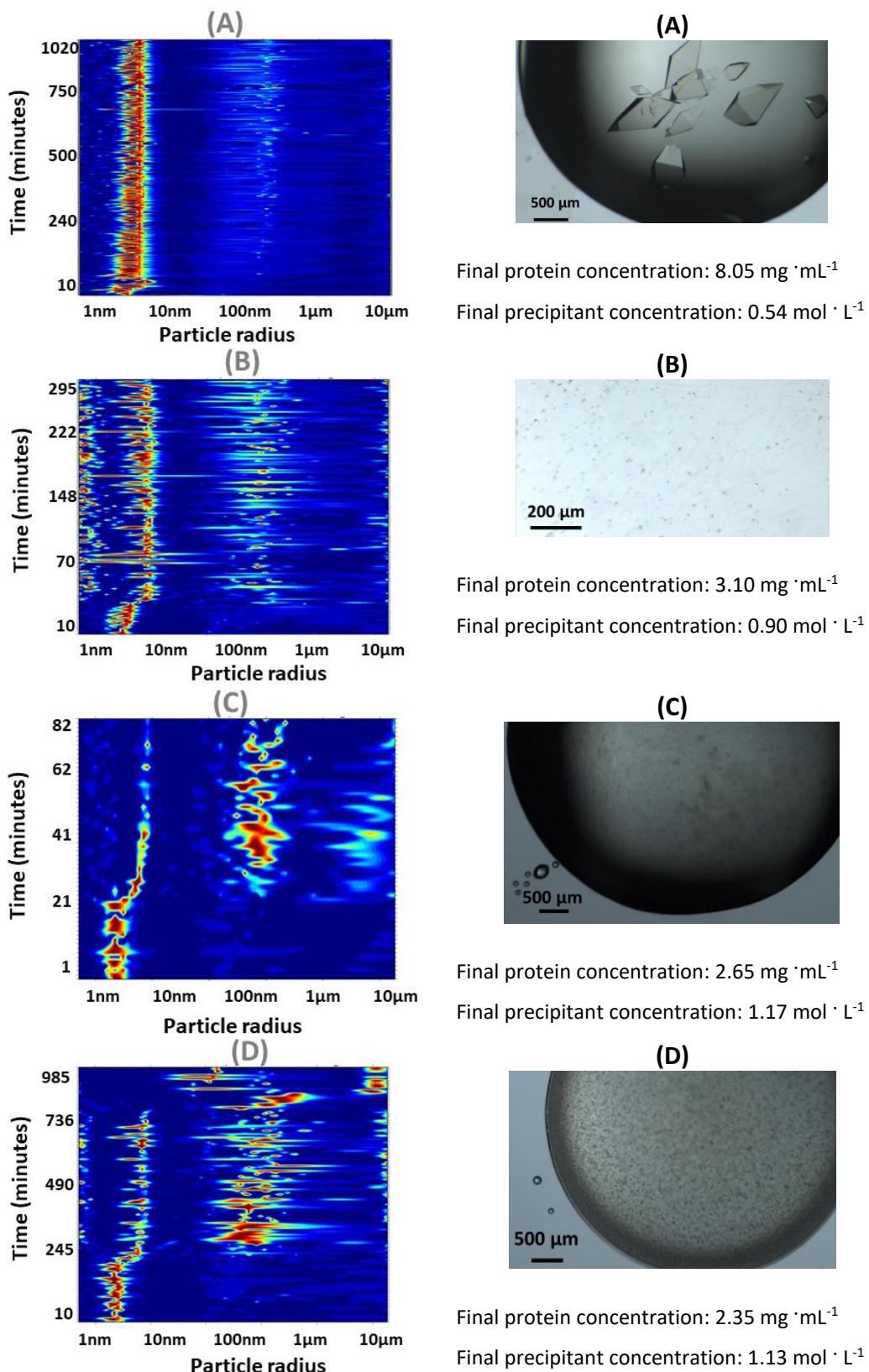
Figure II-10: Radius size distribution maps as a function of time and final state for the thaumatin crystallization experiments

The radius distribution maps and pictures (A) and (B) correspond to the crystallization experiments **THM_1X** and **THM_2X**. The two additional experiments resulting in precipitation **THM_3** and **THM_4** are displayed with the radius distribution maps (C) and (D).

Upon crystallization the SP sample showed a clear jump in particle size as nucleation took its course towards crystal growth. In Figure II-11 A and B is showed how upon crystal growth, the initial protein peak is disappearing, as the second radii fraction (attributed to nuclei) is showing a progressive growth towards 1 – 3 microns.

When the crystallization trials resulted in aggregation and precipitation, the gap in the size distribution plots between soluble protein and a second particle fraction was not present anymore. Shortly after adding precipitant the protein fraction showed an exponential increase to larger sizes, which could potentially show at first the disintegration of the protein (at sizes from a few nanometers growing to approx. 100 nm) and further formation of aggregates towards a few microns and to sizes larger than 10 µm that could not be measured by DLS anymore (Figure II-11 C and D). This is due to the larger particles (above 10 µm) that cannot follow the Brownian motion (by settling to the bottom of the droplet). For SP_4 the fast disappearance of the initial protein shows a double peak formation developing at the same time: a stable fraction at approximately 100 nm and a second fraction at approximately 10 µm that according to the aggregates visible in the droplet, could be attributed to precipitated protein. In this case, the DLS measurements that are attributed to the formation of precipitation for SP_4 can also indicate a false output of the ACF, as the intercept falls down by 10 seconds when no intensity from particles is detected. Furthermore, this could also be due to the DLS measurements taken for SP_3 starting at minute 127 and onwards. In these particular cases, the independent development of a new particle fraction with distinct dimensions is not encountered. The *in situ* DLS map (Figure II-10 C) shows a R_h distribution different from the experiments resulting in protein crystals and similar to SP_4, where the final outcome was protein precipitation.

(A)

Final protein concentration: 6.85 mg $\cdot$ mL^{-1}

Final precipitant concentration: 12.70 mol $\cdot$ L^{-1}

(B)

Final protein concentration: 6.10 mg $\cdot$ mL^{-1}

Final precipitant concentration: 12.13 mol $\cdot$ L^{1}

(C)

Final protein concentration: 20 mg $\cdot$ mL^{-1}

Final precipitant concentration: 13.74 mol $\cdot$ L^{1}

(D)

Final protein concentration: 9.58 mg $\cdot$ mL^{-1}

Final precipitant concentration: 5.10 mol $\cdot$ L^{1}
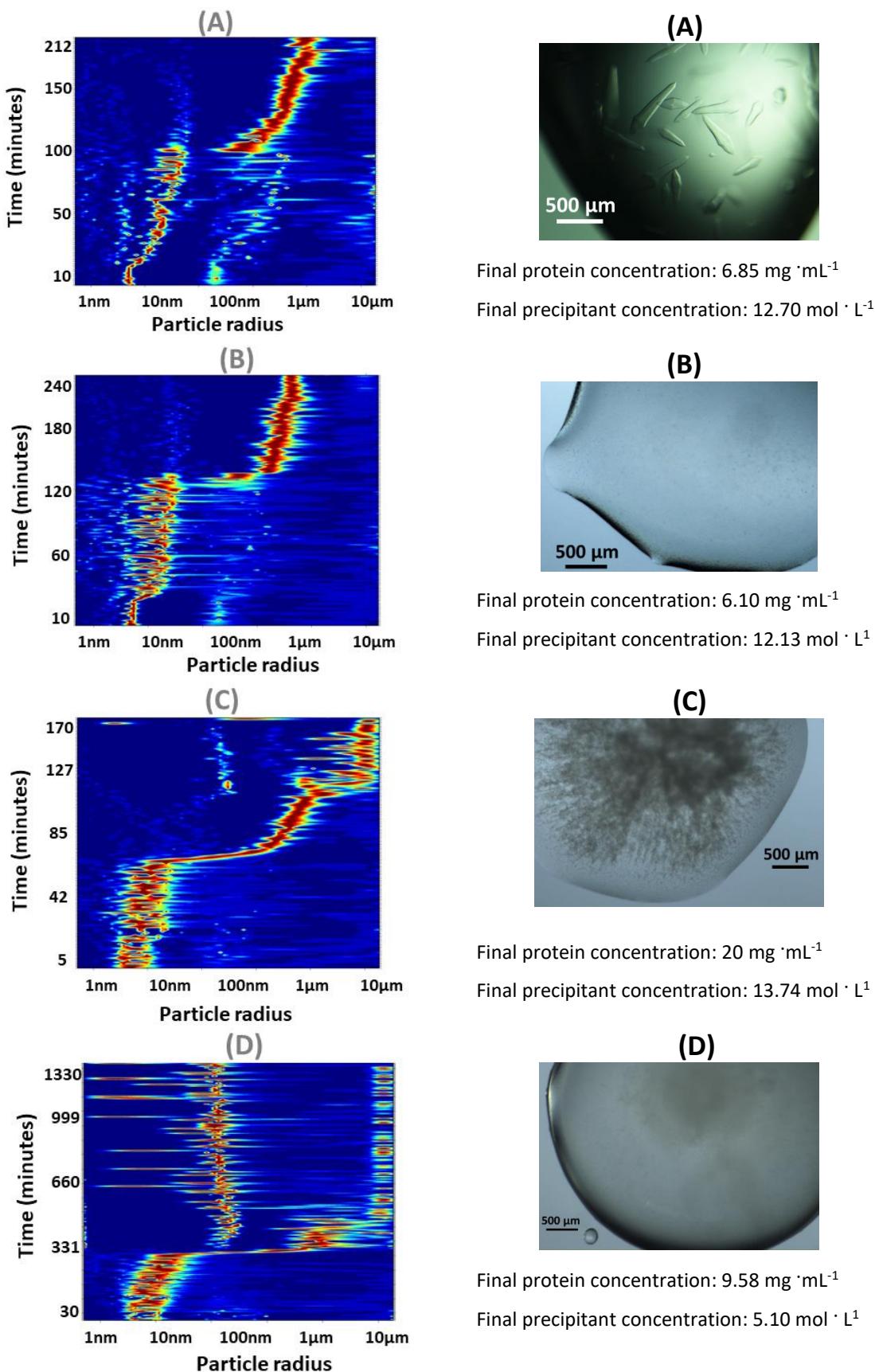
Figure II-11: Radius size distribution maps as a function of time and the final state for the SP protein
The radius distribution maps and pictures (A) and (B) correspond to **SP_1X** and **SP_2X**. The two additional experiments **SP_3** and **SP_4**, resulting in precipitation are shown in (C) and (D).

Similar to thaumatin, PfGST experiments showed a distinct development of the protein particles in solution as the droplet followed protein precipitation. The protein fraction followed supersaturation upon addition of ammonium sulphate, showing a slight increase in the dimer fraction (due to molecule ionization), while the presence of a secondary fraction at approx. 200 – 300 nm could be detected as well (Figure II-12 C). However, the supersaturation of the droplet was too high and as a result the protein started to aggregate forming particles ranging from approximately 30 to 3000 nm and further to objects visible with the CCD camera (Figure II-12). For the experiments PfGST_6 and PfGST_12 (Figure II-12 C and D) the *in situ* DLS maps show at first a similar R_h development to that found when the final outcome of the experiments was protein crystals. This development was interrupted after 2000 minutes for PfGST_6 and after approximately 500 minutes for PfGST_12. At that stage, the dimeric protein vanished as bigger particles began forming. Such a behavior was previously reported for protein precipitation, where for some proteins nucleation must proceed first, being the initiating condition for precipitation (Berrill et. al. 2011).

In general, it can be said that when the conditions in the crystallization droplet are suitable and undergo nucleation followed by crystal growth, the soluble protein is observed to slowly diminish over time, as new particles of specific dimensions are forming and developing (Boistelle and Astier 1988). The formation of the second radii fraction takes place spontaneously showing a clear gap in size from the initial protein in solution (Figure II-10 to Figure II-12, A and B).

However, when the resulting outcome is precipitation the hydrodynamic radii maps show a different mechanism for the assembly of molecules. Regardless of the initiation of precipitation (during precipitant addition or afterwards in the protein – precipitant metastable state) the formation of aggregates can be observed prior to visual observation by analysing the development of soluble protein in solution. As precipitation take places in the droplet, the hydrodynamic radii maps show how new radii fractions are developing, while the protein fraction (at approximately 2-3 nm) disappears. The main aspect and difference from crystal growth is the absence of the double particle fraction that is attributed to crystal growth.

(A)

Final protein concentration: 6.38 mg · mL^{-1}

Final precipitant concentration: 2.17 mol · L^{1}

(B)

Final protein concentration: 2.716 mg · mL^{-1}

Final precipitant concentration: 1.46 mol · L^{1}

(C)

Final protein concentration: 2.3 mg · mL^{-1}

Final precipitant concentration: 2.2 mol · L^{1}

(D)

Final protein concentration: 13.5 mg · mL^{-1}

Final precipitant concentration: 2.58 mol · L^{1}
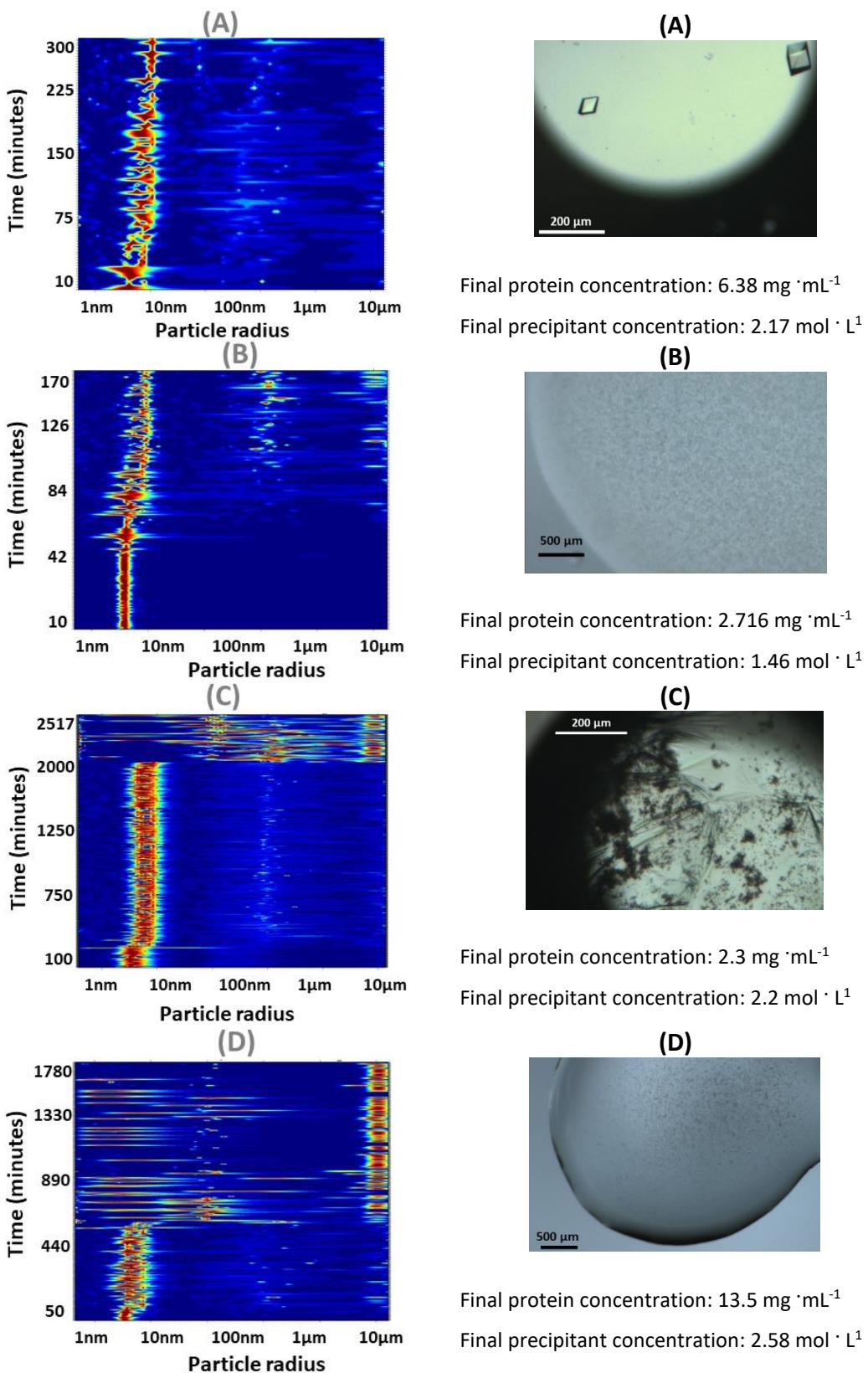
Figure II-12: Radius size distribution maps as a function of time and final state for the PfGST crystallization experiments

The distribution maps and pictures (A) and (B) correspond to **PfGST_1X** and **PfGST_2X,** respectively. The additional experiments **PfGST_6** and **PfGST_12**, resulting in precipitation are shown in (C) and (D).

For precipitation, the aggregates are based on the disintegration of the main protein fraction. In crystallization experiments, this is due to chemical degradation coming from changes in pH, temperature, mother liquor components and is manifested by breaking the van der Waals and hydrophobic attractions between side-chain and backbone atoms, minimizing or maximizing electrostatic repulsions (Roberts 2014). Given that the denaturation of a protein leads to exposed hydrophobic areas resulting in protein aggregation or precipitation, the development of such phenomena can be monitored by *in situ* DLS and found to be very different from DLS radius maps recorded for nucleation and crystal growth of protein crystals.

II.4. Conclusions

The XtalController900 proved to be a reliable method for controlled protein crystallization by precisely monitoring the protein droplet conditions throughout precipitant addition and crystal growth. The results obtained confirmed that the application of *in situ* DLS is capable of distinguishing the crystallization outcome based on differences arising in the hydrodynamic radii of protein molecules undergoing crystallization. The ability to measure the change in particle radii inside a protein droplet can help understand the outcome of a protein crystallization experiment. When a crystallization droplet undergoes crystal growth, the *in situ* DLS maps show the formation of two distinct R_h fractions that are gradually disappearing as crystals continue to grow. For experiments where the final result is protein aggregation and/or precipitation, the in situ DLS maps show a wide R_h distribution indicating disintegration of the protein molecules in solution. All three proteins investigated during crystallization trials showed a clear distinction of the molecular assembly between successful protein crystallization and protein aggregation and/or precipitation.

Chapter III

Nucleation and crystal growth mechanisms in macromolecular crystallography

III.1. Protein liquid dense clusters – precursors in protein crystallization

The topic of protein crystallization was extensively discussed over the years, with a particular focus on the nucleation mechanisms (Rein ten Wolde and Frenkel, 1997; Oleg Galkin and Vekilov, 1999; Ferrone, Ivanova, and Jasuja, 2002). According to the classical theory of nucleation, the appearance of nuclei is determined by an equilibrium between bulk and surface energy of the new phase. A high supersaturated solution is required in order to promote a nucleus, the smallest crystalline entity from which crystals will eventually grow (Gibbs 1978; Sear 2007).

Based on these fundamentals, new theories about the process of protein nucleation have been raised, suggesting that nucleation is initiated by a two-step mechanism, where crystal nuclei are preceded by formation of dense liquid phases (Vekilov, 2010). The model implies that in a first step a spontaneous formation of a new phase called liquid dense clusters takes place. The new phase, which is rich in protein, does not show any internal order but coexists in a metastable phase with the supersaturated protein solution. As the supersaturation increases and the energy barrier for crystal formation is overcome, the liquid clusters will undergo a transition towards internal order. The volume fraction of these clusters is considered to be lower than 10^{-3} % of the solution, and their size was reported to range between one hundred and a few hundred nanometers (Maes et al. 2015; Gliko et al. 2007). Further discoveries propose that the cluster formation is driven by entropy providing a thermodynamic stability of these entities among much more abundant protein in solution (Hagmeyer et al. 2012).

It is argued that nucleation often occurs within such liquid dense clusters environments, where after the initial clusters have coalesced into larger fractions (from an initial size of 250 nm to larger entities of approx. 750 nm) they can gain internal order and act as nuclei from which crystals can grow (Jawor-Baczynska,

Sefcik, and Moore D. 2013; Jawor-Baczynska et al. 2013). The role of the liquid clusters is considered to be that of crystallization precursors and a prerequisite for the crystal formation of particular proteins (Gliko et al. 2007; Pan et al. 2007; F. Zhang et al. 2012; Byington et al. 2016). Analysis of supersaturated solutions by neutron scattering showed the presence of larger clusters enforcing the initiation of crystallization (Hughes et al. 2007).

Considering the size of the clusters (from approx. 50 nm to a few hundreds) and their low concentration in solution, scattering techniques can help investigating their appearance and development within a protein droplet. Recent studies with the XtalController allowed the investigation of liquid dense clusters that could be scored by means of transmission electron microscopy (Schubert et al. 2017). Therefore, using systems such as *in situ* DLS, the particle evolution monitoring in real-time during crystallization can offer the possibility of conducting in depth-studies of liquid dense clusters and their role in the nucleation and crystal growth mechanisms of certain proteins.

In the following sections of this chapter, the XtalController900 was at first used for monitoring phase transitions within the crystallization droplets by *in situ* DLS investigation. The resulting droplets were then transferred for further analysis applying cryo-electron microscopy. In order to check if the mechanical force involved in sample transferring had affected the thermodynamic stability of the droplet, aliquots were also placed in a microcrystallization batch plate covered with paraffin oil. The microcrystallization plate was then investigated with *in situ* DLS for a period of two weeks to track the development of the R_h distribution during crystal growth.

In order to understand the mechanisms of crystal growth, a series of experiments were performed with the XtalController900 to investigate how the compensation of droplet evaporation through small additions of water in picolitre droplets could influence the development of crystal growth. In the last section of this chapter, an overview of the crystallization experiments is presented by means of experimental phase diagrams. The main reason of summarizing the results in a phase diagram plot is to gain a better understanding of why certain crystallization paths provide crystal

outcomes while other experiments following similar conditions will give different results.

III.2. Materials and methods

III.2.1. Sample preparation

The proteins thaumatin, PfGST and SP were investigated to probe and understand the behavior of proteins upon nucleation assembly and crystal growth. Prior to crystallization experiments, the samples were obtained as described in chapter II.2.1. The initial protein concentrations and precipitant compositions are provided in Table 5. Prior to all experiments, the samples were first centrifuged at 16,100 x *g* for 15 minutes in order to remove any unwanted aggregates or impurities.

Table 5: Conditions of the solutions involved in the XTC900 experiments

Protein	Initial protein conc. [mg·mL^{-1}]	Precipitant composition
Thaumatin	10	1.2 – 1.5 M Sodium Tartrate
SP	10	0.1 M Ammonium Sulphate, 0.1 M HEPES, pH=7.3, 10% PEG3, 350; 10% glycerol
PfGST	7	1.6 – 2.1 M Ammonium Sulphate

III.2.2. XtalController crystallization experiments

The nucleation and the crystal growth mechanisms were investigated using the XtalController900, previously described in chapter II.2.2. The formation of pre-nuclei entities (cluster populations) can be detected and followed with *in situ* DLS by means of the hydrodynamic radius distribution of particles over time. Due to the provided feedback conditions, reporting the concentrations of protein and precipitant solutions undergoing the phase transition, new information and experimental mapping of the phase diagrams can be achieved.

The crystallization experiments carried out for investigating the radius distribution of particles during nucleation and crystal growth were performed at a temperature of 20°C and a relative humidity of 80%, unless stated otherwise. The values were kept constant throughout the entire duration of the experiment.

III.2.3. *In situ* Dynamic Light Scattering investigation in crystallization plates

Dynamic light scattering monitoring of thaumatin experiments were performed using the SpectroLight600 from Xtal Concepts GmbH. The device is a fully automated non-invasive in-drop DLS and UV imaging system, designed to utilize standard plates using sub-microliter volumes. The laser diode provides a wavelength of 660 nm and an optical power of 100 mW. The detector is a photomultiplier tube (dark count rate less than 300 Hz) and the scattered light is detected at an angle of 142°. The visualization of the plate wells (droplets) is done with the use of a built-in microscope coupled to a CCD camera. The refractive index used for these studies was 1.33 (refractive index for water) and the viscosity was considered to be η = 1,016 cP.

The protein thaumatin was first crystallized with the XtalController900 and afterwards the droplet was transferred and placed into 2 wells of a Nunc microbatch plate (Douglas Instruments Ltd, England) which had before been coated with paraffin oil. Three series of DLS measurements were taken with the aim of obtaining long-term data.

The first series of DLS measurements were initiated immediately after transferring the droplet aliquots. The measurements were carried out for a total length of 48 hours, with a duration of 60 seconds per measurement and a waiting time between two measurements of 30 minutes. The second series of measurements was done 7 days after transferring the sample aliquots and consisted of 30 measurements per droplet, each measurement having a duration of 60 seconds. The third series of measurements was done 14 days after transferring the droplet aliquots, following the same methodology as before.

III.2.4. Cryo – Electron Microscopy

Cryo-electron microscopy images were taken in order to characterize the outcome from the XtalController900 experiments. The investigations were carried out at the Center for Cellular Imaging and Nano Analytics Biozentrum (Basel, Switzerland) in collaboration with Dr. Thorsten Blum (research group of Prof. J. P. Abrahams). The crystallization samples obtained from the XtalController900 were transferred from the siliconized coverslip to a 0.5 mL Eppendorf tube for proper transportation purposes.

The sample carrier – a 3.05 mm metallic grid on which a fenestrated carbon film is attached was at first glow-discharged for a period of 60 seconds to render the surface hydrophilic, in order to assist sample adsorption. Further, a 2 µL volume of sample was placed onto the surface of the grid and with the use of tweezers the grid was mounted for vitrification into a Vitrobot FEI robot. After the sample had migrated evenly onto the grid surface, the remaining volume of the droplet was blotted away using absorbing paper. The grid containing the sample was then automatically dropped into liquid ethane and afterwards recovered in liquid nitrogen.

Prior to data acquisition, each grid containing the sample was mounted onto a special Talos cryo-holder which was pre-cooled to -180°C. With the use of this holder, each sample was mounted into the electron microscope vacuum chamber for investigation. The data acquisition was done using a Talos F200C electron microscope, using 200 kV voltage and an exposure time of 100 ms. The pictures showing the content of each sample are shown and discussed in the results section.

The thaumatin experiment THM_15 was analyzed using two individual copper grids. For the preparation of the first grid, an aliquot of 1 µL was taken from the top of the crystallized droplet. The second aliquot was taken after mixing the droplet with the use of a pipette. This mixing step was done in order to resuspend larger particles that might have sedimented to the bottom of the crystallization droplet.

III.3. Results and discussions

III.3.1. Monitoring nucleation and crystal growth using *in situ* DLS

Thaumatin crystallization insights by monitoring with *in situ* DLS

The thaumatin experiment showed an initial hydrodynamic radius of approx. 2.5 nm indicating an initially monodisperse sample. Immediately after adding precipitant, the protein exhibited signs of supersaturation (Figure III-1 A). This could be observed by an increase in radius size for the protein fraction and by the slight appearance of larger particles at approximately 20 to 50 nm. The increase in particle dimensions continued even after the final precipitant concentration was reached (4.5 mol $\cdot$ L^{-1}). One hour after experiment initiation, the distribution of the particle radii changed; the monomeric fraction began to decrease in concentration along with the formation of a second radii band. As can be seen from the R_h distribution plot, the second population slightly starts to increase from approx. 100 nm towards 1000 nm. At this stage the experiment was stopped and the droplet was divided into 3 aliquots that were handled as follows: two aliquots were placed in a Nunc crystallization plate, previously coated with paraffin oil, while the third aliquot was placed in a 0.5 mL Eppendorf tube. In order to prevent evaporation of the crystallization droplet, a small quantity of paraffin oil was pipetted into the Eppendorf tube after the aliquot had been added. The evolution of the first two aliquot droplets was then investigated with DLS using the SpectroLight600 device (Xtal Concepts GmbH, Germany).

The crystallization aliquots were measured with *in situ* DLS for a total period of 50 hours. Figure III-1 B shows the development of R_h for the two thaumatin aliquots. The radii distribution plots show the presence of protein and a second fraction at approx. 100 nm. Prior to sample transferring, the radii distribution plot indicated the presence of protein solution and a second particle radii fraction at approx. 1000 nm. One reason for not detecting the larger particles in solution in the aliquot droplets immediately after starting the DLS measurements is that the larger particles dissolve under mechanical stress, releasing the protein into solution. This fraction is also believed to be nuclei precursors - liquid dense clusters, that over time decrease in mean cluster size, as the crystal phase commences (Streets M. Aaron Quake R.

Stephen 2010). Other studies suggest that such entities are believed to be metastable with respect to crystal nuclei and therefore their stability prior to nucleation is poor (Sleutel and Van Driessche 2014). Considering the last assumption, it is believed that upon transferring the aliquots, the population at approx. 1000 nm dissolved and reassembled after regaining stability in regards to the new environment.
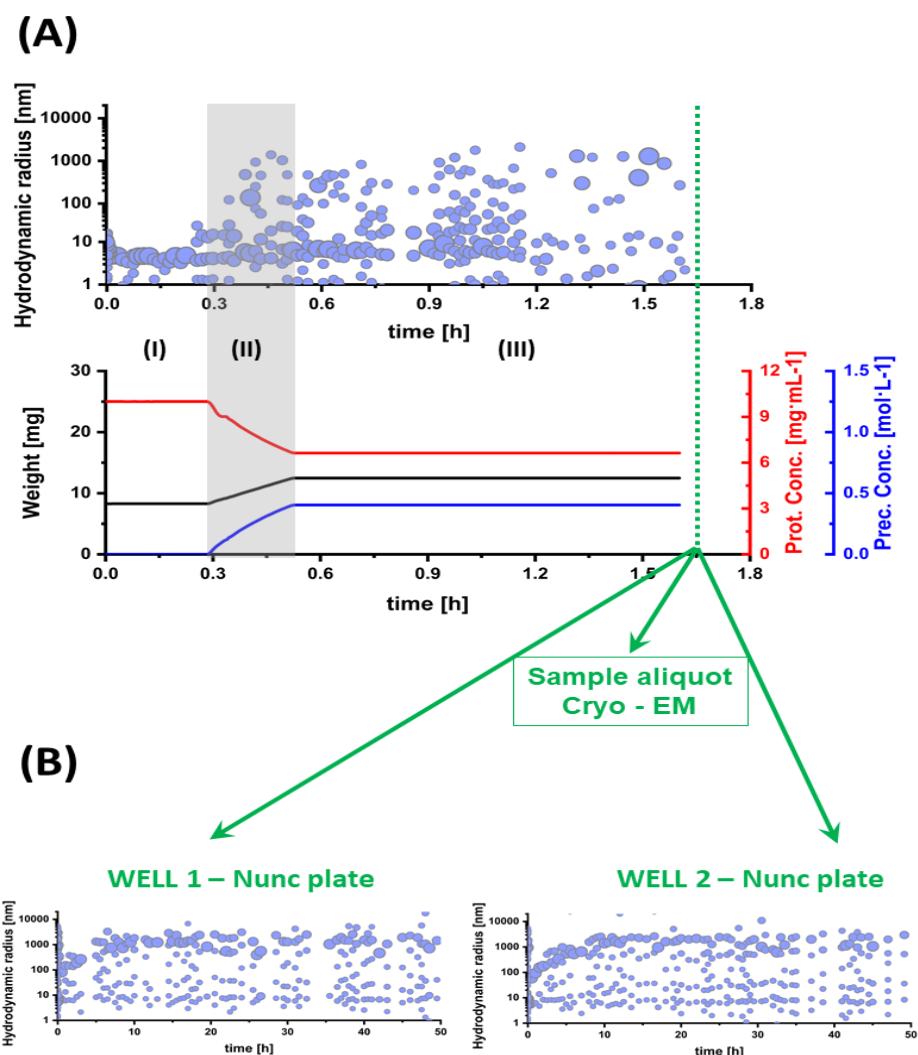

Figure III-1: Thaumatin crystallization with the XTC900 and *in situ* DLS plots for THM_15
(A) Hydrodynamic radius evolution and plots for the measured weight, protein and precipitant concentrations. The dotted green line represents the transferring of the droplet into an Eppendorf tube for EM investigation; (B) R$_h$ distribution plots for the aliquots transferred into 2 wells of a Nunc plate.
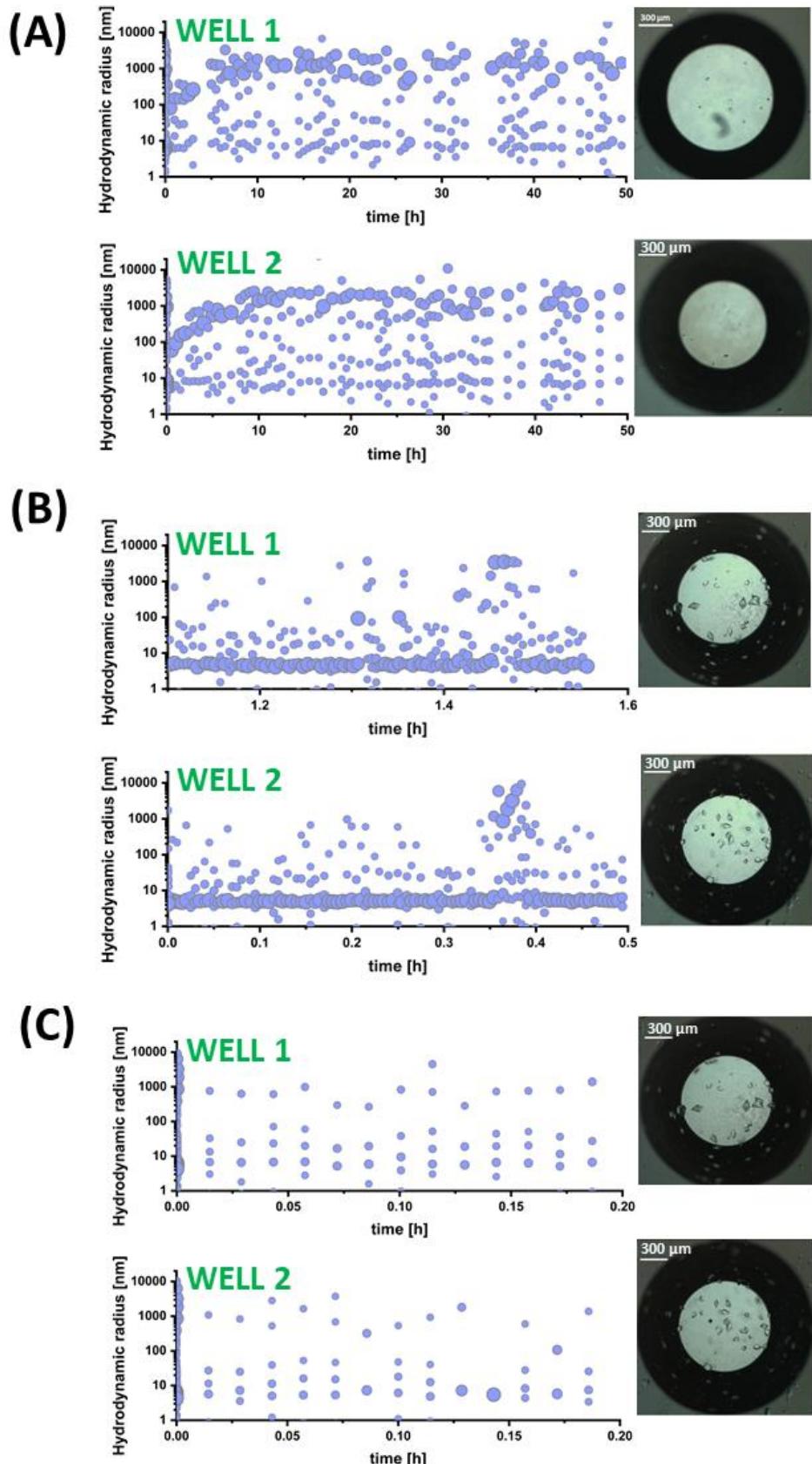
Figure III-2: Radius distribution plots and droplet images from the SPL600 for THM_15
The *in situ* DLS measurements and droplet images collected in wells 1 and 2 (A) after droplet transfer, (B) one week later and (C) two weeks after crystallization.

The two independent sets of measurements carried out in well 1 and well 2 showed a similar pattern of cluster evolution. Within a period of 10 hours, the droplets exhibited an increase in the second radii fraction until new particles formed at approx. 800 – 1200 nm. Consequently, the protein band slowly diminished during this time period. Most probably the monomer was consumed by the formation of protein precursors for nucleation. A few particles with radii variating in size were also present in the droplets, showing independent events ranging from 10 to 100 nm.

Within a period of one week, the fraction at approx. 1000 nm disappeared as crystals grew and reached equilibrium in solution. The radius fraction at approx. 4 nm indicates that among the crystals, protein solution is still available in the crystallization droplet. A specific increase in particle radii was noticed at 0.35 h on the time axis for well 2 (Figure III-2 B), and after one additional week, this fraction was attributed to protein disintegration, as protein aggregates could be seen in the crystallization wells (Figure III-2 C). The *in situ* DLS measurements showed traces of protein fraction present in the droplets after two weeks. Hence, the protein crystals remained of the same size and did not grow any further, meaning that the remaining protein solution followed partial denaturation as an aging effect, forming small aggregates that settled at the bottom of the wells.

In a second experiment with thaumatin (THM_16), the sample was at first concentrated using controlled evaporation until a final protein concentration of 17 mg $\cdot$ mL^{-1} was reached (Figure III-3). The initial DLS measurements showed the protein to be slightly polydisperse in buffer from the beginning onwards. This could be an effect of protein aging, showing partial aggregates among soluble protein. During precipitant addition the sample did not respond to the presence of sodium tartrate. However, the presence of sodium ions caused an increase in the protein R_h fraction from 2.5 nm to approx. 8 nm. Shortly after the droplet reached the final precipitant concentration of 5.5 mol $\cdot$ L^{-1} in solution, the monomer fraction exhibited a non-ordered dispersion in particle size.

After the addition of precipitant was finished, the protein fraction started to diminish as two particle radii bands developed from the protein radii fraction. The two

additional bands at approx. 10 nm and approx. 50 nm increased exponentially in size to a radius of approx. 30 nm and 200 nm within a time period of 30 minutes. At this stage, the droplet was recovered from the coverslip and transferred into a 0.5 mL Eppendorf tube for further investigation applying electron microscopy. As before, in order to prevent undesired evaporation, the droplet was covered with a thin layer of paraffin oil.
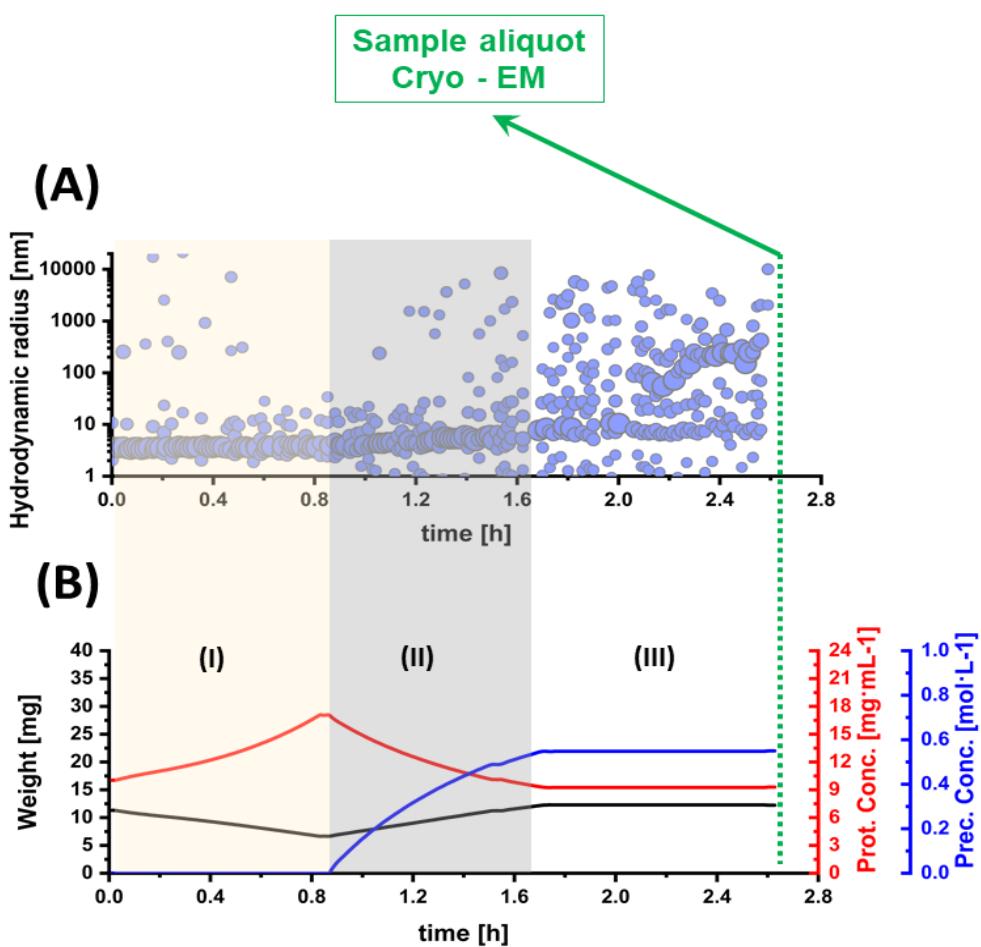

Figure III-3: Thaumatin crystallization with the XTC900, experiment THM_16
(A) Radius distribution evolution and (B) Derived balance plots for the measured weight, protein and precipitant concentrations. The dotted green line represents the transferring of the droplet into an Eppendorf tube for EM investigation. The balance plots and the R_h are divided into three main areas: (I) DLS measurements during controlled evaporation for increasing the protein concentration; (II) Main precipitant addition phase highlighted in grey showing the monitoring of the particles in solution (A) as the precipitant concentration in the crystallization droplet increases over time (blue line); (III) The droplet conditions are kept constant while monitoring the evolution of the R_h distribution prior to sample transfer.

Crystallization evolution of the PfGST and SP proteins

The SP crystallization experiment was conducted with an initial set of DLS measurements for cross-checking the sample quality. As the precipitant solution was gradually added to the droplet, the appearance of a second fraction at approx. 100 to 200 nm was observed (Figure III-4 A). With a further increase in precipitant concentration, a third fraction of particles started developing at approx. 1000 nm, growing at the expense of the second radii band.

As the crystallization droplet reached the final precipitant concentration of 8 mol $\cdot$ L^1 (Figure III-4 B) the crystallization droplet was further evaporated in order to induce nucleation.
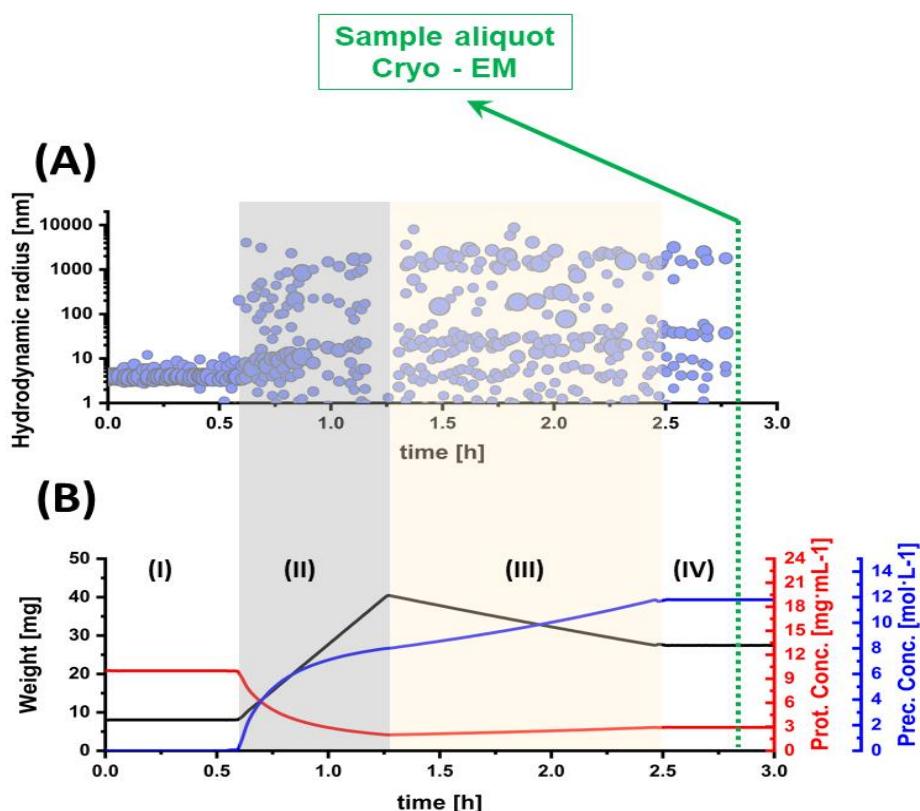
Figure III-4: SP crystallization with the XTC900, experiment SP_5

(A) Radius distribution evolution and (B) Derived balance plots for the measured weight, protein and precipitant concentrations. The dotted green line represents the transferring of the droplet into an Eppendorf tube for EM investigation. The balance plots and the R$_h$ are divided into four main areas: (I) Initial DLS measurements to assess the quality of the protein prior to crystallization; (II) Main precipitant addition phase highlighted in grey showing the monitoring of the particles in solution (A) as the precipitant concentration in the crystallization droplet increases over time (blue line); (III) Induced evaporation of the crystallization droplet with the aim of promoting nucleation. (IV)The droplet conditions are kept constant while monitoring the evolution of the R$_h$ distribution prior to sample transfer.

After 2 hours from initiating the experiment, the second fraction at 100-200 nm showed a proportional transition into the third fraction and 20 minutes later, only the protein fraction and the radii band at approx. 1000 – 2000 nm could be detected. This growth process shows that the mass increase during cluster evolution follows a step-wise mechanism which is mainly limited by particle diffusion. Since the protein fraction has reduced considerably during formation of the second and third fraction, it can be assumed that the clusters formed between approx. 100 to 200 nm grew at the expense of the soluble protein available in solution. As the hydrodynamic radius plot shows, the third fraction of larger particles formed over time, as the second fraction slowly disappeared. If the second fraction is considered to be pre-nucleation clusters, it could mean that the larger fraction grew on behalf of the liquid clusters following the theory of Ostwald ripening in the attempt of minimizing the surface free energy (Niethammer 2008). After the fractions remained stable in solution, the droplet was transferred into a 0.5 mL Eppendorf tube and was covered by a thin layer of paraffin oil for further investigations by electron microscopy.

In the experiment using the protein PfGST, no additional evaporation was applied during crystallization. The initial protein solution showed to be monodispersed until precipitant addition was initiated, when the protein molecules responded immediately to the presence of ammonium sulphate. However, the addition of precipitant caused the formation of a larger fraction at 1000 nm, which could be attributed to the formation of aggregates upon direct contact between the protein molecules and the positively charged precipitant ions (Figure III-5 A). Immediately after the precipitant reached the final concentration of 1 mol $\cdot$ L^{-1} (Figure III-5 B), the protein transitioned from supersaturation towards nucleation while a second fraction at 200 nm was observed. Although the fraction is poor in radii events and the assumption that this could be nucleation precursors is rather feeble, it has been previously reported that the volume fraction of protein clusters in solution is lower than 10^{-3} % (Gliko et al. 2007). The droplet was further kept constant and what is assumed to be a liquid clusters fraction grew over time to about 1000 – 2000 nm. After 24h, the protein as well as the larger fractions were stable in solution.

Considering that no formation of additional particles was observed and the protein droplet remained clear in solution it can be assumed that the larger fraction with a radius size at approx. 1 to 2 μm could represent microscopic crystals. Therefore, the sample was further transferred into an Eppendorf tube as described before and later analyzed by electron microscopy.
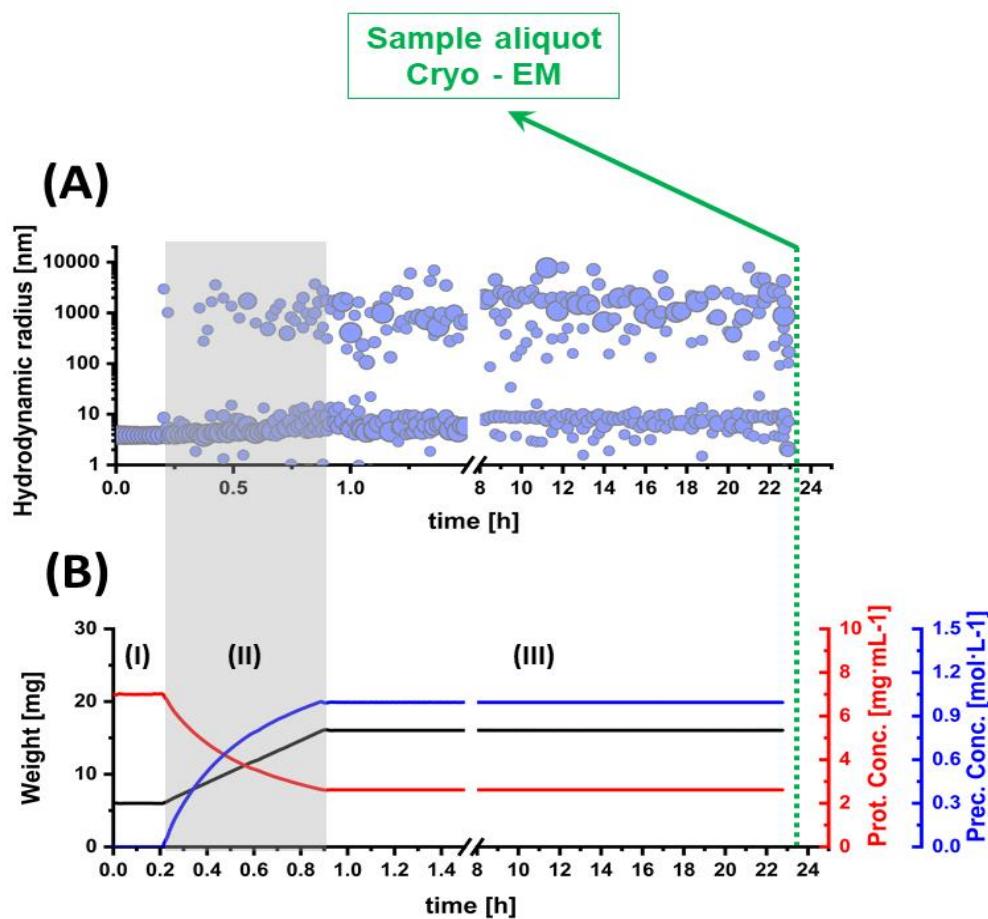

Figure III-5: PfGST crystallization with the XTC900, experiment PfGST_5
(A) Radius distribution evolution and (B) Derived balance plots for the measured weight, protein and precipitant concentrations. The dotted green line represents the transferring of the droplet into an Eppendorf tube for EM investigation. The balance plots and the R_h are divided into three main areas: (I) Initial DLS measurements to assess the quality of the protein prior to crystallization; (II) Main precipitant addition phase highlighted in grey showing the monitoring of the particles in solution (A) as the precipitant concentration in the crystallization droplet increases over time (blue line); (III) The droplet conditions are kept constant while monitoring the evolution of the R_h over time. The cut line of 6.5 hours during sample monitoring was done in order to highlight the balance plots and the R_h distribution prior as well as during precipitant concentration.

III.3.2. Analysis of pre-nuclei and crystal growth by electron microscopy

The thaumatin crystal suspension from the experiment THM_15 (Figure III-1) was further investigated by electron microscopy using two individual copper grinds.

The recorded images from the first copper grid are illustrated in Figure III-6. The size of the objects found in these pictures are in good agreement with the hydrodynamic radii measured by DLS, with particles ranging from approx. $100 - 1000$ nm. The image shows how these particles did not evenly distribute on the grid surface, but rather formed distinct populations. Using a higher magnification step, it could be observed that these entities exhibit two particular shapes. In Figure III-6 B two distinct types of shapes are distinguished: a square shape with defined angles and a second population of particles with a similar shape but without defined edges. However, the sizes of all the objects were found to be similar, between approx. 200 and 250 nm.

Research on cluster evolution and nuclei formation proved that clusters with internal order(crystal nuclei) arise from mesoscopic particles based on a competition between short-range attractive (induced dipoles) and long-range repulsive forces; the latter being the dominating force in crystal nuclei (Groenewold and Kegel 2001; Hutchens and Wang 2007). Based on research previously discussed in the introduction section, the objects with rather poor shape can be attributed to polymorphs or so-called protein liquid dense clusters. Since they are not entirely of a defined shape and do not show precise limitation of the edges, it could be assumed that they represent the vital point towards nucleation in a two-step mechanism of crystalline self-assembly. Equally, two of these objects were observed to merge into each other, displaying a more precise square shape than the third polymorph in their vicinity (Figure III-6 B). Van Driessche et al. reported new results earlier this year showing a few protein nucleation mechanisms where the formation of nuclei is driven by oriented attachments between subcritical clusters that already exhibit a degree of crystallinity. Considering the wide number of space groups investigated in their research, this could mean that Figure III-6 B and C show a building-block mechanism of self-assembly for thaumatin nuclei (Van Driessche et al. 2018). A similar theory of crystal formation was previously discussed by other research

groups, stating that for some proteins, crystal formation is based on a two-step mechanism of nucleation (Vekilov 2010; A. Sauter, Roosen-Runge, Zhang, Lotze, Feoktystov, et al. 2015; A. Sauter, Roosen-Runge, Zhang, Lotze, Jacobs, et al. 2015). Earlier in 2017, investigating the nucleation mechanism of proteins with the use of DLS and electron microscopy revealed how prior to nucleation the proteins form metastable liquid dense clusters that further transit to crystal nuclei (Schubert et al. 2017). The data presented here suggest that a similar mechanism occurs in the case of thaumatin, where prior to nuclei formation the protein clusters show signs of a geometrically defined shape. However, in depth studies such as applying electron diffraction should be further pursued before stating with full certainty that the transition of clusters to highly ordered objects indeed represent crystal nuclei.
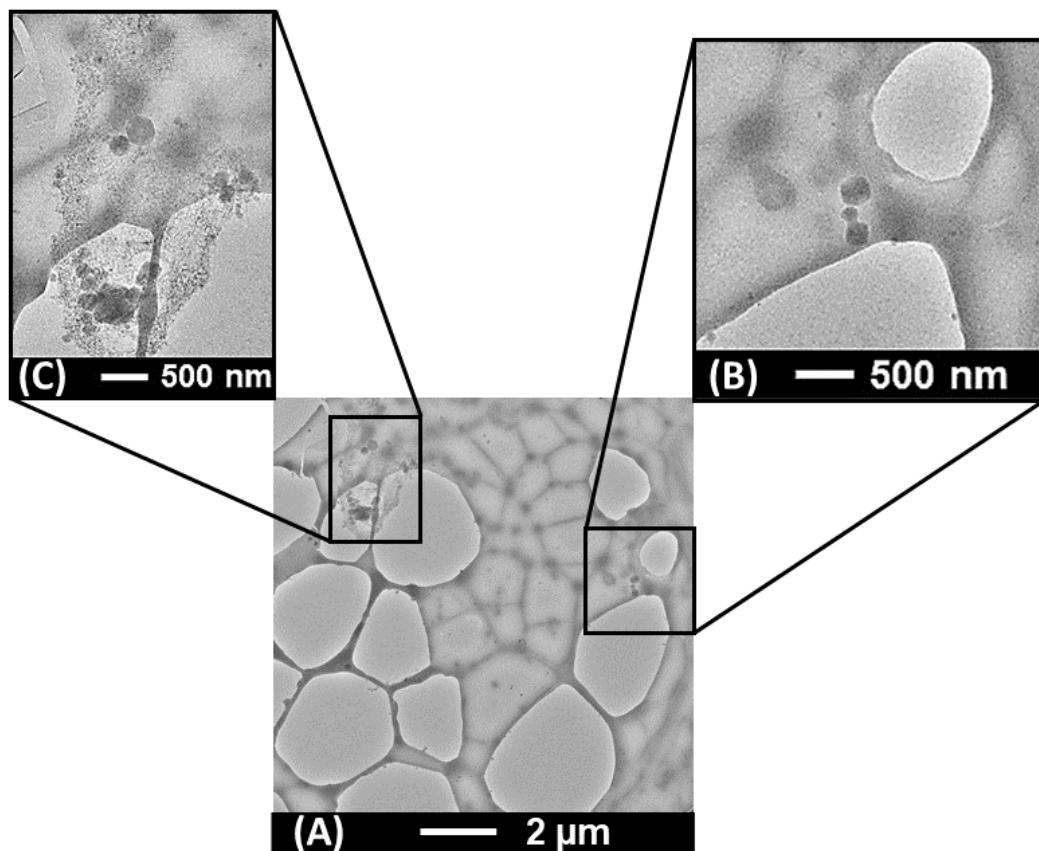
Figure III-6: Cryo-Electron Microscopy images of thaumatin (THM_15) taken from the first grid

The images obtained from the second copper grid containing the THM_15 sample are shown in Figure III-7. Since the aliquot used in this case was collected from the bottom of the droplet, it was expected that the particles would most probably be of larger sizes. The first image (A) and zoomed in picture (C) show parts of thaumatin crystals. The prior mixing of the droplet with the pipette tip most probably caused a high mechanical stress on the crystals and as a consequence they cracked and fragmented. Albeit, if the sample would not have been immediately vitrified, the broken crystals would most probably have dissolved, releasing protein into the solution.

In Figure III-7 B an object with a well-defined shape and edges similar to those reported earlier was found among the crystals. Thaumatin crystals are of tetragonal shape, belonging to the space group $P4_12_12$. Considering the position and shape of the object and the crystalline outcome of the XTC experiment, this could represent a thaumatin crystal nucleus. Since there are no further objects in the surrounding, this could support the earlier hypothesis earlier listed for THM_15 in Figure III-6 B. Based on the lattice points and planes of symmetry for crystals, and that nuclei represent the smallest crystalline entity, this could represent a view of the *001* plane top view of a thaumatin nucleus. Considering its size (approx. 500 nm), this could explain the arbitrary transition of particles between 100 – 500 nm that are always presented in the hydrodynamic radii distribution when crystallizing thaumatin.

The second thaumatin experiment (THM_16) that was evaluated with electron microscopy, (Figure III-3) showed amorphous protein and large aggregates. The copper grid containing the sample was entirely covered with fibers-like shaped objects that lack any shape or distinct size. (Figure III-8 A). Considering the wide irregular hydrodynamic radius distribution of similar particles during crystallization, and the absence of a clear DLS crystallization map (as previously shown in chapter II), these entities could represent oligomer formations by partial protein denaturation. By further scanning the grid, elongated objects could be observed among with ice crystals. Given that sodium tartrate dries into needle-like shaped crystals, it is most probable that the objects shown in Figure III-8 B correspond to some broken sodium tartrate crystals.
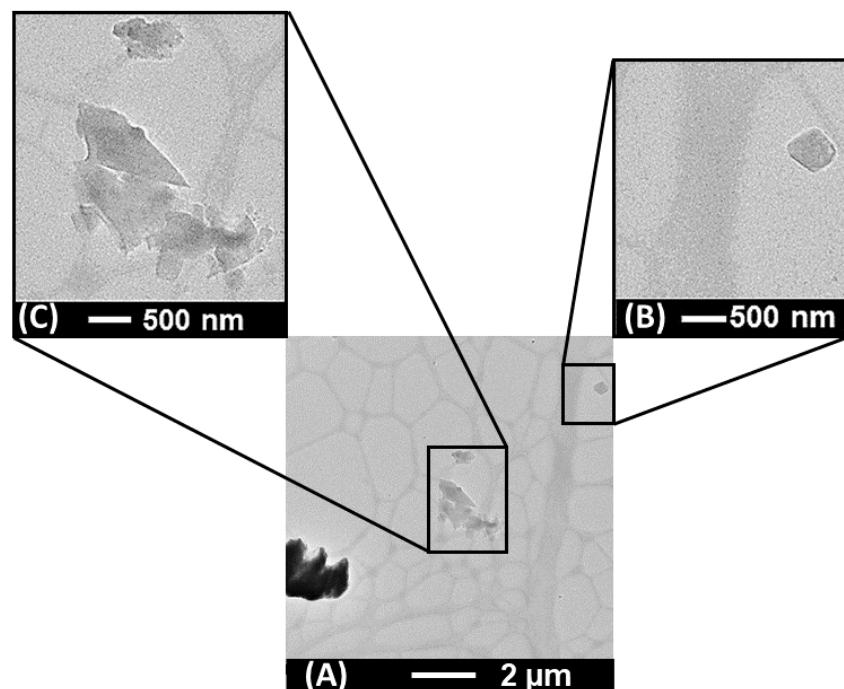
**Figure III-7: Cryo-Electron Microscopy images of thaumatin (THM_15) from the second grid**
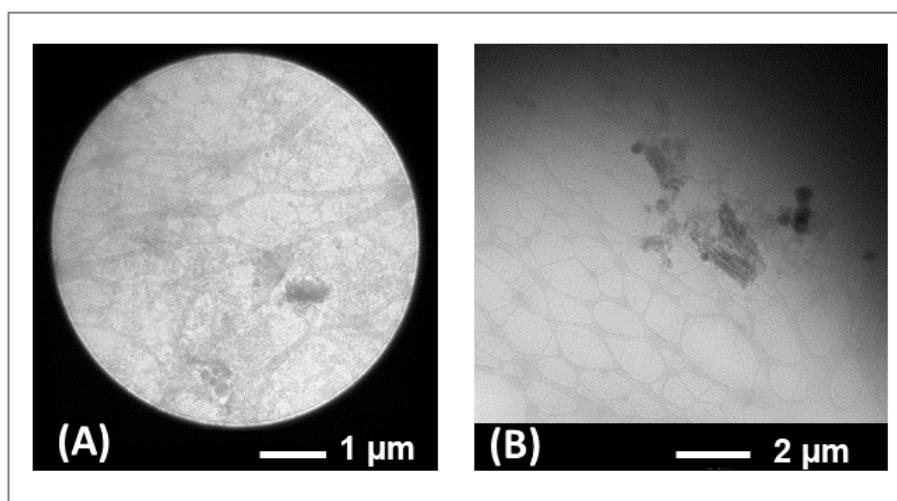
Figure III-8: Cryo-Electron Microscopy image of THM_16 showing (A) aggregated thaumatin and (B) sodium tartrate salt crystals

The crystallization samples produced with the XtalController900 using the proteins SP and PfGST were also analyzed with electron microscopy upon vitrification using plunge freezing. In the case of SP (SP_5) the images show a protein crystal with a diameter of approx. 2 µm (Figure III-9). The *in situ* DLS measurements prior to aliquot collection for this sample showed a clear band for particles at approx. 1 µm in size. Since the algorithm for radius calculation from DLS models all particles as being spherical, the crystal sizes displayed in the maps show how the DLS measurements could detect the crystals but measure only one dimension and hence consider it as a particle with a radius of approx. 1 µm (Figure III-9 A). In the case of the PfGST experiment (PfGST_5), no successful images could be taken as a consequence of the choice in precipitant solution. Ammonium sulphate is known to create a high background noise in the electron microscope. Due to these limitations, the visualization of the grid was possible only at low magnification, with one exception at a higher magnification field. However, the images taken are most probably showing dried ammonium sulphate crystals (Figure III-10 A) and some partially denatured protein among ice crystals (Figure III-10 B).
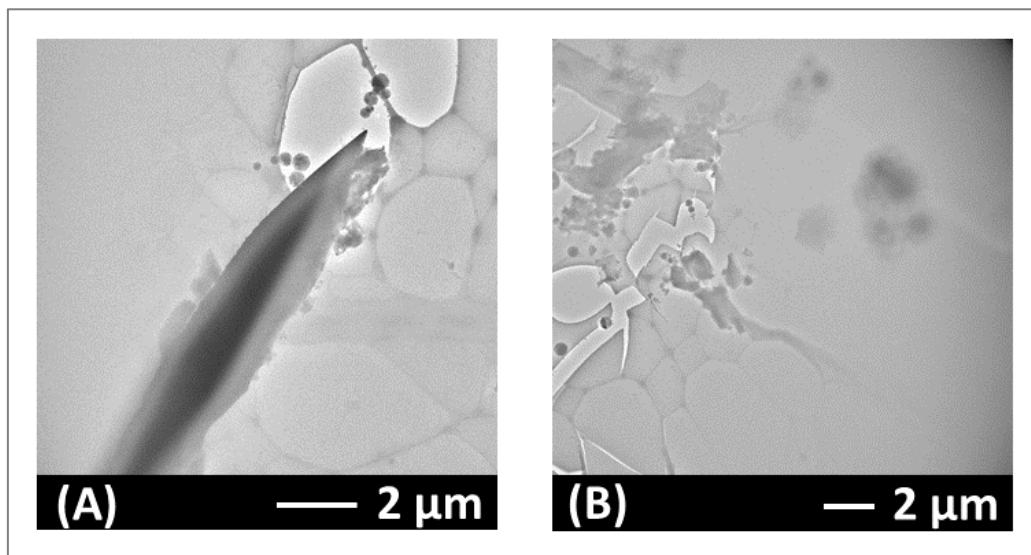

**Figure III-9: Cryo-Electron Microscopy image of SP_5 showing (A) SP crystals and (B) ice crystals**
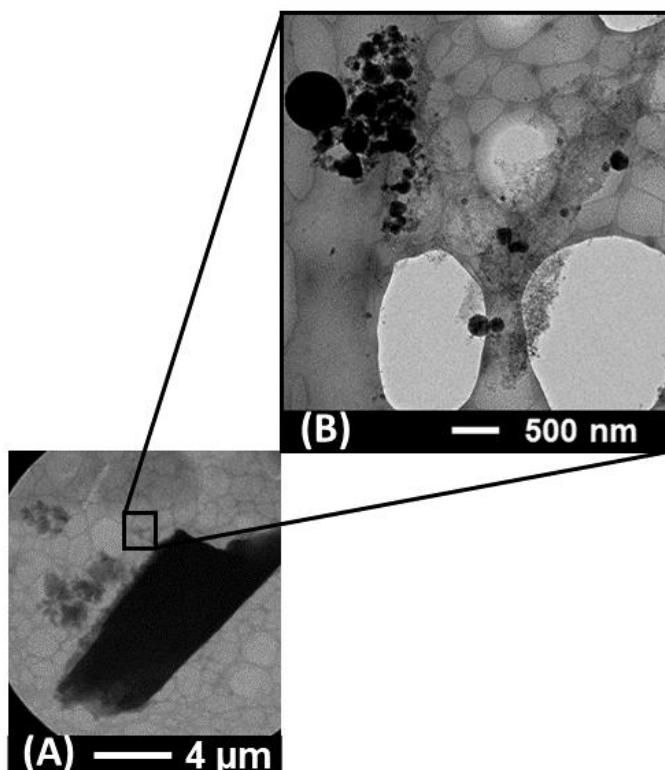
Figure III-10: Cryo-Electron Microscopy image of PfGST (PfGST_14) showing (A) ammonium sulphate crystals and (B) denatured protein and ice crystals

III.3.3. The effect of slow stirring upon protein crystal growth

As outlined before in chapter II, the protein experiments using the XtalController900 follow a modified vapor-diffusion technique where the crystallization droplet rests on a coverslip in an experimental chamber with controlled environmental conditions. Therefore, in order to prevent the crystallization droplet from evaporating, two approaches can be used. The first requires a high relative humidity in the chamber which will cause only a slow evaporation of the droplet – requiring almost no addition of water, or if necessary, at a very slow rate. The second method employs a rather low relative humidity in the chamber and therefore the droplet supplements evaporation with automated water addition using the micro-dosage dispenser; this implies the use of a high repetition rate and high frequency for the water pico-droplets addition. The effect of the second method upon the droplet will

be that of a gentle stirring, which is expected to pose an influence upon the course of crystallization and final outcome.

In the context of the experiments performed, two different relative humidity values were used for a series of two experiments: at first a maximum relative humidity of 99% was used for each experiment, followed by a duplicate experiment carried out at a lower relative humidity of 50%. The temperature was kept constant at 20°C for the entire duration of the experiment.

Previous studies have shown that protein crystallization is significantly influenced at specific stirring flows and found to favor the quality of protein crystals by giving better diffraction images (Adachi et al. 2004; Maki et al. 2008). In order to investigate how the stirring effect can affect the crystallization droplet, experiments were conducted in two independent series, following the same crystallization path with only one difference: the dew point for the environmental humidity in the experimental chamber, as outlined in the materials and methods section. The overall information of the crystallization experiments is summarized below in Table 6.

Table 6: Experimental conditions for XTC900 crystallization at different environmental humidities

	XTC at high humidity		XTC at low humidity	
Experiment Acronym	THM_6	THM_7	THM_8	THM_9
Initial Protein Conc. (mg·mL^{-1})	11.30	11.30	11.30	11.30
Precipitant stock solution (mol·L^{-1})	1.20	1.20	1.20	1.20
Precipitant addition Conc. (mol·L^{-1})	0.70	0.70	0.80	0.80
Precipitant addition time (minutes)	30	30	30	30
Induced evaporation time (minutes)	30	30	-	-
Final Protein Conc. (mg·mL^{-1})	7.20	7.20	4.10	4.10
Final Precipitant Conc. (mol·L^{-1})	1.01	0.98	-	-

Evaluation of the results obtained for XTC crystallization at high humidity

The array of protein molecules for the experiment THM_6 (Figure III-12 A) where the environmental humidity was kept at almost 100%, reveals a favorable nucleation and crystal growth pattern. The process becomes visible in the R_h distribution map as another particle fraction starts developing at approx. 100 nm. By following the development of the R_h it can be observed how the fraction at approx. 100 nm immediately develops towards a distinct radii band (approx. 1000 to 2000 nm) within a period of approximately 120 minutes. As nucleation and crystal growth proceed, the fraction of soluble protein in solution diminishes gradually as it serves for the growth of a few very large protein crystals, as seen in Figure III-12 A1.

 In the case of THM_7 where the precipitant addition and sample evaporation were similar to the first experiment, the results are different in terms of crystal size and the abundance of crystals (Figure III-12 B1). Although the two R_h distribution maps look similar, there are three key differences between the two experiments. By following the development of the particles in solution, it can be seen how for THM_7 the transition of particles from 100 to 1000 nm happens much faster than in the case of THM_6 (Figure III-12 A and B). One explanation could be that the addition of precipitant in increments and the water compensation in small droplets are favoring the diffusion of protein and precipitant molecules into the protein droplet. This will ultimately accelerate the supersaturation rate and consequently promote nucleation at a faster rate. Previously, it has been reported that stirring methods can advance the start of crystallization by inducing supersaturation, and hence, a faster de-supersaturation phase (Feliciano, Dias, and Prazeres 2000).

Another important difference is given by the width of the second band corresponding to the development of particles. The broad distribution observed in the latter case (THM_7) can be explained by considering a competition between nucleation and crystal growth. Due to the constant addition of water to prevent evaporation, an additional mechanical stress was caused to the crystallization droplet; an effect that could not be found in the case of THM_6, since no compensation of evaporation was required. As a consequence, a stirring effect is

encountered in the droplet of THM_7, enhancing probably the formation of new nucleation events at the same time with the crystal growth phase.

As the crystal growth proceeds further towards an equilibrium between particles in solution, a third main difference can be observed: the bands attributed to protein molecules in solution are not similar between the two experiments. In the later situation, the particles attributed to protein molecules are gradually disappearing over time as crystals are forming, leaving only traces of protein in solution. Considering the outcome of the experiment – an abundance of small crystals, it is clear that the protein was consumed completely by the nuclei to form stable crystals. Given the number of events detected between 100 and 1000 nm and the final outcome, the particle events can be attributed to a high number of nuclei, while in the experiment that was conducted at a low environmental humidity, only a couple of nuclei could grow to large crystals from the second radii band. Furthermore, the crystals continued growing until no other nuclei or possible nanocrystals were present in solution. For THM_6, the evolution of particles in solution was investigated until no changes in the R_h distribution could be detected anymore (Figure III-11).
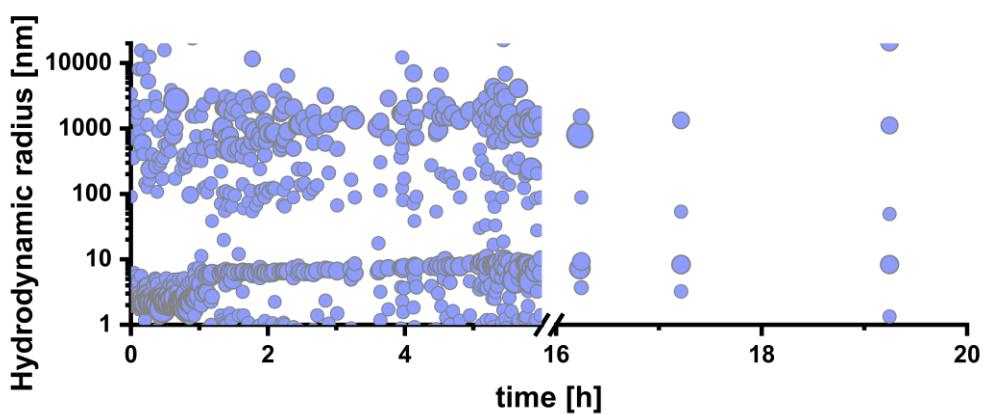
Figure III-11: Radius distribution plot for THM_6 experiment at high humidity conditions
The hydrodynamic radii versus time show the disappearance of both – the protein and the second particle band over time. The cut in the data from 6 to 16 hours for the time axis was done in order to highlight the difference in R_h distribution between the first hours of the experiment and the final state of the droplet. The evolution of the R_h shows a loss in the soluble protein band and the radii band (100 – 1000 nm) over time, as crystals grow at the expense of the remaining protein or nuclei in solution.
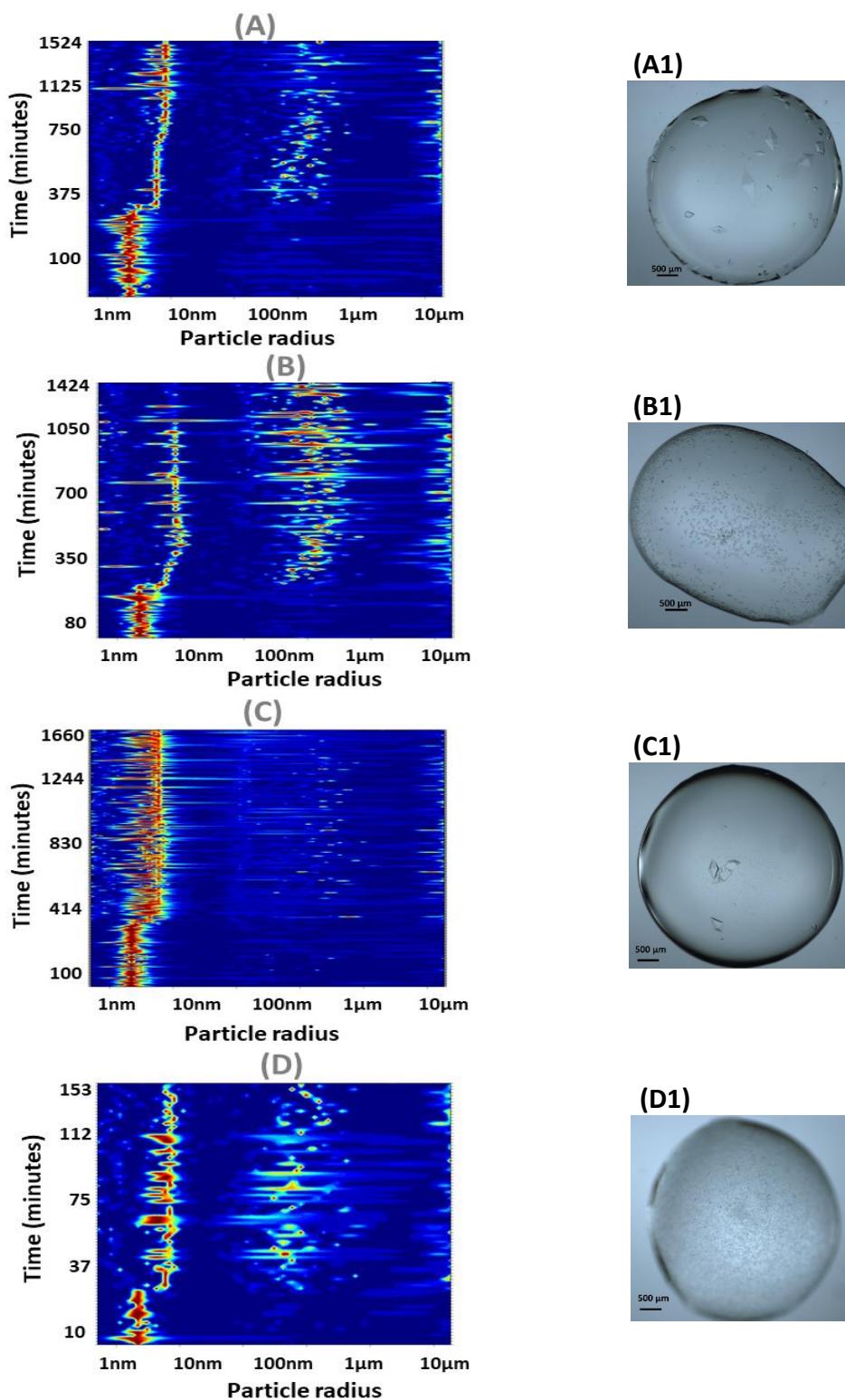
Figure III-12: Particle radius size distribution maps and pictures showing the final crystallization outcome as a function of time for the XTC crystallization in different humidity environments
The results for the R_h distribution maps and droplet crystallization outcome are shown as follows: (A) **THM_6** and (B) **THM_7** for XTC crystallization at high humidity; (C) **THM_8** and (D) **THM_9** for XTC crystallization at low humidity.

Evaluation of the results obtained for the XTC crystallization at low humidity

For the high humidity experiment – THM_8 the outcome displayed in Figure III-12 C1 only shows a few crystals, while in the second experiment THM_9, the droplet was saturated with small crystals. By analyzing the distribution maps, the outcome of these two cases could also be predicted based on the previously discussed results. When crystallizing at high humidity without external perturbations, almost no nucleation and crystal growth could be detected. Nevertheless, the R_h map (Figure III-12 C) shows the occurrence of a few events (between approx. 100 and 1000 nm) that are traceable but not abundant. According to this, it was expected that no crystals or only a few will grow towards the end of the experiment. For THM_9, where the droplet encountered a stirring effect caused by the evaporation compensation, the distribution map displays a very strong second radii band between approx. 100 and 1000 nm. As this fraction continues developing, the intercept of the ACF fell down after 10 seconds providing invalid measurements. These invalid measurements were hindered by the formation of larger particles that could not follow the Brownian motion anymore. The final outcome indicates that these particles were protein crystals that reached stable dimensions of approximately 100 µm.

Discussions and conclusions

Similar phenomena have been reported in different crystallization methods (Bergfors 2009). It was found that in one case, bulk crystallization stirring in a tubular reactor can be applied as an alternative to purification (Hekmat et al. 2017). Additionally, the stirring speed plays a major role in the final outcome (Adachi et al. 2004; Smejkal et al. 2013). In other studies, different solution stirring techniques were employed in order to understand the benefits upon crystallization. The results showed that protein crystallization is significantly influenced by controlling the stirring flow at small Reynolds number (Yaoi et al. 2004b). In another study, the same group showed how nucleation could be enhanced by wave stirring and/or slowed down when using a rotary stirring approach (Yaoi et al. 2004a).

In general, the use of the micro-dosage dispensers and the enhanced stirring effect produced by the compensation of evaporation favored the formation of small protein crystals. Due to the addition of precipitant in increments, a concentration gradient occurred when the precipitant droplet and the protein solution came into contact, producing a fast-localized supersaturation promoting the necessary energy for nuclei formation. Usually, due to the slow diffusion of protein molecules in a crystallization droplet, this localized nucleation will eventually vanish as the nuclei move within an unsaturated area of the droplet. This phenomenon is known as the metastable phase, where a competition between supersaturation and nucleation takes place (O. Galkin and Vekilov 2000; Asherie 2004). However, due to the stirring effect caused by the constant addition of water into the droplet, an equilibrium between supersaturation and nucleation could be reached faster. The frequency of the micro-dosage is directly related to the volume of the droplet and therefore, it will grow exponentially with the degree of evaporation. According to the DLS theory outlined in chapter I, and considering the volume of the droplet as well as the addition of the precipitant in increments, the protein molecules reach an equilibrium within 2.5 seconds. Nevertheless, the precipitant compound (in this case sodium tartrate) would equilibrate about 50 times faster than the protein molecules.

As previously observed, the nucleation phase is strongly enhanced in crystallization droplets where additional stirring caused by water compensation was applied. The stirring method follows a wave motion creating an ultra-fast equilibration of the droplet. As a consequence, the supersaturation is achieved faster and more crystal nuclei can equilibrate within a shorter period of time. As precipitant addition follows, the nucleation events multiply and saturate the droplet. Therefore, a higher number of nuclei benefit the crystal growth phase, the final result being a droplet abundant with crystals.

III.3.4. Experimental phase diagrams derived from the XtalController setup

As described previously, in order to obtain protein crystals the sample must at first undergo crystallization, from solubility to supersaturation, initiate nucleation and further support crystal growth (Asherie 2004). When certain crystallization outcomes are desired, a basic understanding of the phase diagram can provide guidance and help in designing optimal conditions for a desired outcome. As per general, different types of phase diagrams can be plotted and used for better understanding purposes: solubility diagrams, temperature or pH dependent diagrams, phase diagrams at different ionic strength values and so on (Mitsuo Ataka and Tanaka 1986; Mitsuo Ataka and Asai 1988; Juarez-Martines et al. 2001).

In the present work, the diagrams represent experimental crystallization conditions that could explain the main differences in the final outcome. The main critical condition and consequent focus is the nucleation phase, because at this point the macromolecules in solution are at high competition between forming either amorphous clusters and/or nuclei that will further grow into crystals.

For these purposes, a series of complementary crystallization experiments for thaumatin and PfGST were conducted with the XtalController900 in order to map out a wider range of conditions. Since the outcome of an experiment can give different results based on the variables involved, a couple of parameters were kept constant at all experimental stages: the temperature was set to 20°C while the pH and the stock solutions were kept constant to the values reported previously in Table 1. The sodium tartrate addition time for thaumatin was set to 30 minutes for each experiment whereas for PfGST the ammonium sulphate addition time was set to 60 minutes for all runs. For some experiments, induced evaporation was applied in order to push the crystallization kinetics further into a longer nucleation time. It is relevant mentioning that only experiments where the proteins exhibited a high degree of initial monodispersity were considered for plotting the experimental phase diagrams.

Table 7: Experimental conditions used for the thaumatin phase diagram and final outcome

Experiment Acronym	Initial Protein Conc. (mg ·mL^{-1})	Final Protein Conc. (mg ·mL^{-1})	Final Precipitant Conc. (mol·L^{-1})	Crystallization outcome
THM_1	13.80	6.00	0.70	Crystals
THM_2	14.00	3.45	0.87	Crystals
THM_3	11.30	2.65	1.17	Precipitation
THM_4	11.40	2.35	1.14	Precipitation
THM_5	10.00	7.10	0.91	Crystals
THM_6	11.30	7.10	0.70	Crystals
THM_7	11.20	7.20	0.70	Crystals
THM_8	11.33	5.50	0.80	Crystals
THM_9	11.20	4.03	0.80	Crystals
THM_10	11.90	2.12	0.90	Crystals
THM_11	19.00	6.45	0.80	Microcrystals
THM_12	27.80	4.69	1.00	Microcrystals
THM_13	6.37	2.37	0.95	Clear droplet
THM_14	11.22	3.33	0.95	Crystals

Table 8: Experimental conditions used for the PfGST phase diagram and final outcome

Experiment Acronym	Initial Protein Conc. (mg ·mL^{-1})	Final Protein Conc. (mg ·mL^{-1})	Final Precipitant Conc. (mol·L^{-1})	Crystallization outcome
PfGST_1	6.65	2.40	1.02	Crystals
PfGST_2	7.00	4.75	1.52	Crystals
PfGST_3	9.70	3.75	1.00	Clear droplet
PfGST_4	8.50	4.45	1.00	Clear droplet
PfGST_5	7.80	6.80	2.17	Crystals
PfGST_6	13.04	2.30	2.20	Precipitation
PfGST_7	5.00	7.35	2.10	Microcrystals
PfGST_8	8.50	7.20	1.10	Spherulites
PfGST_9	10.00	3.77	2.10	Crystals
PfGST_10	10.35	5.17	1.00	Crystals
PfGST_11	8.48	7.156	1.33	Gelation
PfGST_12	10.50	13.54	2.58	Precipitation
PfGST_13	5.00	7.35	2.35	Precipitation
PfGST_14	15.00	2.50	1.00	Clear droplet

Experimental phase diagram for thaumatin

The initial and final conditions of the experiments used for plotting the experimental phase diagram for thaumatin are shown in Table 7. Based on the protein R_h evolution throughout the experiments and considering the final outcome of the experiments (Figure III-14 and Figure III-15), the phase diagram was separated into different phase transition areas: solubility and supersaturation, nucleation, crystal growth and precipitation. An additional zone called induced nucleation was also placed in the graphic representation as seen in Figure III-12. Since the additional zone of induced nucleation is subject to different kinetic phenomena, no precise boundaries could be assumed but only a rough localization of occurrence. Based on the R_h distribution maps and final outcome of the droplets, at high protein concentration the samples gave no response in the presence of small quantities of precipitant, meaning that the protein solution at this stage was still localized in an unsaturated area.
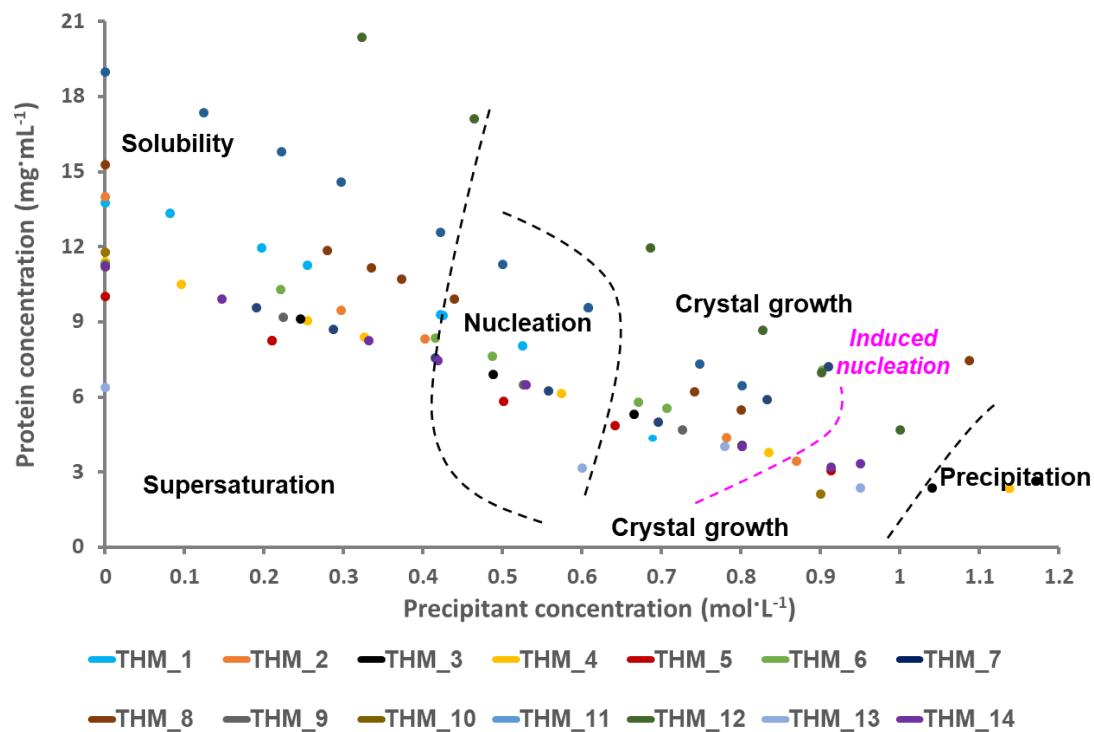
Figure III-13: Experimental phase diagram for the thaumatin XTC900 crystallization

When the protein solution was below 10 mg · mL^{-1}, supersaturation was achieved immediately. The results show that for experiments where the protein concentration was below 15 mg · mL^{-1} and a precipitant concentration between 0.6 – 0.9 mol · L^{-1}, the outcome will most probably be protein crystals, but their number and size would not be similar. On one side, the stirring effect previously discussed enforced the supersaturation and formation of multiple crystal nuclei. On the other side, the experiments THM_1 and THM_6 followed a similar path through the phase diagram with one exception: after the precipitant addition finished, the droplet THM_6 was further evaporated to reach a similar protein and precipitant concentration as in THM_1. In regard to the final outcome, large crystals of approximately 800 – 1000 μm are present in both droplets. However, the latter also showed a rich amount of small crystals among the large ones, which could have been cause by a second step of nucleation, triggered by the induced evaporation.
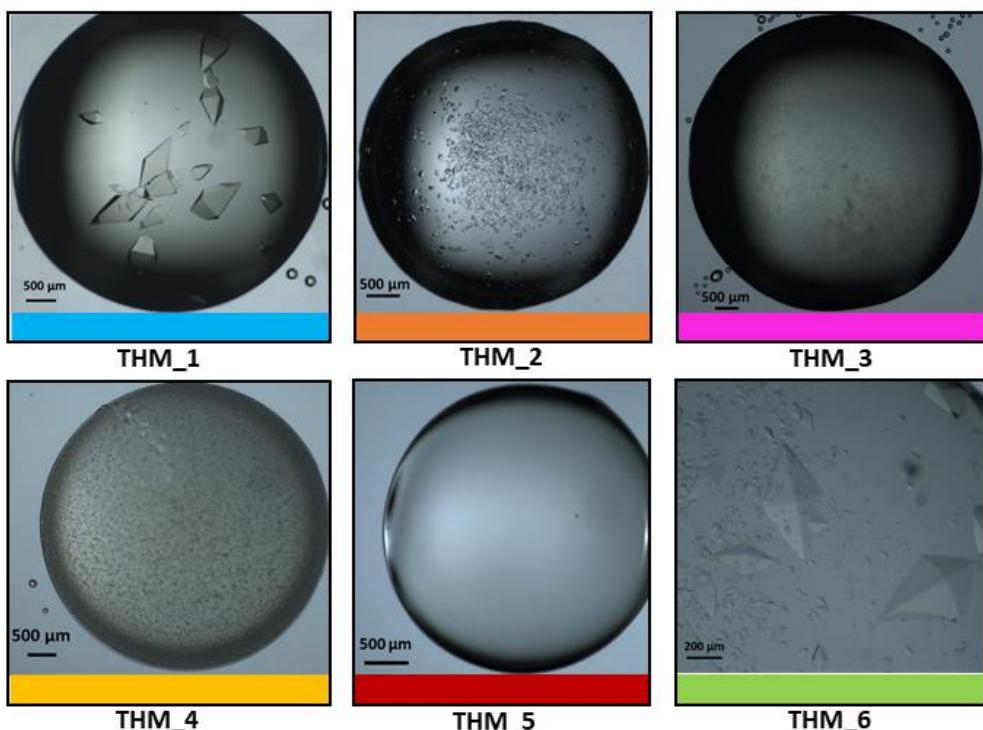
Figure III-14: Crystallization outcome for THM experiments (THM_1 to THM_6) described by the phase diagram

The pictures show the final outcome of the crystallization experiments used in the experimental phase diagram; **THM_1**: Large bipyramidal crystals; **THM_2**: Protein crystals stacking due to their high number in the droplet volume; **THM_3**: protein aggregation; **THM_4**: heavy protein precipitation; **THM_5**: clear droplet; **THM_6**: large protein crystals saturated by small crystals.
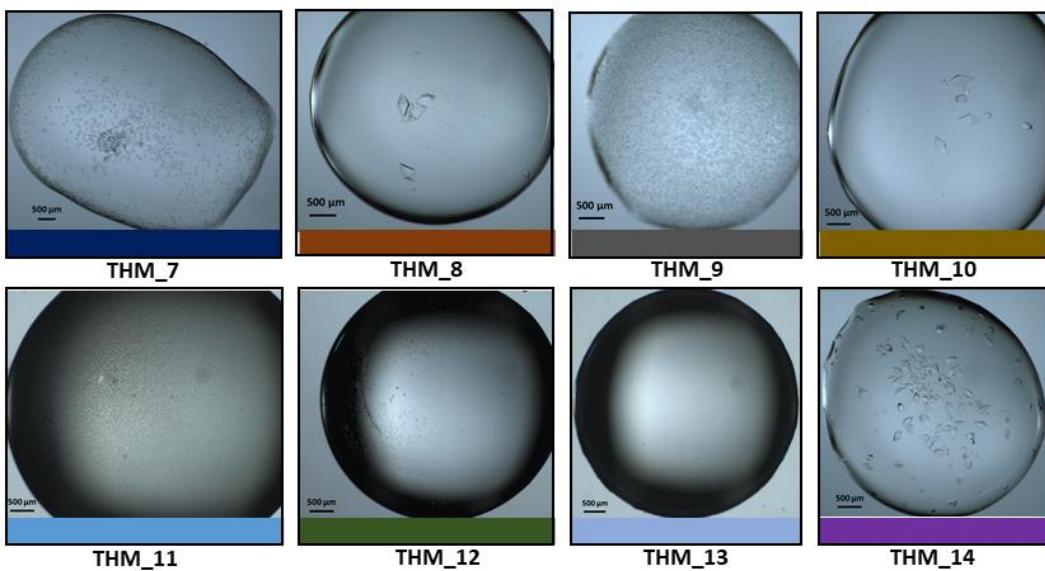
Figure III-15: Crystallization outcome for THM experiments (THM_7 to THM_14) described by the phase diagram

The pictures show the final outcome of the crystallization experiments used in the experimental phase diagram; **THM_7**: small protein crystals; **THM_8**: protein crystals sheered by crystal growth; **THM_9**: droplet saturated in small crystals; **THM_10**: protein crystals saturated by nanocrystals; **THM_11**: droplet saturated in microcrystals; **THM_12**: microcrystal shower; **THM_13**: clear droplet; **THM_14**: protein crystals varying in size.

Equally, when the precipitant concentration was increased to higher values the outcome was a large amount of protein crystals. Previous studies have reported the ionic strength to be the main driving parameter that influences the crystallization of thaumatin (Juarez-Martines et al. 2001). In this case, when the protein concentration was further increased to values above 18 mg · mL^{-1}, the outcome was found to be microcrystallization, as shown for THM_10 and THM_11. For both of these cases, the radius distribution showed a fast jump from 100 nm towards 1000 nm. Since the availability of protein in solution is higher, the probability of forming a higher number of crystal nuclei is proportional to the degree of supersaturation. It was previously reported that at high supersaturation, spontaneous or fast nucleation is promoted, favoring the formation of sub-microscopic crystals (Bergfors 2003). The crystalline quality of the experiment THM_10 and THM_11 are further analyzed and discussed in chapter IV. As precipitant was further added while the available protein in solution was high, the competition between nuclei decreased as the protein solution sufficed for stabilization and equilibration of most probably all crystal nuclei, disfavoring the

Ostwald ripening phenomena that probably took place in experiments such as THM_1 and THM_6. At even higher precipitant concentrations (approx. 1 mol · L⁻¹) the outcome was protein precipitation, as seen for THM_3 and THM_4 in Figure III-14. These particular results from sample precipitation were discussed in more details in chapter II.3.2., as well as the experiments THM_6 to THM_9 in chapter III.3.1.

Experimental phase diagram for PfGST

Using as experimental input the feedback data given by the microbalance during crystallization, an experimental phase diagram was drawn for a comprehensive understanding of the PfGST crystallization process. A schematic representation comprising of the main reaction phases found in PfGST crystallization are shown in Figure III-16. The final outcome of the experiments is shown in Figure III-17 and Figure III-18 while the conditions used for crystallization are provided in Table 8. Similar to thaumatin, the experiments show that PfGST microcrystals form at higher protein concentrations.
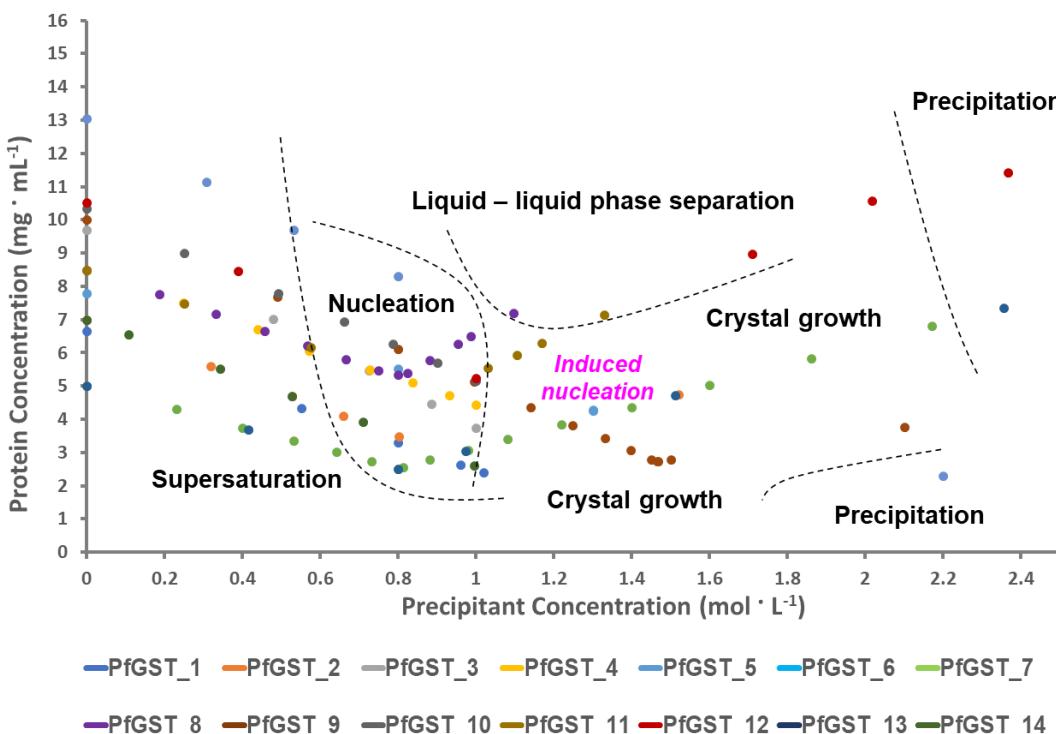
Figure III-16: Experimental phase diagram for PfGST XTC900 crystallization

Some PfGST experiments revealed a critical liquid-liquid phase separation area where the droplet showed spherulites or stronger phase separation, such as gelation. Other authors reported that while protein crystals are present at a thermodynamic equilibrium, the liquid – liquid phase separation was observed to occur in the absence of crystal nucleation (Dumetz et al. 2008). In the current results, the separation phase was found to be located at high protein concentrations, past the nucleation region. Based on the experimental conditions and final outcome, it was found that at a moderate protein concentration (approx. 7 mg · mL) and a medium precipitant concentration (1 mol · L^{-1}) the liquid – liquid phase separation in the form of protein spherulites was promoted during sample evaporation, resulting in protein spherulites (experiment PfGST_8).
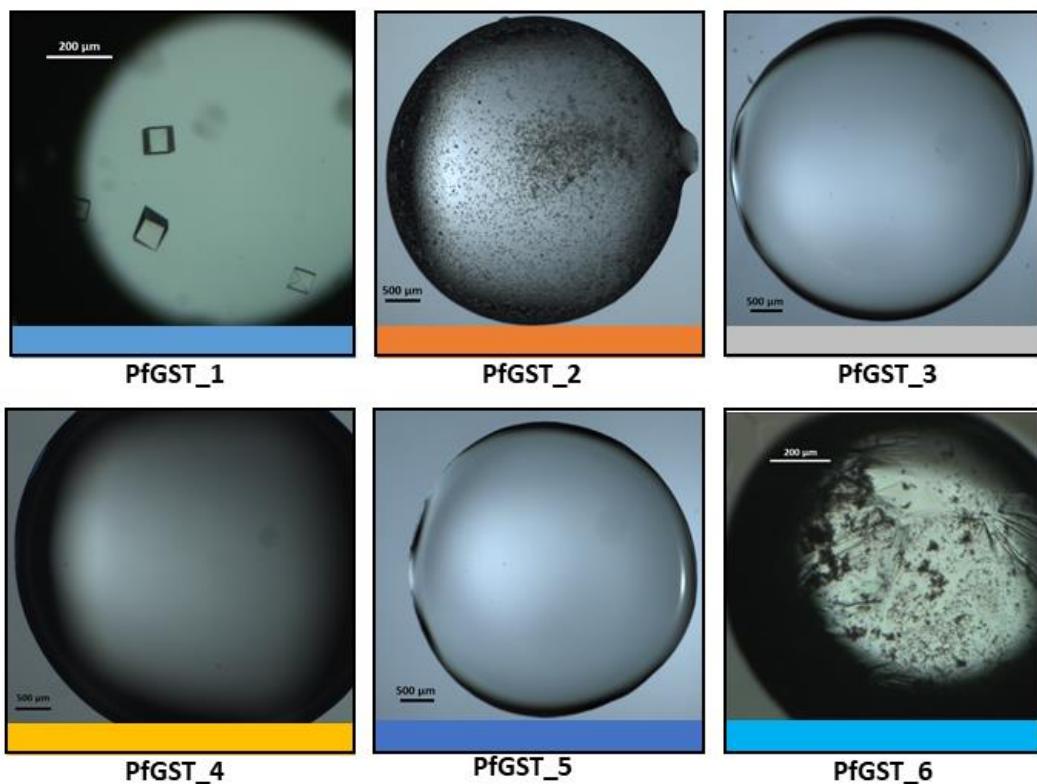
Figure III-17:Crystallization outcome for the PfGST experiments (PfGST_1 – PfGST_6) described by the phase diagram

The pictures show the final outcome of the crystallization experiments used in the experimental phase diagram; **PfGST_1**: tetrameric crystals observed 48h after precipitant addition; **PfGST_2**: formation of small tetrameric crystals that further developed into dark spherulites during induced evaporation; **PfGST_3**: clear droplet of supersaturated protein in solution; **PfGST_4**: clear droplet of supersaturated protein in solution; **PfGST_5**: clear droplet of supersaturated protein in solution; **PfGST_6**:Precipitated protein in solution that could be observed 2h after precipitant addition.
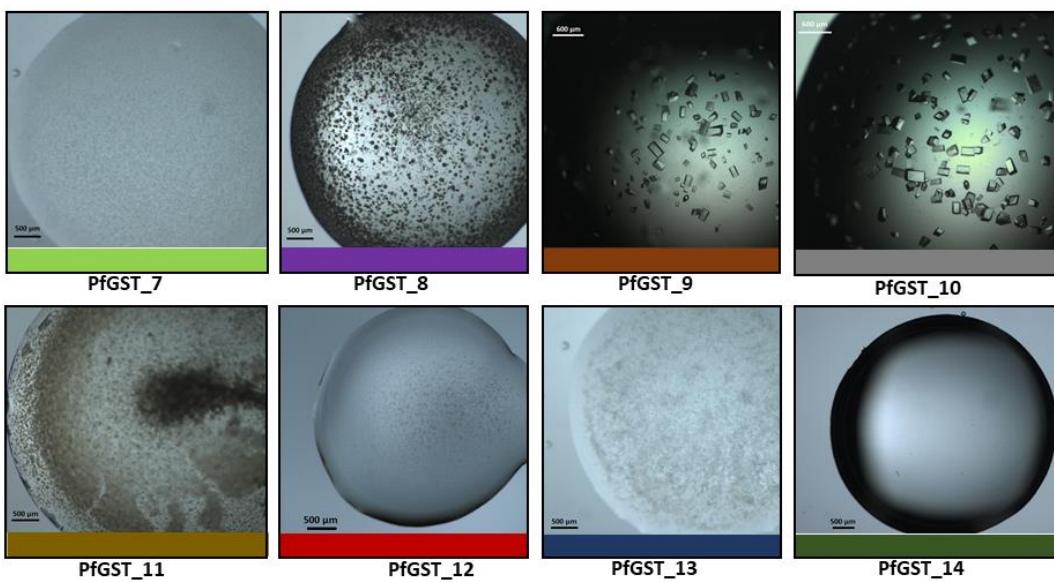
Figure III-18: Crystallization outcome for the PfGST experiments (PfGST_7 – PfGST_14) described by the phase diagram

The pictures show the final outcome of the crystallization experiments used in the experimental phase diagram; **PfGST_7**: tetrameric microcrystals obtained during induced evaporation of the crystallization droplet; **PfGST_8**: protein spherulites in solution among precipitated protein; **PfGST_9**: tetrameric crystals observed 48h after precipitant addition; **PfGST_10**: tetrameric crystals observed 48h after precipitant addition; **PfGST_11**: protein spherulites among liquid-liquid phase separation; **PfGST_12**: soft protein aggregation in solution; **PfGST_13**: tetrameric microcrystals among a small layer of precipitated protein caused by induced droplet evaporation; **PfGST_14**: clear droplet of supersaturated protein in solution.

According to the crystallization stages drawn for the PfGST phase diagram, the experiment PfGST_8 could be localized in the nucleation area during precipitant addition and sample evaporation. Most probably nucleation did occur, since the R_h distribution map showed a typical crystallization trend.

However, it might be that the step of droplet evaporation disrupted the necessary equilibrium for crystal growth and as a consequence, the droplet followed liquid-liquid phase separation. These findings fall into the theory previously cited, where nucleation is believed to occur in the absence of liquid – liquid phase separation. Albeit, changing the crystallization direction towards a phase separation process can cause a denaturation of the pre-existing nuclei. It was also found that at a slightly higher precipitant concentration (1.33 mol $\cdot$ L^{-1}), the outcome was a heavy phase separation as seen for PfGST_11. The existing data suggest that the phenomenon of liquid-liquid phase separation in the form of protein spherulites or heavy protein

gelation happens at a medium precipitant concentration (approx. 1 mol · L^{-1}) and relatively high protein concentration (approx. 7 mg · mL^{-1}).

To obtain protein crystals of approx. 200 µm that can be used for conventional crystallography, a regular step of precipitant addition for inducing nucleation and further crystal growth sufficed. However, the formation of PfGST microcrystals was only observed to happen at induced nucleation via evaporation. For both experiments PfGST_7 and PfGST_13 the appearance of microcrystals could be observed immediately after the precipitant concentration in the droplet was increased by evaporation to 2 mol · L^{-1}. The DLS measurements also showed the rapid formation of a radius band at approx. 1000 – 2000 nm. Since the formation of nuclei is formed on behalf of the protein solution while its concentration increases via evaporation, the availability of protein in solution becomes poor; crystals cannot grow any further, but they do stabilize in solution as microcrystals.

III.4. Conclusions

The XtalController900 provided valuable insights about the nucleation and crystal growth phases during protein crystallization. The *in situ* DLS measurements provided information about the evolution of the crystallization droplets showing particularities in the radius distribution maps, as the protein molecules developed upon precipitant addition. The formation of a second hydrodynamic radius fraction at approx. 100 – 1000 nm could be detected for all proteins undergoing crystallization, while the further investigation with the use of the SpectroLight600 confirmed the final formation of this particular radii fraction into protein crystals. The size and development of this fraction is in good agreement with other studies where particles of similar sizes (from one hundred to several hundred nanometers) were identified as crystallization precursors for some specific proteins (Gliko et al. 2007; Pan et al. 2007; Oleg Galkin et al. 2007).

The characterization of the crystallization droplets with Cryo – Electron Microscopy suggested that for thaumatin, the second band at approx. 200 – 500 nm represents

a two-step mechanism of nucleation where the formation of nuclei is preceded by protein clusters with low internal order. Two different assemblies of the clusters could be identified: geometrically ordered shapes with well-defined edges and a population of objects with rough edges merging into well-defined clusters. Based on previous results (Schubert et al. 2017) this population might show the transition of protein clusters into crystal nuclei by a consequent growth with internal order. Further images of crystallization droplets showed crystalline entities for thaumatin and SP. The size of the crystals was in perfect agreement with the radii fraction above 1000 nm, showing how the particle distribution revealed by these two proteins during crystallization is due to microcrystal formation.

The data obtained from protein crystallization could be used to obtain a better understanding of the crystallization pathway by mapping its course as an experimental phase diagram. By keeping several parameters constant that could influence the crystallization kinetics, it was found that ionic strength and protein concentration are the main variables that influence the crystallization outcome. The use of induced evaporation proved to be very helpful when aiming for protein microcrystallization. If a sample is still undersaturated after precipitant addition this can be changed by controlled sample evaporation. This will promote an increase in the protein and precipitant concentration, pushing towards supersaturation. With the use of DLS the evaporation can be precisely manipulated until a specific change in the hydrodynamic radius size is observed. This change implied the formation of a second radius band between approx. 100 – 1000 nm, showing the presence of crystallization precursors, while the protein was slowly consumed.

Chapter IV

Optimization and production of protein microcrystals for SMX

IV.1. Sample preparation for serial diffraction data collection

The growth of protein micro- and nanocrystals has become an active area of research for the protein crystallography community, especially with the continuous development of serial data collection. The high peak brilliance of FELs and third generation synchrotron radiation sources have allowed the achievement of successful results using far smaller crystals in suspension than applied for conventional, rotational methods (Kirian et al. 2011; Chapman et al. 2011; Kupitz, Basu, et al. 2014). As a result, the production of protein micro- and nanocrystals has become of high relevance, posing a high demand on the preparation of such crystalline suspensions for serial data collection (Schubert et al. 2015; Lee et al. 2018). Due to the small crystal size range that can be used for successful data collection at X-ray free electron lasers (XFELs) or at synchrotron radiation sources, as well as the limited availability of experimental beamtime, sample characterization prior to data collection is most essential. The common techniques for characterizing protein crystals in suspension until now have been electron microscopy and X-ray powder diffraction (Nederlof et al. 2013; Stevenson et al. 2014). Since the amount of crystalline material has to be large (a few microliters) adaptations have been made for the most common crystallization methods with the aim of producing bulk amounts of crystals in the small nano- and micrometre range.

The batch method harbours a fast mixing of highly concentrated protein and precipitant solutions, thus forcing the sample solution to a highly supersaturated phase where nanocrystallization might be favoured (Schlichting 2015a). This represents a fast crystallization approach, but equally, the crystallization conditions must be precisely established beforehand. Other methods include crushing large protein crystals to obtain a nanocrystalline suspension which can then be used for serial data collection (Stevenson, Makhov, et al. 2014). The sensitive part about such

techniques is that the outcomes might result in decreased diffraction quality, as hard mechanical forces can deteriorate and lower the internal order of the crystals. Nanocrystallization based on free interface diffusion is also an available alternative, where protein solution is added in small amounts to a highly concentrated precipitant solution (Kupitz, Grotjohann, et al. 2014). Among all of these techniques, the most efficient methods until now appear to be the batch crystallization and more innovative manipulative techniques using vapor-diffusion methods in sitting drops (Meyer et al. 2012).

When preparing crystals for data collection at an XFEL or at a synchrotron for serial crystallography studies, there are a series of sample requirements prior to data acquisition. One important aspect is that the crystals need to be of a reasonable size, from a few hundred nanometres to a few micrometres. Equally, the crystals in the sample need to be homogeneous in size, in order to prevent clogging of the sample delivery system or damage to the X-ray detector (Chavas, Gumprecht, and Chapman 2015).

In order to prepare bulk amounts of protein microcrystals for serial crystallography studies, two main approaches were used. At first, crystallization experiments were performed with the XtalController900 to track the evolution and formation of microcrystals. The degree of crystallinity of the resulting micro- and nanocrystals were tested by X-ray powder diffraction. Since the maximum volume that can be obtained during controlled crystallization is about 100 µL, the crystallization conditions had to be further optimized and scaled up. Based on the information achieved from the hydrodynamic radius distribution plots and from the experimental phase diagrams, the batch method was further used to optimize the size of the crystals, the concentration of the crystals in solution, as well as the final volume of the crystal suspension.

IV.2. Materials and methods

IV.2.1. Sample preparation

The sample used for microcrystallization experiments was thaumatin from *T. daniellii* that was prepared to a final concentration of 45 mg ·mL^{-1} as previously described in chapter II.2.1. Prior to crystallization, the sample was centrifuged at 16,100 x *g* for 15 minutes in order to remove any unwanted aggregates or impurities present in the sample. The protein buffer and precipitant solutions used for the XtalController900 experiments and for the batch optimization are summarized in Table 9.

Table 9: Crystallization stock solutions used for thaumatin microcrystallization

Crystallization Solutions	Composition
Precipitant stock 1	0.5 M sodium tartrate
Precipitant stock 2	1.2 M sodium tartrate
Precipitant stock 3	1.8 M sodium tartrate
Washing buffer	1.025 M sodium tartrate, 12.5 mM Bis-Tris

IV.2.2. Protein microcrystallization using the XtalController900

The XtalController900 previously described in chapter. II.2.2 was used to follow the evolution of the hydrodynamic radius distribution of the thaumatin molecules during crystallization. The environmental conditions were set to 20°C and the relative humidity to approximately 80% and both were kept constant for the entire duration of the crystallization experiments.

IV.2.3. X-ray powder diffraction for sample quality

The experimental samples obtained from the XtalController900 were checked for crystallinity applying X-ray powder diffraction (XRPD) in capillaries. At the end of each crystallization experiment the droplet was collected with a pipette from the

coverslip and transferred into a quartz capillary with a 0.5 µm inner diameter (HR6–108, Hampton Research, Aliso Viejo, USA). In order to obtain a dense crystal pellet from the suspension, the capillary was centrifuged for 10 minutes at 2400 x g. The supernatant was then removed and the pellet was prevented from drying by applying wax to the open end of the rod (Harvard Dental International). The capillary was mounted onto a 3-axis goniometer and the length and position of the crystalline pellet were aligned to the X-ray beam intercept. The XRPD images were collected at the IµS home source from *Incoatec*, with a sample to detector distance of 250 mm. Each sample was exposed for a period of 900 seconds, while rotating each capillary through 10°.

IV.3. Results and discussions

IV.3.1. Production of nano- and microcrystals using the XtalController900

The results obtained by the *in situ* DLS hydrodynamic radius plots (Figure IV-1 and Figure IV-2) for the two thaumatin experiments show a detailed evolution of particles in solution during crystallization. Prior to precipitant addition the protein exhibited a monodispersed behaviour in solution with a constant particle size of approximately 2.5 nm. As precipitant was added, the droplet developed two main hydrodynamic radii fractions in solution and did not develop further once the droplet entered the constant mode. For experiment THM_10, after the precipitant reached the final sodium tartrate concentration of 0.9 mol · L^{-1} in solution, the sample showed a very broad band of particle radii between approx. 100 and 1000 nm that further increased to 2000 nm. Over time the appearance of a few crystals could be detected with the CCD camera and 35 hours later the crystals grew to their final size (Figure III-15). As learned from previous experiments, it was assumed that the fraction at approx. 100 nm is attributed to crystal nuclei, whereas the fraction at 2000 nm could be crystalline. However, due to their size, the crystallinity degree of the fraction at 2000 nm could not be appreciated under an optical microscope. The most interesting aspect is that this fraction remained constant in solution over a period of 34 hours, without being consumed by the formation of the larger crystals. In order to probe

the quality of this material, the sample was removed from the coverslip and investigated by X-ray powder diffraction. Since the main focus was on the fraction that cannot be investigated by optical microscopy, the crystals that could be detected by the microscope were removed from the droplet prior to data collection. The XRPD image showed that the material was indeed crystalline, giving powder diffraction rings up to a resolution of 5.6 Å (Figure IV-3 A). In comparison to the higher resolution obtained earlier for THM_11, for this experiment the crystals size was too small for the inset energy, causing a loss in resolution. Nevertheless, the resolution rings that were obtained were enough to correlate the large radii distribution to the presence of sub-microscopic crystals.
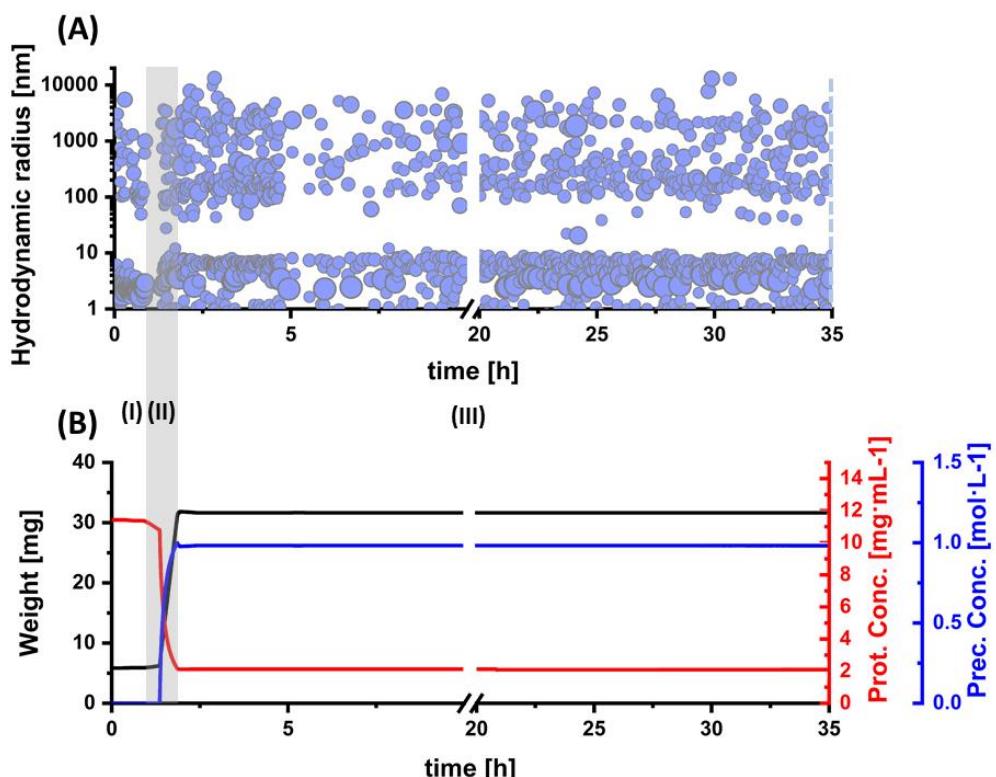
Figure IV-1: THM_10 (A) Radius Distribution and (B) Parameter plots derived from the XTC900
The balance plots and the R_h distribution are divided into three main areas: (I) Initial DLS measurements to assess the quality of the protein prior to crystallization; (II) Main precipitant addition phase highlighted in grey showing the monitoring of protein particles in solution (A) as the precipitant concentration in the crystallization droplet increases over time (blue line); (III) The droplet conditions are kept constant while monitoring the evolution of the R_h distribution over time. The cut line of 5 hours during sample monitoring was done in order to highlight the balance plots and the R_h distribution prior as well as during precipitant concentration.
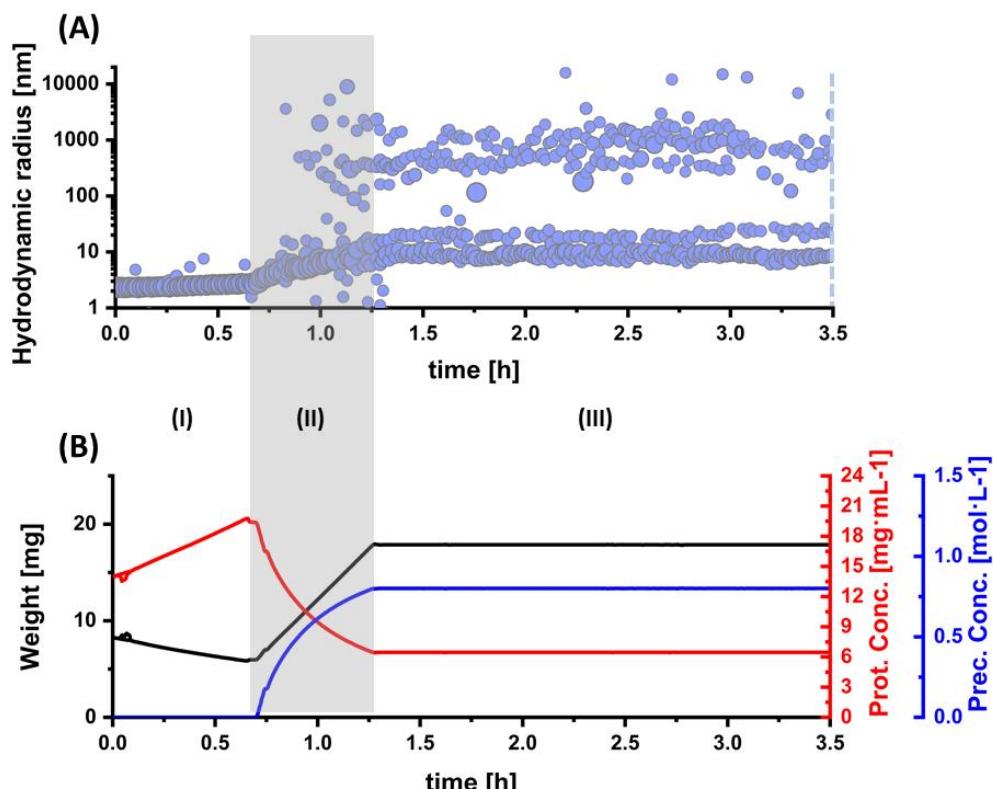
Figure IV-2: THM_11 (A) Radius Distribution and (B) Parameter plots derived from the XTC900
The balance plots and the R_h distribution are divided into three main areas: (I) DLS measurements during controlled evaporation for increasing the protein concentration (red line); (II) Main precipitant addition phase highlighted in grey showing the monitoring of the particles in solution (A) as the precipitant concentration in the crystallization droplet increases over time (blue line); (III) The droplet conditions are kept constant while monitoring the evolution of the R_h distribution over time.

The second experiment performed (THM_11) was at first initiated with an induced evaporation in order to obtain a protein concentration of 19 mg $\cdot$ mL^{-1} prior to precipitant addition. According to the radius plot the sample entered the supersaturation area immediately after initiating the precipitant addition. When the sodium tartrate concentration reached 0.6 mol$\cdot$L^{-1}, the nucleation phenomena could be identified and tracked. As more nuclei formed, the initial crystalline entities grew in size, reaching a radius distribution between approx. 800-1300 nm (Figure IV-2). It can be concluded that at this stage, the crystal nuclei advanced into the crystal growth stage, while the protein fraction gradually became poorer as microcrystals formed. With the increase in precipitant concentration, the formation of small particles could be easily identified as the radius distribution increased towards 500-1500 nm. The particles became visible at a precipitant concentration of 0.7 mol $\cdot$ L^{-1}

and by the end of precipitant addition (0.9 mol · L^{-1}) the droplet was saturated with small visible objects (Figure III-15). Due to their small size, no preliminary conclusions could be drawn regarding the internal order of these particles. To identify crystallinity, the sample was subjected to powder diffraction data experiments. The XRPD image confirmed that the investigated particles were crystalline, diffracting to a resolution of approx. 2.8 Å. The strong powder diffraction rings confirmed that the droplet contained microcrystals.
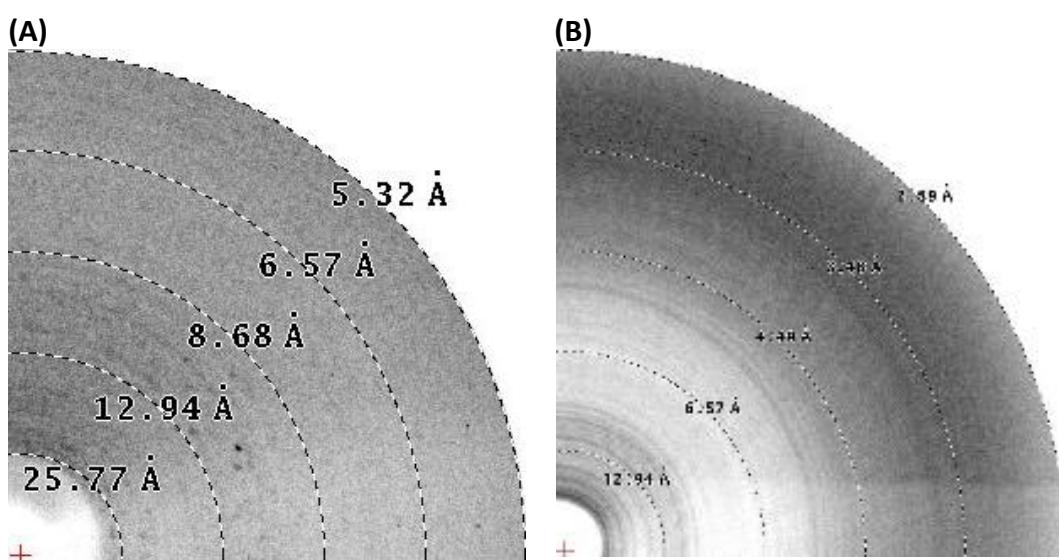
Figure IV-3: X-ray powder diffraction images collected for THM_10 and THM_ 11
(A**) THM_10** XRPD showing a maximal resolution of 5.3 Å and (B) **THM_11** XRPD showing a maximal resolution of 2.8 Å.

IV.3.2. In batch optimization for larger volumes of microcrystal suspensions

A volume of 1 mL thaumatin solution (45 mg · mL^{-1}) was added to an equal volume of sodium tartrate stock of 0.5 mol · L^{-1}. The protein and precipitant solutions were allowed to contact for a period of 10 minutes in order to achieve supersaturation. In a second step, a volume of 2 mL sodium tartrate stock of 1.8 mol · L^{-1} was added to the tube. Considering the previously discussed data (chapter III.3.4 and chapter IV.3.1.), by using such a high stock concentration it was expected that a fast nucleation would commence immediately after precipitant addition.

In order to obtain a high number of nucleation events and to assure homogeneity, the solution was gently mixed with the use of a pipette for a period of 2 minutes until the appearance of the solution became cloudy. Since this indicated that the nucleation was initiated at a high rate with success, the solution was allowed to rest for 30 minutes in order for crystal growth to proceed. Afterwards, an aliquot was taken and analyzed under a high-resolution optical microscope.

The solution indeed showed a high amount of crystals in solution (Figure IV-4). However, small aggregates could be also distinguished among crystals. The reason for this might be the high concentration of stock precipitant solution causing a partial protein aggregation. Since DLS measurements could not be performed in the crystallization tube, it could not be verified whether the crystals had already reached the final state or if they would continue to grow at the expense of the protein available in solution. Therefore, in order to eliminate the possibility of Oswald ripening and to ensure that the crystals will remain of similar size, the suspension was further washed in order to remove any protein available in solution. At first, the crystals were centrifuged at 200 x g for 5 minutes in order to settle the crystals into a pellet. The supernatant was then collected and replaced with 5 mL washing buffer (Table 9). The new mother liquor was gently mixed until all of the crystals were resuspended. The washing procedure was performed 3 times in order to remove all aggregates or soluble protein that could support the further growth of thaumatin crystals. At the end of the crystal washing steps, an aliquot was inspected under the microscope. The thaumatin crystals looked intact, and most importantly, the aggregates previously detected were mostly removed. To ensure that the new mother liquor would not result in long-term damage to the crystals, 24 hours later another aliquot was inspected under the microscope. The crystals remained of the same size and no visual crystal damage could be observed.
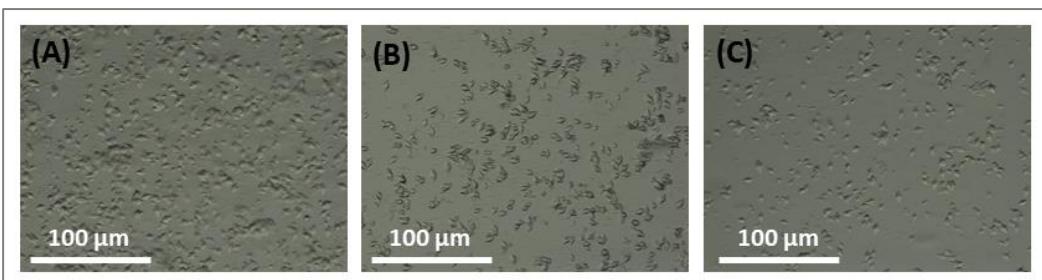
Figure IV-4: Thaumatin microcrystals
(A) after crystal growth, (B) after washing with washing buffer and (C) after 24h hours

IV.4. Conclusions

The results obtained for thaumatin crystallization with the Xtalcontroller900 indicate that the concentration of protein and precipitant addition are important parameters in the formation of crystals and their dimensions. When specific parameters were kept constant, such as temperature and pH, precipitant stock solution and time of addition, an increase in the ionic strength resulted in the formation of nanocrystals. This was indicated by the large hydrodynamic radii that formed above 200 nm and continued growing as crystal growth followed. These statements are supported by previous DLS studies on protein crystallization where an increase in precipitant solution resulted in microcrystals (Juarez-Martines et al. 2001). By performing crystallization at high protein concentrations, a high supersaturation was achieved resulting in spontaneous nucleation. This lead to the formation of a very large number of nuclei growing further, while shifting the equilibrium towards the formation of microcrystals. The crystallinity of both thaumatin droplets was cross-checked with X-ray powder diffraction and the presence of nano- and microcrystals giving diffraction rings to a maximal resolution of 2.8 Å, was confirmed.

Sample optimization for serial crystallography implied using the crystallization batch approach for preparing protein microcrystals in high amounts, that will suffice for a serial data collection approach. For this method, thaumatin at high concentration and high ionic strength (sodium tartrate) at room temperature lead to the production of microcrystals. In order to ensure equilibrium, the crystals were washed

with buffer solution. The action implied removing any soluble protein left in solution that could support Ostwald ripening or further crystal growth. The final batch of crystal suspension resulted in thaumatin microcrystals with a size of 10 to 20 μm and a crystal concentration of 40 to 50% in solution.

Chapter V

X-ray crystallography – data collection and analysis

V.1. Serial Femtosecond and Millisecond Crystallography

Serial femtosecond crystallography (SFX) at X-ray free electron laser facilities has been a revolutionary alternative for solving challenging protein structures while operating at room temperature under the principle of diffraction before destruction (Neutze et al. 2000). The use of ultra-short X-ray pulses allows the collection of diffraction data before the onset of significant radiation damage from small crystals (Chapman, H. N., et al. 2011; Tenboer et al. 2014). Due to the nature of the data collected (thousands of images where one diffraction image is collected from a single crystal in a random orientation) along with the intensity and the photon energy of the pulses, new programs for data analysis have been developed along as well (Foucar et al. 2012; White et al. 2012b; Kern et al. 2013, 2014; Kabsch 2014; Kirian et al. 2010). The serial collection of data has also been adapted for use at synchrotron radiation sources. Despite not having ultrashort pulses, the data can be collected with limited radiation damage due to the single X-ray exposure per crystal. (Sabine Botha et al. 2015; Beyerlein et al. 2017a; Weinert et al. 2017; S Botha et al. 2018).

However, regardless of the source or method employed for collecting diffraction data, the collected diffraction patterns contain only information about the amplitudes of the complex structure factors, missing information about the phase angle. Without the phase information the Fourier transformation for the electron density map calculation is not possible. This is known as the phase problem, where the phases of the recorded amplitudes must be known for solving the structure. At the moment, molecular replacement (MR) is the most commonly used choice, where a homologous protein structure is used to calculate the missing phase information (Rossmann 1990). Although the PDB database is more than generous nowadays, many biological samples do not have a homolog to substitute the missing phases. Depending on the light source and data available, experimental phasing methods

have been developed for protein phase retrieval. In a simplified manner, the phasing methods can be divided into two categories: those relying on the differences between crystals, such as single or multiple isomorphous replacement (SIR / MIR) and methods working with anomalous differences, such as single or multiple wavelength anomalous diffraction (SAD / MAD) (Hendrickson, Horton, and LeMaster 1990). Furthermore, methods from both categories can be combined where the phase information can be obtained by single or multiple isomorphous replacement combined with anomalous scattering (SIRAS / MIRAS). There are also particular methods, such as native SAD, where the phases can be retrieved from native sulfurs inherent to the protein, that can give anomalous scattering signal to determine the initial protein substructure without the need of introducing heavy atoms into the crystals.

The most effective method used for *de novo* structure determination, when working with serial data, has been proven to be SIRAS (single isomorphous replacement with anomalous scattering). The first structure obtained by *de novo* phasing using single-wavelength anomalous data collected at a FEL was published in 2014 where a lysozyme gadolinium derivative was used (Barends et al. 2014). Another successful SIRAS phasing was reported using an isomorphous mercury derivative where 200,000 images were used for structure determination (Yamashita et al. 2015). Shortly after, it has been reported that a number of 12,000 images could suffice for the structure determination of bacteriorhodopsin using a heavy atom derivative (Nakane et al., 2016).

The first time SIRAS was attempted with serial data collected at a synchrotron facility it failed and only MIRAS could be done applying lysozyme crystals using an iodine and gold derivative (Sabine Botha et al. 2015). Recently, native SAD phasing with serial data was proven to be possible with synchrotron radiation where a g-coupled protein receptor was solved using 186,688 serially collected images (Weinert et al. 2017). In a most recent study the gap between MIRAS and SAD was bridged by successfully applying SIRAS to a mercury derivative of proteinase K. A number of only 12,000 images were successfully contributed towards obtaining the structure using serially collected synchrotron data (Botha et al. 2018). It has been

reported that by using serially collected data at a FEL, SAD phasing was successful with 70,000 images and native SAD for thaumatin with 125,000 images (Nass et al. 2016). However, another recent study reported that for lysozyme 80,000 patterns were enough for SAD phasing (Yamashita et al. 2017).

For structure determination based on anomalous scattering, SAD requires only a single set of diffraction data. Native SAD phasing uses the intrinsic anomalous scattering signal present in a macromolecule using single wavelength diffraction. The use of specific atoms such as sulfur as an endogenous source of anomalous signal (phasing markers in the cysteine and methionine residues) circumvents the problems related to the use of heavy atoms derivatives. Usually, the challenge arises from accurately measuring the very small anomalous signal. The most suitable way to maximize the emitted anomalous signal is by taking multiple measurements of every unique reflection at an appropriate wavelength. This will provide a high multiplicity and therefore, accurate information of the anomalous difference (Taylor 2010).

All of the serial millisecond crystallography approaches aforementioned used a free standing lipidic cubic phase (LCP) stream to deliver the crystals into the X-ray interaction region. Although LCP is known to be an ideal delivery medium due to its compatibility with both soluble and membrane protein (Huang et al. 2015; Weierstall et al. 2014; Sabine Botha et al. 2015), the mixing procedure can sometimes result in crystal crushing that cannot be accounted for during data collection. Furthermore, due to the high viscosity of LCP large pressures are required for extruding the crystal containing sample (Conrad et al. 2015; Johansson et al. 2012). Moreover, crystals from soluble proteins tend to be more stable in their native solutions and a change in the carrier medium can sometimes result in reduced diffraction quality. Therefore, new methods such as the tape drive approach is a friendly manner for serial data collection of microcrystals at synchrotrons. The tape drive data collection method (Beyerlein et al. 2017a) for SMX has already proven to be successful for mix and diffuse studies of ligand binding, where the diffusion of ligand molecules into the protein could be observed 2 seconds after solution mixing with the protein microcrystals. This new approach for time resolved serial data collection at synchrotrons shows a high potential for structural

enzymology and high-throughput drug screening and moreover, the method could also be used for further structure determination using *de novo* phasing methods such as SIRAS or native SAD.

V.2. Materials and methods

V.2.1. Sample preparation and crystallization

The proteins PfGST from *Plasmodium Falciparum* and thaumatin from *Thaumatococcus daniellii* used for crystallization experiments were prepared as outlined in chapter II.2.1. For PfGST the crystallization was carried out as described in chapter III, whereas the protein crystallization and optimization protocol for thaumatin was discussed before in chapter IV.

V.2.2. Conventional data collection at a synchrotron radiation source

The PfGST crystals previously obtained and discussed in chapter III were further analyzed to obtain information about crystal quality and internal order by X-ray diffraction at the synchrotron facility PETRA III, DESY. Diffraction data were collected at the EMBL beamline P13 in Hamburg (Cianci et al. 2017) in April 2018.

Prior to data collection, the crystals were treated with a cryoprotectant solution comprising of the crystals mother liquor solution supplemented with 15% glycerol. Each crystal was first mounted onto a cryo-nylon loop (Mounted CryoLoop, Hampton Research, US). The cryoprotection was performed by soaking the crystals into the aforementioned solution for a period of 30 seconds followed by flash freezing in liquid nitrogen.

All X-ray diffraction measurements were conducted using a cryogenic stream to prevent high radiation damage. The diffraction images were collected using a crystal to detector distance of 293 mm and a wavelength of 0.991 Å. A total number of 2200 images were collected using a rotation range per image of 0.1° and an exposure time

of 4 milliseconds. The maximum resolution obtained during data collection was found to be between 1.5 and 1.62 Å.

V.2.3. Serial data collection using synchrotron radiation

Thaumatin microcrystals obtained by optimized batch crystallization were applied for serial millisecond crystallography data collection at PETRA III, DESY (Hamburg, Germany). The crystal delivery approach was done using the previously mentioned tape-drive method. The instrumentation was installed by Dr. Dominik Oberthür and his team, together with engineers and beamline scientists of the P11 beamline at PETRA III, DESY. The beamtime for collecting the serial data of thaumatin crystals was kindly provided by Dr. Dominik Oberthür and his group.

The main principle of the tape-drive method follows the technique of a classic cassette player, where with the use of a rotor, a tape is moved from one side to the other. For protein crystallography, a matt polyimide film, also known as Kapton® tape (DuPont 2013), was used as the tape carrier material for the protein crystal suspension, since it produces low background scattering and minimal X-ray absorption. With the use of a rotor, the tape (width of 6 mm and thickness of 12 μm) is pulled from the initial supply roll towards the second roll which collects the drawn and used tape (Figure V-1). The crystal suspension is delivered to the tape with the use of a fused silica capillary that is placed vertically next to the feeder roll, depositing crystals in suspension onto the tape, as it is transported away from the first roll. The crystal suspension is then moved along with the tape in a thin layer oriented towards the detector side, directly passing through the X-ray interaction region.

For this setup, a rotating beam chopper was placed at P11 after the focusing optics in the X-ray path, in order to generate X-ray pulses with a duration of 4.03 ms and a repetition rate of 25 Hz. The detector used at P11 at this time was a Pilatus 6M that was synchronized with the X-ray pulses in order to collect a single diffraction pattern per X-ray exposure. The purpose of the chopper is to ensure that the crystalline sample is prevented from any induced boiling that could raise from the intense X-ray

exposure. Furthermore, this also allows establishing control over the exact exposure time of crystals.
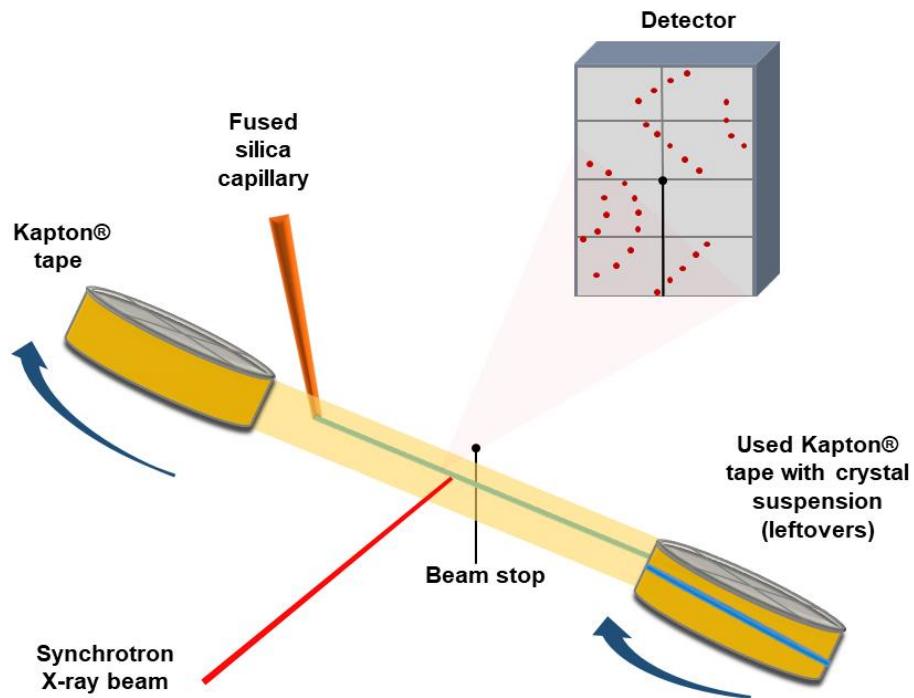
Figure V-1: Experimental setup for serial data collection using the tape drive method
Experimental setup for serial data collection of the native thaumatin microcrystals using the tape drive approach. The crystal suspension is delivered to the polyimide tape through a fused silica capillary of 100 µm inner diameter. The tape is then moving from the initial roll to the collector roll, passing the crystals through the interaction region of the X-ray beam.

Tape-drive Data Collection for thaumatin microcrystals

Native thaumatin microcrystals suspension was transported using a hydraulic setup based on air pressuring the sample reservoir and thus expelling sample. The tape speed was set to 1 mm per second and a sample flow rate of 0.2 to 3 µL per second was used during data collection, depending on the abundance of crystals in solution at a certain time interval. At most synchrotrons, radiation is tunable within an energy range of 6 to 12 KeV. Since the sulfur K edge is found at 2.47 KeV corresponding to a wavelength of 5.015 Å and the anomalous scattering signal defined by $\Delta f''$ ranges

from 0.13 to 0.95 e⁻ (Figure V-2), data should be collected at a photon energy as close to the edge as possible (Merritt 2012). The ultimate choice, however, will be a result of a compromise between the limitations set by beamline transmission and optics, the resolution of the diffraction data and also photon absorption of the anomalous atoms. Studies have shown that when measuring at an energy far from the sulfur K edge, data collected at 6 keV and 7 KeV provided enough phase information for structure retrieval applying native Sulphur SAD (Mueller et al. 2015; Nakane et al. 2016; Nass et al. 2016).
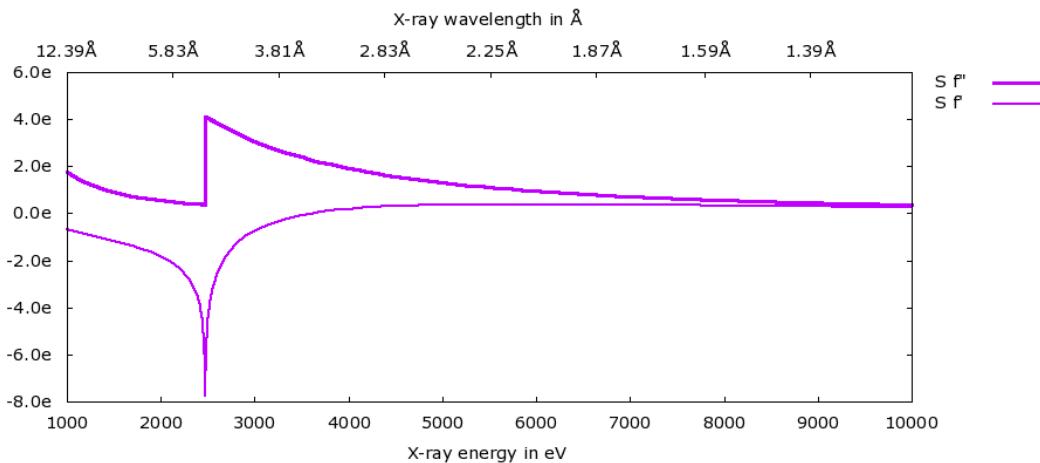
Figure V-2: X-ray anomalous scattering plot for sulfur

The SAD data for thaumatin were collected at beamline P11, PETRA III DESY in April 2018, at room temperature, using a photon energy of 7.0 keV (1.76 Å wavelength), with a measured photon flux of approximately $3.7 \cdot 10^{12}$ photons per second at 100% transmission.

The amount of used crystal solution was 5 mL containing approx. 50% settled crystal pellet, resulting in a hit rate of 30 to 40% for the majority of the time. However, due to their size, the crystals started to settle after approximately 40 minutes. This could be noticed by a steady drop in hit rate. After resuspending the sample by gentle shaking, the hit rate recovered to 30 to 40%. Pressures of 100 to 600 mbar were used

in order to achieve flow rates between 0.5 and 3 µL per minute. The crystal to detector distance was set to 200 mm and in total, a number of 842,796 images were collected in 12 hours.

V.3. Results and discussions

V.3.1. PfGST structure solution

A complete data set to 1.6 Å resolution was collected and further analyzed for crystal quality. The diffraction data were integrated and processed using the XDS program package (Kabsch 2010), with the data statistics summarized in Table 10. The scaling was done using the AIMLESS program of the CCP4 software package (Winn et al. 2011; Evans and Murshudov 2013). The structure solution and refinement were achieved using the software package Phenix (Adams et al. 2010). The PDB search model 3FR6 (Burmeister et al. 2003) corresponding to PfGST crystallization using similar sample expression and purification protocols, was used for structure validation.

The spatial resolution was cut at 1.8 Å in order to achieve 96 % completeness for the high-resolution shell. Since the aim of this study was crystal quality, no manual refinement was used for structure improvement. Hence, if aiming for optimal structure determination, performing manual refinement is necessary and will most probably bring these parameters to optimal values.

The resulting PDB file and the PDB model 3FR6 were further superimposed using an algorithm for sequence-order independent protein structure alignment (Y. Zhang and Skolnick 2005). The similarity of the two compared PDBs is then evaluated by means of optimal superimposition and a TM-score value representing the structural similarity. The resulting TM-alignment (Figure V-3) gave a score of 0.98 being very close to a perfect match which is represented by a value of 1.0. When inspecting the superimposition using Pymol (Schrödinger 2015), some slight variations could be detected, as seen in Figure V-3. These displacements are caused by loops regions which are usually formed by protein flexibility.

Table 10: Summary of data statistics for the PfGST data collection
Statistics showing values for data collection, data processing and refinement scores for the PfGST structure determination using molecular replacement. The values between brackets refer to the highest resolution shell.

	PfGST data
Wavelength (Ã)	0.98
Resolution range (Ã)	74.87 - 1.88 (1.947 - 1.88)
Space group	P 21 21 2
Unit cell parameters	
a / b / c (A°)	62.17 / 74.87 / 88.42
α / β / γ (°)	90.00 / 90.00 / 90.00
Total reflections	133199 (13928)
Unique reflections	33035 (3342)
Multiplicity	4.0 (4.2)
Completeness (%)	96.38 (98.84)
Mean I/sigma(I)	19.32 (4.64)
Wilson B-factor	30.4
R-merge	0.03 (0.19)
R-meas	0.04386
CC1/2	0.999 (0.968)
CC*	1 (0.992)
R-work	0.19 (0.2)
R-free	0.23 (0.31)
Number of non-hydrogen atoms	3593
Protein atoms	3254
water	339
Protein residues	404
R.m.s.d., bonds (Å)	0.005
R.m.s.d., angles (°)	0.80
Ramachandran favoured (%)	98
Ramachandran outliers (%)	0.26
Clash score	3.04
Average B-factor	36.3
Macromolecules	35.8
Solvent	41.3
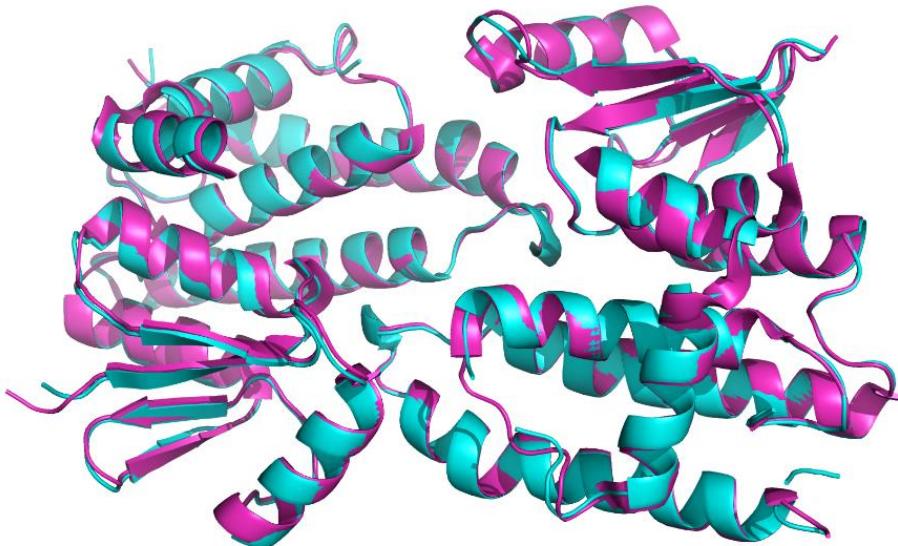
Figure V-3: Superimposition of the PfGST structure with the PDB model 3FR6 of PfGST

V.3.2. Thaumatin SMX data processing

The serially collected data using the tape drive approach were indexed and integrated using the CrystFEL suite of programs version 0.6.3. (White et al. 2012a). The indexing of all diffraction patterns from crystals in random orientations was performed using the programs dirax (Duisenberg 1992), mosflm (Powell, Johnson, and Leslie 2013) and XDS (Kabsch 2010). After optimizing the peak-finder parameters, a total number of 84,085 diffraction patterns were indexed as crystal hits, accounting for 10% of the total number of images collected.

For native SAD data collection, it is common practice to collect data sets of high multiplicity since the error associated with a measurement decreases with the square of the number of observations (Cianci, Helliwell, and Suzuki 2008; Weiss 2001). The multiplicity value for thaumatin in the current case was relatively high, as a result of serial data collection, which also increased the anomalous signal-to-noise level in the data. When collecting serial data, only a single diffraction pattern is collected per one crystal, significantly decreasing the radiation damage per crystal, which is currently a bottleneck for conventional crystallography native SAD. In general, data collection should be carried out under minimal radiation conditions,

since radiation damage is detrimental to the quality of the electron density maps and the quality of the final obtained structure (Diederichs and Karplus 1997). The data statistics of the thaumatin microcrystals are shown in Table 11.

The distribution of cell constants from all the crystals were well fitted by a Gaussian curve giving the correct space group and unit cell parameters for thaumatin crystals with unit cell parameters of a = 58.28 Å, b = 58.34 Å, c = 151.10 Å and α =β = γ = 90° (Figure V-4). Although some crystals diffracted to a maximal resolution of 1.8 Å with a completeness of 75%, the data was cut to 2.6 Å, in order to ensure 100% completeness. Although high resolution can provide further information, for SAD phasing it is important that the data are of high quality, even if this implies a slightly lower amount of data that is further used for structure determination.
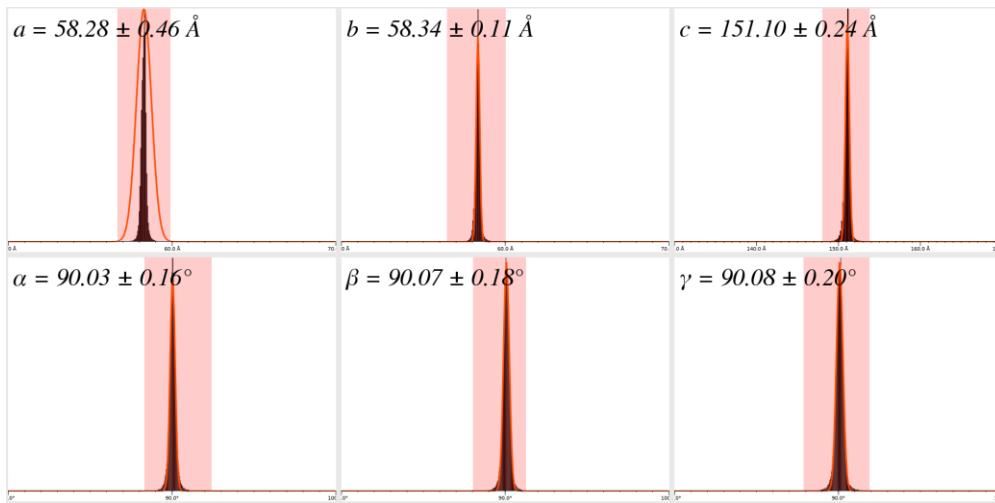
Figure V-4: Unit cell constant distributions of thaumatin obtained by CrystFEL
The unit cell parameters for all indexed patterns were calculated using the cell-explorer routine from CrystFEL. The histogram data were fitted with a Gaussian function returning the constants mean distribution along with the standard deviation.

Table 11: Summary of data statistics for the serially collected thaumatin data
Data statistics showing data for serial collection, data processing and refinement scores of the final determined structure using *de novo* sulfur phasing applying SAD. The values between brackets refer to the highest resolution shell.

	Native thaumatin
Wavelength (A°)	1.776
Resolution range (A°)	25.7 – 2.59 (2.66 – 2.59)
Space group	P41212
Unit-cell parameters **a / b / c (A°)** **α / β / γ (°)**	58.30 / 58.30 / 151.10 90.00 / 90.00 / 90.00
Total images collected	842,796
Total indexed hits	84,085
Total reflections	21,666148 (453009)
Unique reflections	15,545 (1,139)
Multiplicity	1,393.8 (397.7)
Completeness (%)	100 % (100%)
Mean I/σ(I)	15.58 (5.41)
Rsplit	5.37 (16.76)
CC*	0,999 (0.989)
CCano	0.0855 (0.266)
Radiation dose / crystal (MGy)	0.73
Rwork/Rfree	12.2/16.8
Total	1736
Protein	1659
Het atoms	10
Water	67
R.m.s.d., bonds (Å)	0.013
R.m.s.d., angles (°)	1.8
Coordinate error (Luzzati) (Å)	0.128
Ramachandran favoured (%)	96.39
Ramachandran outliers (%)	0
Wilson B-factor (Å²)	15.3
Overall	28.0
Protein	28.3
Solvent	28.8

V.3.3. Thaumatin native sulfur phasing

The processed diffraction data were prepared for native SAD phasing using SHELXD for substructure determination (location of sulfur heavy atom positions). Thaumatin contains 17 sulfur atoms with 8 cysteine and one methionine residues. The program found a total number of 20 marker atoms with occupancies of at least 25% and a CFOM of 38.8. In general, SHELXD tries to find more marker positions than requested in the input "find" file (Sheldrick 2010). Although the output sites in this case had 3 additional potential sulfur sites, a major drop in occupancy was detected after the 17th marker site, indicating the low possibility of these sites to be accurately present in the substructure (Figure V-5). Subsequently, 10 cycles of density modification and auto-tracing of the thaumatin backbone were done using SHELXE. The maximum correlation coefficient was found to be 15.7 for the inverse density map (the original hand was 12.9) and the maximum number of residues built for this hand was 93. Because the values obtained were very low, the positions of the marker sites were verified in Coot (Emsley et al. 2010) using the PDB model 5MJG (Meents, Oberthuer, and Srajer 2017). From all marker sites found by SHELXD only 4 marker positions could be identified as potential di-sulfide bonds, whereas the rest of the predicted Sulphur atoms were found in wrong positions.
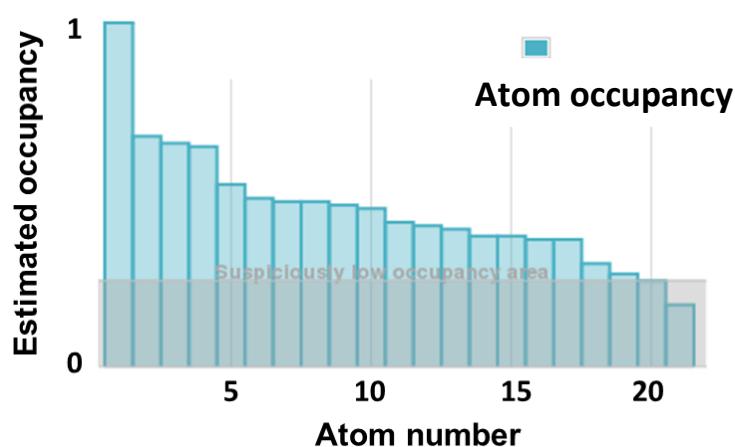
Figure V-5: Occupancy of Sulphur marker atoms found from the best search
The highlighted grey area indicates a suspiciously low occupancy for the identified positions.

The thaumatin PDB model was therefore used to identify the correct marker positions. After implementing the correct coordinates for the sulfur sites to the serially collected data, SAD phasing was proceeded using Parrot (Cowtan 2010), automatic refinement with refmac5 (Murshudov et al. 2011) and subsequent model building with buccaneer (K 2006). After 70 cycles of combined iterative model building with density modification and phase refinement the FOM was improved to 0.90 resulting in a built structure of 202 residues from the 207 amino acids present in the thaumatin structure. The results were inspected in Coot and the electron density at the C terminus indicated the absence of the last 5 residues (Phe, Cys, Pro, Thr and Ala). Using alternate cycles of manual refinement in Coot and automatic refinement with Refmac5 resulted in a structure with a final Rwork/Rfree of 12.2/12.7. The 2 Fo-Fc maps calculated for native SAD phasing and the final refined map contoured at 1σ are shown in Figure V-6.

An anomalous difference density map was calculated using the structure factors from the native data collection (all of the 84,085 indexed images collected along with the final refined phases). Figure V-7 shows the anomalous density around the sulfur marker sites contoured at 4.5σ. The anomalous difference density map shows clear anomalous density around all 8 di-sulfide bonds even when contoured at 4.5σ, the only exception being the sulfur from the methionine residue where no anomalous density could be observed at this value. Although typically a difference density map is contoured at 5.0σ, in the present case the value was chosen to optimally show the anomalous density surrounding the cysteine residues of the refined structure. These results are clearly showing that for a reasonable amount of serially collected data, the scattered anomalous signal of sulfur can provide enough information for applying *de novo* SAD phasing for structure determination in the future.
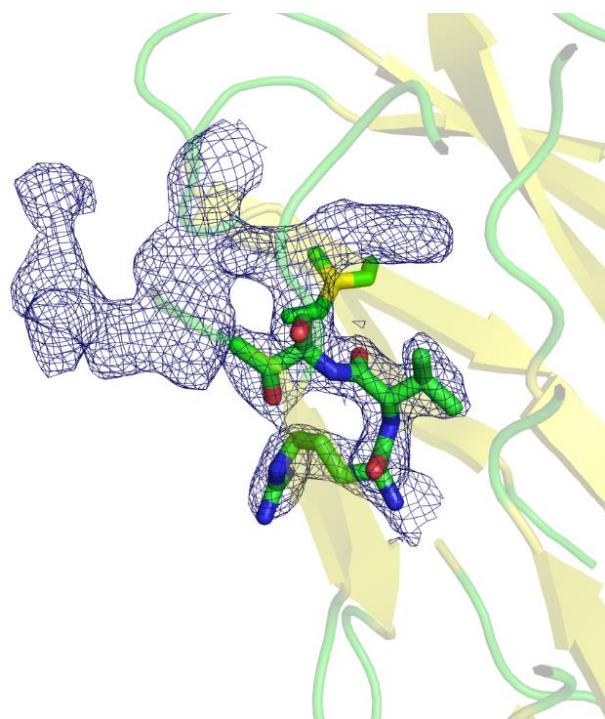
(A) Fo-Fc after model building
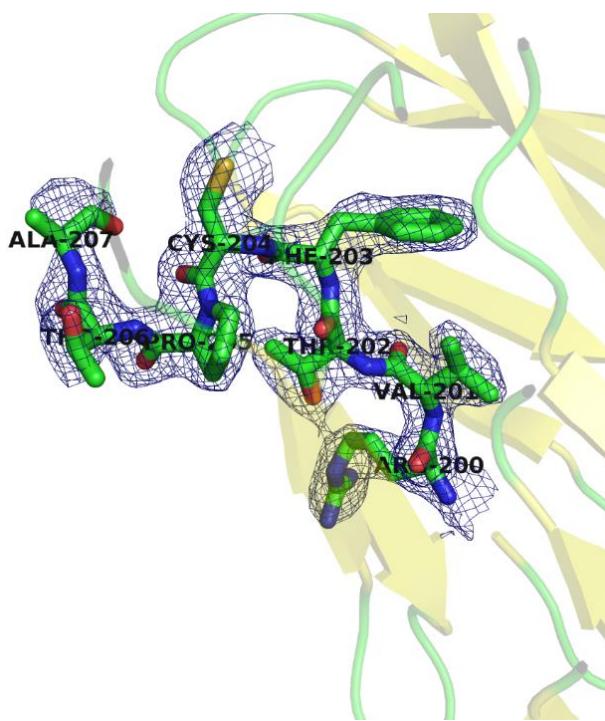
(b) Fo-Fc after final refinement

Figure V-6: Electron density maps at two different stages of the phasing process
The 2Fo-Fc electron density maps calculated using the phases after applying (a) iterative model building with phased refinement and (b) the final refined structure with the missing 5 remaining residues at the C terminus
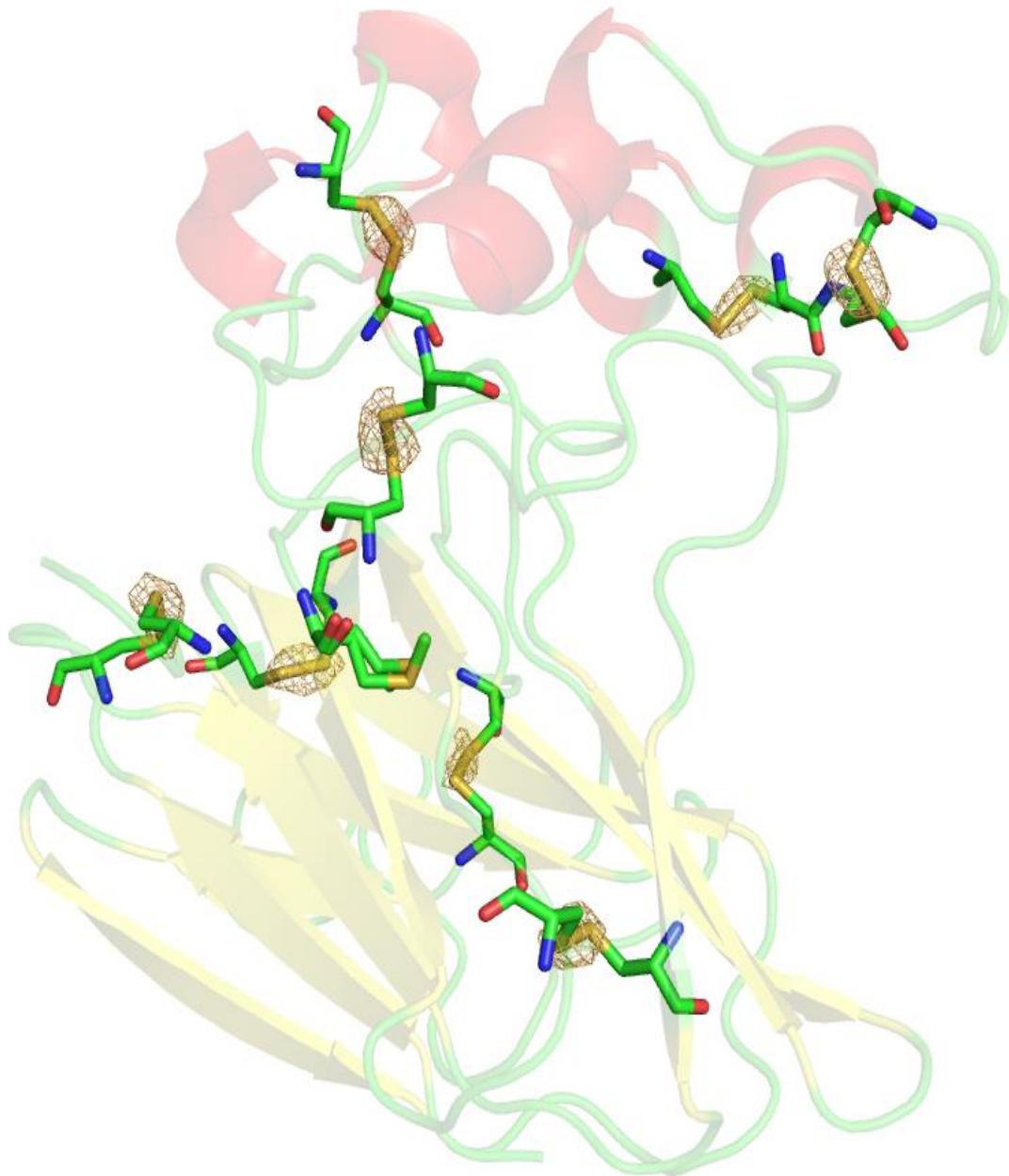
Figure V-7: Thaumatin anomalous difference density map
Phased refined model of thaumatin with the anomalous difference density map contoured at 4.5σ.

V.4. Discussions

The results presented in this chapter have shown the collection of serial data using the tape drive approach having the potential for native SAD phasing. Although some promising attempts to find the substructure were done, ultimate *de novo* phasing, including automatic determination of the substructure, failed.

The data however, was of high enough quality for further studies when using the known sulfur coordinates. The phases that were determined from the native data were sufficient for automatically building the structure – 202 out of 207 amino acids using automatic model building tools. The peak heights from the anomalous density difference map (Figure V-7) are clearly situated around the sulfur atoms and disulfide bonds but the level is reasonably low and the anomalous difference map showed no density around one of the cysteine bonds when contoured at 5σ. The low anomalous intensity is probably due to an insufficient amount of data when compared to the successful FEL thaumatin SAD phasing experiment, where a minimum number of 125,000 patterns were needed for phasing (Nass et al.,2016). In the present study, the final number of indexed patterns was 84,085 which is substantially lower than previously described elsewhere. In another study, serially collected data for a g-coupled receptor could only be phased by native SAD when using 186,688 images measured at 6 keV at a synchrotron radiation source (Weinert et al. 2017). In the present study, the data collection was carried out at an energy limited by the instrumentation at the P11 beamline – 7 keV, whereas by comparison, the FEL thaumatin data collection was carried out at 6 keV.

Nevertheless, attempts were done to collect more diffraction data, but unfortunately PETRA III did not deliver a constant X-ray beam at this time. A second study was planned to be conducted for collection at a higher energy (approximately 12 keV) in order to extend the resolution limit of the data. Due to machine problems, this also was not be possible, since the X-ray beam was not delivered anymore at this time.

A further issue contributing to the rather low number of patterns obtained was the clogging of the sample during delivery, which resulted in ineffective data collection

with downtime consequences. Preventions such as a lower concentration of crystals in solution and sample shaking at different time intervals were taken, but it did not resolve the challenge. If otherwise resolved, this could have improved the final number of diffraction data images and consequently serve to finalize the thaumatin sulfur *de novo* phasing.

V. 5. Conclusions

PfGST crystals obtained from crystallization with the XtalController900 proved to be of high quality, diffracting to similar resolution as previously reported for crystals grown in standard crystallization plates.

Thaumatin data were successfully collected at 7 keV in a serial tape drive approach from microcrystals that were grown from conditions identified from the XtalController900. The data was successfully processed and the 84,085 images used for structure determination had full completeness. Initial native SAD phasing attempts resulted in the identification of the expected 17 sulfur sites. However, unfortunately direct phasing using this substructure failed. When implementing the correct coordinates for the sulfur sites without any additional information, the structure could be phased successfully and built almost completely using automatic model building tools. The calculated anomalous difference density map using the structure factors from all of the processed images and the final refined phases showed clear anomalous density around all 8 cysteines present in the thaumatin structure. To my knowledge, this is the first time that SAD phasing was attempted from serially collected data using the tape drive setup. From previous studies, both SFX and SMX applying SAD, it is known that the amount of data needed surpasses the data collected during the beamtime due to technical issues. Therefore, should an opportunity arise for additional data collection this study would most certainly provide the first *de novo* phased structure using the tape drive setup, providing the opportunity for directly phasing time resolved studies.

Chemicals and hazards

List of chemicals and GHS classification

Table 12: Chemicals used in this work (in alphabetic order)

Chemical	GHS hazard	Hazard statement	Precautionary Statements
Ammonium Sulphate	-	-	-
Ampicillin	GHS08	334, 317	280, 261, 302+352, 342+311
Bis Tris	GHS07	315-319, 335	261, 305+351+338
DTT	GHSO7	302, 315, 319, 335	261, 302+352, 305+351+338, 501
Glutathione Agarose	GHS02	-	-
Glutathione reduziert	GHS02	-	-
Glycerol	-	-	-
HEPES	-	-	-
IPTG	GHSO7, GHSO8	319, 351	281, 305+351+338
NaCl	-	-	-
Paraffin oil	-	-	-
PBS	-	-	-
PEG350	-	-	-
Sodium tartrate	-	-	-
Tris - HCl	GHS07	315-319-335	261, 305+351+338

GHS risk symbols

GHSO2   **GHSO7**   **GHSO8**

Hazard and precautionary statements

GHS precautionary statements

P261	Avoid breathing dust/fume/gas/mist/vapours/spray
P280	Wear protective gloves/protective clothing/eye protection/face protection
P311	Call a POISON CENTER or doctor/physician
P501	Dispose of contents/container in accordance with
P302+P352	IF ON SKIN: Wash with soap and water
P305+P351+P338	IF IN EYES: Rinse cautiously with water for several minutes. Remove contact lenses if present and easy to do - continue rinsing
P342+P311	Call a POISON CENTER or doctor/physician

GHS hazards statements

H302	Harmful if swallowed
H315	Causes skin irritation
H316	Causes mild skin irritation
H317	May cause an allergic skin reaction
H318	Causes serious eye damage
H319	Causes serious eye irritation
H334	May cause allergy or asthma symptoms or breathing difficulties if inhaled
H335	May cause respiratory irritation

List of abbreviations

α	Alpha
Å	Angstrom
ACF	Autocorrelation function
AFM	Atomic force microscopy
Ala	Alanine
APS	Advanced Photon Source
AUC	Analytical centrifugation
Bis-Tris	2,2-Bis(hydroxymethyl)-2,2',2"-nitrilotriethanol
β	Beta
C	Celsius
CCD	Charged-coupled device
Conc	Concentration
CMC	Critical micelle concentration
cP	Dynamic viscosity
Cu	Copper
CX	Classic crystallography
Cys	Cysteine
Da	Dalton
DLS	Dynamic light scattering
DTT	Dithiothreitol
EM	Electron microscopy
ESRF	European synchrotron radiation facility
FELs	Free electron lasers
γ	Gamma
GUI	Graphic user interface
h	Hours
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hz	Hertz

KeV	Kiloelectronvolt
kV	kilovolt
L	Liter
LCP	Lipidic cubic phase
M	Molar
MAD	Multiple wavelength anomalous diffraction
mg	Milligrams
min	Minutes
MIR	Multiple isomorphous replacement
MIRAS	Multiple isomorphous replacements with anomalous signal
mL	Milliliters
mM	Millimolar
mm	Millimeter
MR	Molecular replacement
ms	Milliseconds
mW	Milliwatt
MWCO	Molecular weight cut-off
µm	Micrometers
NaCl	Sodium chloride
nm	Nanometer
NMR	Nuclear magnetic resonance
OD	Optical density
PAGE	Native polyacrylamide gel electrophoresis
PBS	Phosphate buffer saline
PC	Portable computer
PDB	Protein data bank
PEG	Polyethylene glycol
PETRA III	Positron electron tandem ring accelerator III
PfGST	Glutathione S-transferase from *Plasmodium falciparum*
PMT	Photomultiplier tube
Pro	Proline

R_h	Hydrodynamic radius
s	Seconds
S	Sulfur
SAD	Single wavelength anomalous diffraction
SFX	Serial femtosecond crystallography
SFX	Serial femtosecond crystallography
SIR	Single isomorphous replacement
SIRAS	Single isomorphous replacements with anomalous signal
SLS	Static light scattering
SMX	Serial millisecond crystallography
SR	Synchrotron radiation
SP	Target sample
σ	Sigma
THM	Thaumatin
Thr	Threonine
Tris HCl	Tromethamine hydrochloride
XRPD	X-ray powder diffraction
XTC900	XtalController900

List of figures

List of equipment

Table 13: Equipment used for the purpose of this thesis (listed alphabetically)

Instrument	Model	Manufacturer
Beamlines	P11	DESY,Petra III, DESY, Hamburg
	P13	EMBL, PETRA III, DESY Hamburg
Centrifuges	5415R / 5415C / 5804R	Eppendorf, Germany
DLS devices	SpectroLight300	Xtal Concepts GmbH
	SpectroLight600	
Electron Microscope	Talos F200C	Thermo Fisher Scientific
FPLC purifier	Äkta purifier 900	GE Healthcare, USA
Freezer (-20°C)	Liebherr premium	Liebherr, Germany
	B35-85	FRYKA-Kältetechnik, Germany
Imaging	Microscope MDG41	Leica Microsystems, Germany
	Microscope SZX12	Olympus, Japan
Incubator	Heraeus B6120	Heraeus, Germany
	RUMED 3003	Rubarth, Germany
Microbalance	CP2245-OCE	Sartorius, Germany
pH meter	SevenEasy	Mettler-Toledo, Switzerland
Photospectrometry	Nanodrop 2000c	ThermoScientific, Peqlab, Germany
Pipettes	Eppendorf Research, 2 µL, 10 µL, 200 µL, 1000 µL	Eppendorf, Germany
Thermomixer	Comfort	Eppendorf
Vitrobot	Mark III	FEI Company
Vortex mixer	VF2	Janke & Kunkel, IKA Labortechnik, Germany
XtalController	XtalController900	Xtal Concepts GmbH

References

Adachi, Hiroaki, Hiroyoshi Matsumura, Ai Niino, Kazufumi Takano, Takayoshi Kinoshita, Masaichi Warizaya, Tsuyoshi Inoue, Yusuke Mori, and Takatomo Sasaki. 2004. "Improving the Quality of Protein Crystals Using Stirring Crystallization." *Japanese Journal of Applied Physics* 43 (No. 4B): L522–25. https://doi.org/10.1143/JJAP.43.L522.

Adams, Paul D, Pavel V Afonine, Gábor Bunkóczi, Vincent B Chen, Ian W Davis, Nathaniel Echols, Jeffrey J Headd, et al. 2010. "{\it PHENIX}: A Comprehensive Python-Based System for Macromolecular Structure Solution." *Acta Crystallographica Section D* 66 (2): 213–21. https://doi.org/10.1107/S0907444909052925.

Annunziata, Onofrio, Luigi Paduano, Arne J Pearlstein, Donald G Miller, and John G Albright. 2006. "The Effect of Salt on Protein Chemical Potential Determined by Ternary Diffusion in Aqueous Solutions." *The Journal of Physical Chemistry B* 110 (3): 1405–15. https://doi.org/10.1021/jp054543c.

Aquila, Andrew, Mark S. Hunter, R. Bruce Doak, Richard A. Kirian, Petra Fromme, Thomas A. White, Jakob Andreasson, et al. 2012. "Time-Resolved Protein Nanocrystallography Using an X-Ray Free-Electron Laser." *Optics Express* 20 (3): 2706. https://doi.org/10.1364/OE.20.002706.

Asherie, Neer. 2004. "Protein Crystallization and Phase Diagrams." *Methods* 34 (3): 266–72. https://doi.org/10.1016/j.ymeth.2004.03.028.

Ataka, M. 1995. "Nucleation and Growth Kinetics of Hen Egg-White Lysozyme Crystals." *Progress in Crystal Growth and Characterization of Materials* 30 (2–3): 109–28. https://doi.org/10.1016/0960-8974(95)00013-W.

Ataka, Mitsuo, and Michihiko Asai. 1988. "Systematic Studies on the Crystallization of Lysozyme." *Journal of Crystal Growth* 90 (1–3): 86–93. https://doi.org/10.1016/0022-0248(88)90302-8.

Ataka, Mitsuo, and Shoji Tanaka. 1986. "The Growth of Large Single Crystals of Lysozyme." *Biopolymers* 25 (2): 337–50. https://doi.org/10.1002/bip.360250213.

Baitan, Daniela, Robin Schubert, Arne Meyer, Karsten Dierks, Markus Perbandt, and Christian Betzel. 2018. "Growing Protein Crystals with Distinct Dimensions Using Automated Crystallization Coupled with In Situ Dynamic Light Scattering." *JoVE*, no. 138: e57070. https://doi.org/doi:10.3791/57070.

Baldwin, E. T., K. V. Crumley, and C.W. Carter Jr. 1986. "Practical, Rapid Screening of Protein Crystallization Conditions by Dynamic Light Scattering." *Biophysical Journal* 49 (1): 47–48. https://doi.org/10.1016/S0006-3495(86)83587-1.

Barends, Thomas R M, Lutz Foucar, Albert Ardevol, Karol Nass, Andrew Aquila, \textbfSabine \textbfBotha, R Bruce Doak, et al. 2015. "Direct Observation of Ultrafast Collective Motions in CO Myoglobin upon Ligand Dissociation." *Science* 350 (6259): 445–50. https://doi.org/10.1126/science.aac5492.

Barends, Thomas R M, Lutz Foucar, Sabine Botha, R. Bruce Doak, Robert L. Shoeman, Karol Nass, Jason E. Koglin, et al. 2014. "De Novo Protein Crystal Structure Determination from X-Ray Free-Electron Laser Data." *Nature* 505 (7482): 244–47. https://doi.org/10.1038/nature12773.

References

Bergfors, Terese. 2003. "Seeds to Crystals." *Journal of Structural Biology* 142 (1): 66–76. https://doi.org/10.1016/S1047-8477(03)00039-X.

———. 2009. *Protein Crystallization*. Edited by Terese Bergfors. Second. International University Line. https://books.google.de/books?id=VAX7TNFLxksC.

Berman, H.M., J. Westbrook, Z. Feng, G. Gilliand, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. 2000. "The Protein Data Bank." Nucleic Acids Research. 2000. www.rcsb.org.

Berrill, Alex, Jamie Biddlecombe, and Daniel Bracewell. 2011. "Chapter 13 - Product Quality During Manufacture and Supply." In *Peptide and Protein Delivery*, edited by Chris Van Der Walle, 313–39. Boston: Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-384935-9.10013-6.

Beyerlein, Kenneth R., Dennis Dierksmeyer, Valerio Mariani, Manuela Kuhn, Iosifina Sarrou, Angelica Ottaviano, Salah Awel, et al. 2017a. "Mix-and-Diffuse Serial Synchrotron Crystallography." *IUCrJ* 4: 769–77. https://doi.org/10.1107/S2052252517013124.

Beyerlein, Kenneth R, Dennis Dierksmeyer, Valerio Mariani, Manuela Kuhn, Iosifina Sarrou, Angelica Ottaviano, Salah Awel, et al. 2017b. "Mix-and-Diffuse Serial Synchrotron Crystallography." *IUCrJ* 4 (6). https://doi.org/10.1107/S2052252517013124.

Blakeley, Matthew P, Paul Langan, Nobuo Niimura, and Alberto Podjarny. 2008. "Neutron Crystallography: Opportunities, Challenges, and Limitations." *Current Opinion in Structural Biology* 18 (5): 593–600. https://doi.org/https://doi.org/10.1016/j.sbi.2008.06.009.

Boistelle, R., and J. P. Astier. 1988. "Crystallization Mechanisms in Solution." *Journal of Crystal Growth* 90 (1–3): 14–30. https://doi.org/10.1016/0022-0248(88)90294-1.

Botha, S, Daniela Baitan, Katharina E J Jungnickel, Dominik Oberthür, C Schmidt, Stephan Stern, Max Wiedorn, Markus Perbandt, Henry N Chapman, and Christian Betzel. 2018. "De Novo Protein Structure Determination by Heavy-Atom Soaking in Lipidic Cubic Phase and SIRAS Phasing Using Serial Synchrotron Crystallography." *IUCrJ* in press: 1–7. https://doi.org/10.1107/S2052252518009223.

Botha, Sabine, Karol Nass, Thomas R.M. Barends, Wolfgang Kabsch, Beatrice Latz, Florian Dworkowski, Lutz Foucar, et al. 2015. "Room-Temperature Serial Crystallography at Synchrotron X-Ray Sources Using Slowly Flowing Free-Standing High-Viscosity Microstreams." *Acta Crystallographica Section D: Biological Crystallography* 71: 387–97. https://doi.org/10.1107/S1399004714026327.

Boutet, Sébastien, Lukas Lomb, Garth J Williams, Thomas R M Barends, Andrew Aquila, R Bruce Doak, Uwe Weierstall, et al. 2012. "High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography." *Science (New York, N.Y.)* 337 (6092): 362–64. https://doi.org/10.1126/science.1217737.

———. 2012. "High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography." *Science* 337 (6092): 362–64. https://doi.org/10.1126/science.1217737.

Bragg, W. H., and W. L. Bragg. 1913. "The Reflection of X-Ras by Crystals." *Proceedings of the Cambridge Philosophical Society* 17 (43): 428 LP-438. https://doi.org/10.1098/rspa.1983.0054.

References

Burkhardt, Anja, Tim Pakendorf, Bernd Reime, Jan Meyer, Pontus Fischer, Nicolas Stübe, Saravanan Panneerselvam, et al. 2016. "Status of the Crystallography Beamlines at PETRA III." *European Physical Journal Plus* 131 (3): 0–8. https://doi.org/10.1140/epjp/i2016-16056-0.

Burmeister, C., M. Perbandt, Ch. Betzel, R. D. Walter, and E. Liebau. 2003. "Crystallization and Preliminary X-Ray Diffraction Studies of the Glutathione S -Transferase from Plasmodium Falciparum." *Acta Crystallographica Section D Biological Crystallography* 59 (8): 1469–71. https://doi.org/10.1107/S0907444903011090.

Byington, Michael C., Mohammad S. Safari, Jacinta C. Conrad, and Peter G. Vekilov. 2016. "Protein Conformational Flexibility Enables the Formation of Dense Liquid Clusters: Tests Using Solution Shear." *Journal of Physical Chemistry Letters* 7 (13): 2339–45. https://doi.org/10.1021/acs.jpclett.6b00822.

Caldwell, G. W., D. M. Ritchie, J.A. Masucci, W. Hageman, and Z. Yan. 2001. "The New Pre-Preclinical Paradigm: Compound Optimization in Early and Late Phase Drug Discovery." *Bentham Science Publishers* 1 (5): 353–66. https://doi.org/10.2174/1568026013394949.

Chapman, H. N., et al. 2011. "Femtosecond {X}-Ray Protein Nanocrystallography." *Nature* 470 (7332): 73–78. https://doi.org/10.1038/nature09750.Femtosecond.

Chapman, Henry N., Petra Fromme, Anton Barty, Thomas A. White, Richard A. Kirian, Andrew Aquila, Mark S. Hunter, et al. 2011. "Femtosecond X-Ray Protein Nanocrystallography." *Nature* 470 (7332): 73–78. https://doi.org/10.1038/nature09750.

Chavas, L. M.G., L. Gumprecht, and H. N. Chapman. 2015. "Possibilities for Serial Femtosecond Crystallography Sample Delivery at Future Light Sources." *Structural Dynamics* 2 (4): 1–14. https://doi.org/10.1063/1.4921220.

Cianci, Michele, Gleb Bourenkov, Guillaume Pompidor, Ivars Karpics, Johanna Kallio, Isabel Bento, Manfred Roessle, Florent Cipriani, Stefan Fiedler, and Thomas R. Schneider. 2017. "P13, the EMBL Macromolecular Crystallography Beamline at the Low-Emittance PETRA III Ring for High- and Low-Energy Phasing with Variable Beam Focusing." *Journal of Synchrotron Radiation* 24 (1): 323–32. https://doi.org/10.1107/S1600577516016465.

Cianci, Michele, John R Helliwell, and Atsuo Suzuki. 2008. "The Interdependence of Wavelength, Redundancy and Dose in Sulfur SAD Experiments." *Acta Crystallographica Section D* 64 (12): 1196–1209. https://doi.org/10.1107/S0907444908030503.

Conrad, Chelsie E, Shibom Basu, Daniel James, Dingjie Wang, Alexander Schaffer, Shatabdi Roy-Chowdhury, Nadia A Zatsepin, et al. 2015. "A Novel Inert Crystal Delivery Medium for Serial Femtosecond Crystallography." *IUCrJ* 2 (Pt 4): 421–30. https://doi.org/10.1107/S2052252515009811.

Cowtan, Kevin. 2010. "Recent Developments in Classical Density Modification." *Acta Crystallographica Section D: Biological Crystallography* 66 (4): 470–78. https://doi.org/10.1107/S090744490903947X.

Diederichs, Kay, and P Andrew Karplus. 1997. "Improved R-Factors for Diffraction Data Analysis in Macromolecular Crystallography." *Nature Structural Biology* 4 (April): 269. http://dx.doi.org/10.1038/nsb0497-269.

Dierks, Karsten, Arne Meyer, Howard Einspahr, and Christian Betzel. 2008. "Dynamic Light Scattering in Protein Crystallization Droplets : Adaptations for Analysis and Optimization of Crystallization Processes & DESIGN 2008."

Driessche, Alexander E S Van, Nani Van Gerven, Paul H H Bomans, Rick R M Joosten, Heiner Friedrich, David Gil-Carton, Nico A J M Sommerdijk, and Mike Sleutel. 2018. "Molecular Nucleation Mechanisms and Control Strategies for Crystal Polymorph Selection." *Nature* 556 (April): 89. http://dx.doi.org/10.1038/nature25971.

Duisenberg, Albert J M. 1992. "Indexing in Single-Crystal Diffractometry with an Obstinate List of Reflections." *Journal of Applied Crystallography* 25 (pt 2): 92–96. https://doi.org/10.1107/S0021889891010634.

Dumetz, André C., Aaron M. Chockla, Eric W. Kaler, and Abraham M. Lenhoff. 2008. "Protein Phase Behavior in Aqueous Solutions: Crystallization, Liquid-Liquid Phase Separation, Gels, and Aggregates." *Biophysical Journal* 94 (2): 570–83. https://doi.org/10.1529/biophysj.107.116152.

DuPont, Circuits and Packaging Materials. 2013. No Title, issued 2013. http://www2.dupont.com/Kapton/en_US/news_events/article20131115.html.

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

Evans, Philip R., and Garib N. Murshudov. 2013. "How Good Are My Data and What Is the Resolution?" *Acta Crystallographica Section D: Biological Crystallography* 69 (7): 1204–14. https://doi.org/10.1107/S0907444913000061.

Ewing, F, E. Forsythe, and M. Pusey. 1994. "Orthorhombic Lysozyme Solubility." *Acta Crystallographica Section D* 50 (4): 424–28. https://doi.org/10.1107/S0907444993014428.

Feliciano, A. S., A. I. Dias, and D. M.F. Prazeres. 2000. "The Effect of Stirring and Seeding on the AcPheLeuNH2synthesis and Crystallization in a Reversed Micellar System." *Enzyme and Microbial Technology* 27 (3–5): 264–69. https://doi.org/10.1016/S0141-0229(00)00201-5.

Ferré-D'Amaré, Adrian R., and Stephen K. Burley. 1994. "Use of Dynamic Light Scattering to Assess Crystallizability of Macromolecules and Macromolecular Assemblies." *Structure* 2 (5): 357–59. https://doi.org/10.1016/S0969-2126(00)00037-X.

Ferrone, Frank A., Maria Ivanova, and Ravi Jasuja. 2002. "Heterogeneous Nucleation and Crowding in Sickle Hemoglobin: An Analytic Approach." *Biophysical Journal* 82 (1): 399–406. https://doi.org/10.1016/S0006-3495(02)75404-0.

Foucar, Lutz, Anton Barty, Nicola Coppola, Robert Hartmann, Peter Holl, Uwe Hoppe, Stephan Kassemeyer, et al. 2012. "CASS - CFEL-ASG Software Suite." *Computer Physics Communications* 183 (10): 2207–13. https://doi.org/10.1016/j.cpc.2012.04.023.

Fowler, Susan B, Stephen Poon, Roman Muff, Fabrizio Chiti, Christopher M Dobson, and Jesús Zurdo. 2005. "Rational Design of Aggregation-Resistant Bioactive Peptides: Reengineering Human Calcitonin." *Proceedings of the National Academy of Sciences* 102 (29): 10105–10. https://doi.org/10.1073/pnas.0501215102.

References

Friedrich, W, Knipping P., and M Laue. 1913. "Interferenzerscheinungen Bei Röntgenstrahlen." *Ann. Phys.* 346 (10): 971–88. http://dx.doi.org/10.1002/andp.19133461004.

Gadomski, A., and J. Siódmiak. 2003. "A Novel Model of Protein Crystal Growth: Kinetic Limits, Length Scales and the Role of the Double Layer." *Croatica Chemica Acta* 76 (2): 129–136. http://pubwww.carnet.hr/ccacaa/CCA-PDF/cca2003/v76-n2/CCA_76_2003_129-136_gadomski.pdf.

Galkin, O., and P. G. Vekilov. 2000. "Control of Protein Crystal Nucleation around the Metastable Liquid-Liquid Phase Boundary." *Proceedings of the National Academy of Sciences* 97 (12): 6277–81. https://doi.org/10.1073/pnas.110000497.

Galkin, Oleg, Weichun Pan, Luis Filobelo, Rhoda Elison Hirsch, Ronald L. Nagel, and Peter G. Vekilov. 2007. "Two-Step Mechanism of Homogeneous Nucleation of Sickle Cell Hemoglobin Polymers." *Biophysical Journal* 93 (3): 902–13. https://doi.org/10.1529/biophysj.106.103705.

Galkin, Oleg, and Peter G. Vekilov. 1999. "Direct Determination of the Nucleation Rates of Protein Crystals." *The Journal of Physical Chemistry B* 103 (49): 10965–71. https://doi.org/10.1021/jp992786x.

Garman, Elspeth F. 2013. "Radiation Damage in Macromolecular Crystallography: What Is It and Why Do We Care?" *NATO Science for Peace and Security Series A: Chemistry and Biology*, 69–77. https://doi.org/10.1007/978-94-007-6232-9-7.

Gibbs, J.Williard. 1978. "On the Equilibrium of Heterogeneous Substances." *American Journal of Science and Arts* 16 (96): 1820–79.

Gliko, Olga, Nikolaus Neumaier, Weichun Pan, Ilka Haase, Markus Fischer, Adelbert Bacher, Sevil Weinkauf, and Peter G. Vekilov. 2005. "A Metastable Prerequisite for the Growth of Lumazine Synthase Crystals." *Journal of the American Chemical Society* 127 (10): 3433–38. https://doi.org/10.1021/ja043218k.

Gliko, Olga, Weichun Pan, Panagiotis Katsonis, Nikolaus Neumaier, Oleg Galkin, Sevil Weinkauf, and Peter G. Vekilov. 2007. "Metastable Liquid Clusters in Super- and Undersaturated Protein Solutions." *Journal of Physical Chemistry B* 111 (12): 3106–14. https://doi.org/10.1021/jp068827o.

Goon, Scarlett, John F Kelly, Susan M Logan, Cheryl P Ewing, and Patricia Guerry. 2003. "Pseudaminic Acid, the Major Modification on Campylobacter Flagellin, Is Synthesized via the Cj1293 Gene." *Molecular Microbiology* 50 (2): 659–71. https://doi.org/10.1046/j.1365-2958.2003.03725.x.

Groenewold, Jan, and Willem K Kegel. 2001. "Anomalously Large Equilibrium Clusters of Colloids." *The Journal of Physical Chemistry B* 105 (47): 11702–9. https://doi.org/10.1021/jp011646w.

Hagmeyer, Daniel, Johannes Ruesing, Tassilo Fenske, Heinz-Werner Klein, Carsten Schmuck, Wolfgang Schrader, Manuel E Minas da Piedade, and Matthias Epple. 2012. "Direct Experimental Observation of the Aggregation of α-Amino Acids into 100–200 Nm Clusters in Aqueous Solution." *RSC Advances* 2 (11): 4690–96. https://doi.org/10.1039/C2RA01352E.

Harding, Stephen E, and Kornelia Jumel. 1998. "Light Scattering." *Current Protocols in Protein Science* 11 (1): 7.8.1-7.8.14. https://doi.org/10.1002/0471140864.ps0708s11.

References

Hekmat, Dariusch, Max Huber, Christian Lohse, Nikolas Von Den Eichen, and Dirk Weuster-Botz. 2017. "Continuous Crystallization of Proteins in a Stirred Classified Product Removal Tank with a Tubular Reactor in Bypass." *Crystal Growth and Design* 17 (8): 4162–69. https://doi.org/10.1021/acs.cgd.7b00436.

Henderson, R, J M Baldwin, T A Ceska, F Zemlin, E Beckmann, and K H Downing. 1990. "Model for the Structure of Bacteriorhodopsin Based on High-Resolution Electron Cryo-Microscopy." *Journal of Molecular Biology* 213 (4): 899–929. https://doi.org/https://doi.org/10.1016/S0022-2836(05)80271-2.

Hendrickson, W a, J R Horton, and D M LeMaster. 1990. "Selenomethionyl Proteins Produced for Analysis by Multiwavelength Anomalous Diffraction (MAD): A Vehicle for Direct Determination of Three-Dimensional Structure." *The EMBO Journal* 9 (5): 1665–72. https://doi.org/10.1002/j.1460-2075.1990.tb08287.x.

Huang, Chia-Ying, Vincent Olieric, Pikyee Ma, Ezequiel Panepucci, Kay Diederichs, Meitian Wang, and Martin Caffrey. 2015. "*In Meso in Situ* Serial X-Ray Crystallography of Soluble and Membrane Proteins." *Acta Crystallographica Section D Biological Crystallography* 71 (6): 1238–56. https://doi.org/10.1107/S1399004715005210.

Hughes, Colan E., Said Hamad, Kenneth D.M. Harris, C. Richard A. Catlow, and Peter C. Griffiths. 2007. "A Multi-Technique Approach for Probing the Evolution of Structural Properties during Crystallization of Organic Materials from Solution." *Faraday Discussions* 136 (0): 71–89. https://doi.org/10.1039/b616611c.

Hutchens, Shelby B., and Zhen Gang Wang. 2007. "Metastable Cluster Intermediates in the Condensation of Charged Macromolecule Solutions." *Journal of Chemical Physics* 127 (8). https://doi.org/10.1063/1.2761891.

Ii, W Nicholson Price, Yang Chen, Samuel K Handelman, Helen Neely, Richard Karlin, Rajesh Nair, Jinfeng Liu, et al. 2010. "Understanding the Physical Properties Controlling Protein Crystallization." *Nat Biotechnol.* 27 (1): 51–57. https://doi.org/10.1038/nbt.1514.Understanding.

Jawor-Baczynska, Anna, Barry D Moore, Han Seung Lee, Alon V McCormick, and Jan Sefcik. 2013. "Population and Size Distribution of Solute-Rich Mesospecies within Mesostructured Aqueous Amino Acid Solutions." *Faraday Discuss.* 167 (0): 425–40. https://doi.org/10.1039/C3FD00066D.

Jawor-Baczynska, Anna, Jan Sefcik, and Barry Moore D. 2013. "250 Nm Glycine-Rich Nanodroplets Are Formed on Dissolution of Glycine Crystals But Are Too Small To Provide Productive Nucleation Sites." *Crystal Growth and Design* 13 (2): 470–78. https://doi.org/10.1021/cg300150u.

Johansson, Linda C, David Arnlund, Thomas A White, Gergely Katona, Daniel P DePonte, Uwe Weierstall, R Bruce Doak, et al. 2012. "Lipidic Phase Membrane Protein Serial Femtosecond Crystallography." *Nature Methods* 9 (3): 263–65. https://doi.org/10.1038/nmeth.1867.

Juarez-Martines, Garza C, Castillo R, and A Moreno. 2001. "A Dynamic Light Scattering Investigation of the Nucleation and Growth of Thaumatin Crystals." *Journal of Crystal Growth* 232 (1–4): 119–31. https://doi.org/10.1016/S0022-0248(01)01081-8.

K, Cowtan. 2006. "The Buccaneer Software for Automated Model Building. 1. Tracing Protein Chains." *Acta Crystallogr D Biol Crystallogr* 62: 1002.

References

Kabsch, Wolfgang. 2010. "{\it XDS}." *Acta Crystallographica Section D* 66 (2): 125–32. https://doi.org/10.1107/S0907444909047337.

———. 2014. "Processing of X-Ray Snapshots from Crystals in Random Orientations." *Acta Crystallographica Section D* 70 (8): 2204–16. https://doi.org/10.1107/S1399004714013534.

Kam, Z, H B Shore, and G Feher. 1978. "On the Crystallization of Proteins." *Journal of Molecular Biology* 123 (4): 539–55. https://doi.org/https://doi.org/10.1016/0022-2836(78)90206-1.

Kashchiev, Dimo, Peter G. Vekilov, and Anatoly B. Kolomeisky. 2005. "Kinetics of Two-Step Nucleation of Crystals." *Journal of Chemical Physics* 122 (24): 1–6. https://doi.org/10.1063/1.1943389.

Kendrew, J. C., G. Bodo, H. M. Dintzis, R. G. Parrish, and H Wyckoff. 1958. "© 1958 Nature Publishing Group." *Nature* 181: 662–66.

Kern, Jan, Roberto Alonso-Mori, Rosalie Tran, Johan Hattne, Richard J Gildea, Nathaniel Echols, Carina Glöckner, et al. 2013. "Simultaneous Femtosecond X-Ray Spectroscopy and Diffraction of Photosystem II at Room Temperature." *Science* 340 (6131): 491–95. https://doi.org/10.1126/science.1234273.

Kern, Jan, Johan Hattne, Rosalie Tran, Roberto Alonso-Mori, Hartawan Laksmono, Sheraz Gul, Raymond G. Sierra, et al. 2014. "Methods Development for Diffraction and Spectroscopy Studies of Metalloenzymes at X-Ray Free-Electron Lasers." *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1647). https://doi.org/10.1098/rstb.2013.0590.

Kirian, Richard A, Xiaoyu Wang, Uwe Weierstall, Kevin E Schmidt, John C H Spence, Mark Hunter, Petra Fromme, Thomas White, Henry N Chapman, and James Holton. 2010. "Femtosecond Protein Nanocrystallography---Data Analysis Methods." *Opt. Express* 18 (6): 5713–23. https://doi.org/10.1364/OE.18.005713.

Kirian, Richard A, Thomas A White, James M Holton, Henry N Chapman, Petra Fromme, Anton Barty, Lukas Lomb, et al. 2011. "Structure-Factor Analysis of Femtosecond Microdiffraction Patterns from Protein Nanocrystals." *Acta Crystallographica Section A* 67 (2): 131–40. https://doi.org/10.1107/S0108767310050981.

Kisselman, Gera, Wei Qiu, Vladimir Romanov, Christine M. Thompson, Robert Lam, Kevin P. Battaile, Emil F. Pai, and Nickolay Y. Chirgadze. 2011. "X-CHIP: An Integrated Platform for High-Throughput Protein Crystallization and on-the-Chip X-Ray Diffraction Data Collection." *Acta Crystallographica Section D: Biological Crystallography* 67 (6): 533–39. https://doi.org/10.1107/S0907444911011589.

Kremer, Werner, and Hans Robert Kalbitzer. 2001. "[1] - Physiological Conditions and Practicality for Protein Nuclear Magnetic Resonance Spectroscopy: Experimental Methodologies and Theoretical Background." In *Nuclear Magnetic Resonance of Biological Macromolecules - Part B*, edited by Thomas L James, Volker Dötsch, and Uli Schmitz, 339:3–19. Methods in Enzymology. Academic Press. https://doi.org/https://doi.org/10.1016/S0076-6879(01)39306-0.

Kupitz, Christopher, Shibom Basu, Ingo Grotjohann, Raimund Fromme, Nadia A Zatsepin, Kimberly N Rendek, Mark S Hunter, et al. 2014. "Serial Time-Resolved Crystallography of Photosystem II Using a Femtosecond X-Ray Laser." *Nature* 513: 261. https://doi.org/10.1038/nature13453.

References

Kupitz, Christopher, Ingo Grotjohann, Chelsie E. Conrad, Shatabdi Roy-Chowdhury, Raimund Fromme, and Petra Fromme. 2014. "Microcrystallization Techniques for Serial Femtosecond Crystallography Using Photosystem II from Thermosynechococcus Elongatus as a Model System." *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1647). https://doi.org/10.1098/rstb.2013.0316.

Lee, Dan Bi, Jong Min Kim, Jong Hyeon Seok, Ji Hye Lee, Jae Deok Jo, Ji Young Mun, Chelsie Conrad, et al. 2018. "Supersaturation-Controlled Microcrystallization and Visualization Analysis for Serial Femtosecond Crystallography." *Scientific Reports* 8 (1): 1–10. https://doi.org/10.1038/s41598-018-20899-9.

Levantino, Matteo, Briony A. Yorke, Diana C F Monteiro, Marco Cammarata, and Arwen R. Pearson. 2015. "Using Synchrotrons and XFELs for Time-Resolved X-Ray Crystallography and Solution Scattering Experiments on Biomolecules." *Current Opinion in Structural Biology* 35: 41–48. https://doi.org/10.1016/j.sbi.2015.07.017.

Liebau, Eva, Bärbel Bergmann, Alison M. Campbell, Paul Teesdale-Spittle, Peter M. Brophy, Kai Lüersen, and Rolf D. Walter. 2002. "The Glutathione S-Transferase from Plasmodium Falciparum." *Molecular and Biochemical Parasitology* 124 (1–2): 85–90. https://doi.org/10.1016/S0166-6851(02)00160-3.

Maes, Dominique, Maria A. Vorontsova, Marco A.C. Potenza, Tiziano Sanvito, Mike Sleutel, Marzio Giglio, and Peter G. Vekilov. 2015. "Do Protein Crystals Nucleate within Dense Liquid Clusters?" *Acta Crystallographica Section:F Structural Biology Communications* 71: 815–22. https://doi.org/10.1107/S2053230X15008997.

Maki, S., R. Murai, H. Y. Yoshikawa, T. Kitatani, S. Nakata, H. Kawahara, H. Hasenaka, et al. 2008. "Protein Crystallization in a 100 Nl Solution with New Stirring Equipment." *Journal of Synchrotron Radiation* 15 (3): 269–72. https://doi.org/10.1107/S0909049508001842.

Martin-Garcia, Jose M., Chelsie E. Conrad, Garrett Nelson, Natasha Stander, Nadia A. Zatsepin, James Zook, Lan Zhu, et al. 2017. "Serial Millisecond Crystallography of Membrane and Soluble Protein Microcrystals Using Synchrotron Radiation." *IUCrJ* 4: 439–54. https://doi.org/10.1107/S205225251700570X.

McPherson, Alex. 2009. *Chapter 1 Introduction to the Crystallization of Biological Macromolecules*. *Current Topics in Membranes*. 1st ed. Vol. 63. Elsevier Inc. https://doi.org/10.1016/S1063-5823(09)63001-5.

Meents, A., D. Oberthuer, and V Srajer. 2017. "Single-Shot Pink Beam Serial Crystallography: Thaumatin." https://doi.org/10.2210/pdb5MJG/pdb.

Merritt, Ethan A. 2012. "X-Ray Anomalous Scattering." 2012. http://skuld.bmsc.washington.edu/scatter/AS_index.html.

Meyer, Arne, Karsten Dierks, Dierk Hilterhaus, Thomas Klupsch, Peter Mühlig, Jens Kleesiek, Robert Schöpflin, Howard Einspahr, Rolf Hilgenfeld, and Christian Betzel. 2012. "Single-Drop Optimization of Protein Crystallization." *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* 68 (8): 994–98. https://doi.org/10.1107/S1744309112024074.

References

Meyer, Arne, Karsten Dierks, Rana Hussein, Karl Brillet, Hevila Brognaro, and Christian Betzel. 2015. "Systematic Analysis of Protein-Detergent Complexes Applying Dynamic Light Scattering to Optimize Solutions for Crystallization Trials." *Acta Crystallographica Section F:Structural Biology Communications* 71: 75–81. https://doi.org/10.1107/S2053230X14027149.

Michel, Hartmut. 1983. "Crystallization of Membrane Proteins." *Trends in Biochemical Sciences* 8 (2): 56–59. https://doi.org/https://doi.org/10.1016/0968-0004(83)90390-0.

Mikol, V, E Hirsch, and R Giege. 1989. "Monitoring Protein Crystallization by Dynamic Light Scattering." *FEBS Lett.* 258 (1): 63–66.

Mueller, C., A. Marx, S. W. Epp, Y. Zhong, A. Kuo, A. R. Balo, J. Soman, et al. 2015. "Fixed Target Matrix for Femtosecond Time-Resolved and in Situ Serial Micro-Crystallography." *Structural Dynamics* 2 (5). https://doi.org/10.1063/1.4928706.

Murshudov, Garib N., Pavol Skubák, Andrey A. Lebedev, Navraj S. Pannu, Roberto A. Steiner, Robert A. Nicholls, Martyn D. Winn, Fei Long, and Alexei A. Vagin. 2011. "REFMAC5 for the Refinement of Macromolecular Crystal Structures." *Acta Crystallographica Section D: Biological Crystallography* 67 (4): 355–67. https://doi.org/10.1107/S0907444911001314.

Nakane, Takanori, Shinya Hanashima, Mamoru Suzuki, Haruka Saiki, Taichi Hayashi, Keisuke Kakinouchi, Shigeru Sugiyama, et al. 2016. "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): 13039–44. https://doi.org/10.1073/pnas.1602531113.

Nass, Karol, Lutz Foucar, Thomas R.M. Barends, Elisabeth Hartmann, Sabine Botha, Robert L. Shoeman, R. Bruce Doak, et al. 2015. "Indications of Radiation Damage in Ferredoxin Microcrystals Using High-Intensity X-FEL Beams." *Journal of Synchrotron Radiation* 22 (2): 225–38. https://doi.org/10.1107/S1600577515002349.

Nass, Karol, Anton Meinhart, Thomas R.M. Barends, Lutz Foucar, Alexander Gorel, Andrew Aquila, Sabine Botha, et al. 2016. "Protein Structure Determination by Single-Wavelength Anomalous Diffraction Phasing of X-Ray Free-Electron Laser Data." *IUCrJ* 3: 180–91. https://doi.org/10.1107/S2052252516002980.

Nederlof, Igor, Yao Wang Li, Marin Van Heel, and Jan Pieter Abrahams. 2013. "Imaging Protein Three-Dimensional Nanocrystals with Cryo-EM." *Acta Crystallographica Section D: Biological Crystallography* 69 (5): 852–59. https://doi.org/10.1107/S0907444913002734.

Neutze, Richard, and Keith Moffat. 2013. "Time-Resolved Structural Studies at Synchrotrons and X-Ray Free Electron Lasers: Opportunities and Challenges." *Current Opinion in Structural Biology* 22 (5): 651–59. https://doi.org/10.1016/j.sbi.2012.08.006.Time-resolved.

Neutze, Richard, Remco Wouts, David van der Spoel, Edgar Weckert, and Janos Hajdu. 2000. "Potential for Biomolecular Imaging with Femtosecond X-Ray Pulses." *Nature*, no. 6796: 752–57. https://doi.org/10.1038/35021099.

References

Niesen, Frank H, Anja Koch, Ulf Lenski, Ulrich Harttig, Yvette Roske, Udo Heinemann, and Klaus Peter Hofmann. 2008. "An Approach to Quality Management in Structural Biology: Biophysical Selection of Proteins for Successful Crystallization." *Journal of Structural Biology* 162 (3): 451–59. https://doi.org/https://doi.org/10.1016/j.jsb.2008.03.007.

Niethammer, Barbara. 2008. "Effective Theories for Ostwald Ripening." *Analysis and Stochastics of Growth Processes and Interface Models*, 1–21. https://doi.org/10.1093/acprof:oso/9780199239252.003.0010.

Nogly, Przemyslaw, Daniel James, Dingjie Wang, Thomas A. White, Nadia Zatsepin, Anastasya Shilova, Garrett Nelson, et al. 2015. "Lipidic Cubic Phase Serial Millisecond Crystallography Using Synchrotron Radiation." *IUCrJ* 2: 168–76. https://doi.org/10.1107/S2052252514026487.

Oberthür, Dominik. 2012. "Detailed Analysis of Protein Crystallization and Aggregation Phenomena Applying Dynamic Light Scattering." http://ediss.sub.uni-hamburg.de/volltexte/2012/5833/.

Pan, Weichun, Oleg Galkin, Luis Filobelo, Ronald L. Nagel, and Peter G. Vekilov. 2007. "Metastable Mesoscopic Clusters in Solutions of Sickle-Cell Hemoglobin." *Biophysical Journal* 92 (1): 267–77. https://doi.org/10.1529/biophysj.106.094854.

Pan, Weichun, Peter G. Vekilov, and Vassiliy Lubchenko. 2010. "Origin of Anomalous Mesoscopic Phases in Protein Solutions." *Journal of Physical Chemistry B* 114 (22): 7620–30. https://doi.org/10.1021/jp100617w.

Pande, Kanupriya, Christopher D M Hutchison, Gerrit Groenhof, Andy Aquila, Josef S Robinson, Jason Tenboer, Shibom Basu, et al. 2016. "Femtosecond Structural Dynamics Drives the Trans/Cis Isomerization in Photoactive Yellow Protein." *Science (New York, N.Y.)* 352 (6286): 725–29. https://doi.org/10.1126/science.aad5081.

Perutz, M F. 1969. "X-Ray Analysis , Structure and Function of Enzymes." *Main* 8: 455–66.

Powell, Harold R., Owen Johnson, and Andrew G.W. Leslie. 2013. "Autoindexing Diffraction Images with IMosflm." *Acta Crystallographica Section D: Biological Crystallography* 69 (7): 1195–1203. https://doi.org/10.1107/S0907444912048524.

Proteau, Ariane, Rong Shi, and Miroslaw Cygler. 2010. "Application of Dynamic Light Scattering in Protein Crystallization." *Current Protocols in Protein Science*, no. SUPPL. 61: 1–9. https://doi.org/10.1002/0471140864.ps1710s61.

Provenche, Stephen W. 1982. "CONTIN : A GENERAL PURPOSE CONSTRAINED REGULARIZATION PROGRAM FOR INVERTING NOISY LINEAR ALGEBRAIC AND INTEGRAL EQUATIONS Stephen W ., PROVENCHER LONG WRITE-UP CONTIN Implements the Constrained Regulari- Zation Algorithm Described in Detail in Ref . [1]." *Computer Physics Communications* 27: 229–42.

Rajendran, Arivazhagan, Masayuki Endo, and Hiroshi Sugiyama. 2012. "Chapter 2 - Structural and Functional Analysis of Proteins by High-Speed Atomic Force Microscopy." In *Structural and Mechanistic Enzymology*, edited by Christo Christov and Tatyana Karabencheva-Christova, 87:5–55. Advances in Protein Chemistry and Structural Biology. Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-398312-1.00002-0.

Rein ten Wolde, Pieter, and Daan Frenkel. 1997. "Enhancement of Protein Crystal Nucleation by Critical Density Fluctuations." *Science* 277 (5334): 1975–78. https://doi.org/10.1126/science.277.5334.1975.

Ricci, Margaret S, and David N. Brems. 2004. "Common Structural Stability Properties of 4-Helical Bundle Cytokines: Possible Physiological and Pharmaceutical Consequences." *Bentham Science Publishers* 10 (31): 3901–11. https://doi.org/10.2174/1381612043382611.

Riekel, Christian, Manfred Burghammer, and Gebhard Schertler. 2005. "Protein Crystallography Microdiffraction." *Current Opinion in Structural Biology* 15 (5): 556–62. https://doi.org/10.1016/j.sbi.2005.08.013.

Roberts, Christopher J. 2014. "Protein Aggregation and Its Impact on Product Quality." *Current Opinion in Biotechnology* 30: 211–17. https://doi.org/10.1016/j.copbio.2014.08.001.

Roedig, Philip, Ramona Duman, Juan Sanchez-Weatherby, Ismo Vartiainen, Anja Burkhardt, Martin Warmer, Christian David, Armin Wagner, and Alke Meents. 2016. "Room-Temperature Macromolecular Crystallography Using a Micro-Patterned Silicon Chip with Minimal Background Scattering." *Journal of Applied Crystallography* 49 (3): 968–75. https://doi.org/10.1107/S1600576716006348.

Röntgen, W C. 1898. "Ueber Eine Neue Art von Strahlen." *Ann. Phys.* 300 (1): 1–11. http://dx.doi.org/10.1002/andp.18983000102.

Rossmann, Michael. 1990. "The Molecular Replacement Method." *Acta Crystallographica Section A: Foundations of Crystallography* 46 (2): 73–82. https://doi.org/10.1107/S0108767389009815.

Rupp, Bernhard. 2010. *Biomolecular Crystallograhy*. First. New York, USA: Garland Science.

Santos, Nuno C., and Miguel A.R.B. Castanho. 1996. "Teaching Light Scattering Spectroscopy: The Dimension and Shape of Tobacco Mosaic Virus." *Biophysical Journal* 71 (3): 1641–50. https://doi.org/10.1016/S0006-3495(96)79369-4.

Saridakis, E.E. G., P. D. S. Stewart, L. F. Lloyd, and D. M. Blow. 1994. "Phase Diagram and Dilution Experiments in the Crystallization of Carboxypeptidase G2." *Acta Crystallographica Section D* 50 (3): 293–97. https://doi.org/10.1107/S0907444993013186.

Sauter, Andrea, Felix Roosen-Runge, Fajun Zhang, Gudrun Lotze, Artem Feoktystov, Robert M.J. Jacobs, and Frank Schreiber. 2015. "On the Question of Two-Step Nucleation in Protein Crystallization." *Faraday Discussions* 179: 41–58. https://doi.org/10.1039/c4fd00225c.

Sauter, Andrea, Felix Roosen-Runge, Fajun Zhang, Gudrun Lotze, Robert M.J. Jacobs, and Frank Schreiber. 2015. "Real-Time Observation of Nonclassical Protein Crystallization Kinetics." *Journal of the American Chemical Society* 137 (4): 1485–91. https://doi.org/10.1021/ja510533x.

Sauter, Claude, Bernard Lorber, Daniela Kern, Jean Cavarelli, Dino Moras, and Richard Giegé. 1999. "Crystallogenesis Studies on Yeast Aspartyl-TRNA Synthetase: Use of Phase Diagram to Improve Crystal Quality." *Acta Crystallographica Section D* 55 (1): 149–56. https://doi.org/10.1107/S0907444998010890.

Schlichting, Ilme. 2015a. "Serial Femtosecond Crystallography: The First Five Years." *IUCrJ* 2 (2013): 246–55. https://doi.org/10.1107/S205225251402702X.

———. 2015b. "Serial Femtosecond Crystallography: The First Five Years." *IUCrJ* 2 (2): 246–55. https://doi.org/10.1107/S205225251402702X.

Schmidt, Marius. 2013. "Mix and Inject: Reaction Initiation by Diffusion for Time-Resolved Macromolecular Crystallography." *Advances in Condensed Matter Physics* 2013 (167276). https://doi.org/doi:10.1155/2013/167276.

Schrödinger, L L C. 2015. "The PyMOL Molecular Graphics System, Version~1.8."

Schubert, Robin, Arne Meyer, Daniela Baitan, Karsten Dierks, Markus Perbandt, and Christian Betzel. 2017. "Real-Time Observation of Protein Dense Liquid Cluster Evolution during Nucleation in Protein Crystallization." *Crystal Growth and Design* 17 (3): 954–58. https://doi.org/10.1021/acs.cgd.6b01826.

Schubert, Robin, Arne Meyer, Karsten Dierks, Svetlana Kapis, Rudolph Reimer, Howard Einspahr, Markus Perbandt, and Christian Betzel. 2015. "Reliably Distinguishing Protein Nanocrystals from Amorphous Precipitate by Means of Depolarized Dynamic Light Scattering." *Journal of Applied Crystallography* 48 (August): 1476–84. https://doi.org/10.1107/S1600576715014740.

Sear, Richard P. 2007. "Nucleation: Theory and Applications to Protein Solutions and Colloidal Suspensions." *Journal of Physics: Condensed Matter* 19 (3): 33101. http://stacks.iop.org/0953-8984/19/i=3/a=033101.

Sheldrick, George M. 2010. "Experimental Phasing with {\it SHELXC}/{\it D}/{\it E}: Combining Chain Tracing with Density Modification." *Acta Crystallographica Section D* 66 (4): 479–85. https://doi.org/10.1107/S0907444909038360.

Shi, Rong, Magda Villarroya, Rafael Ruiz-Partida, Yunge Li, Ariane Proteau, Silvia Prado, Ismaïl Moukadiri, et al. 2009. "Structure-Function Analysis of Escherichia Coli MnmG (GidA), a Highly Conserved TRNA-Modifying Enzyme." *Journal of Bacteriology* 191 (24): 7614–19. https://doi.org/10.1128/JB.00650-09.

Sichun, Yang. 2014. "Methods for SAXS-Based Structure Determination of Biomolecular Complexes." *Advanced Materials* 26 (46): 7902–10. https://doi.org/10.1002/adma.201304475.

Skouri, M, B Lorber, R Giegé, J.-P. Munch, and J S Candau. 1995. "Effect of Macromolecular Impurities on Lysozyme Solubility and Crystallizability: Dynamic Light Scattering, Phase Diagram, and Crystal Growth Studies." *Journal of Crystal Growth* 152 (3): 209–20. https://doi.org/https://doi.org/10.1016/0022-0248(95)00051-8.

Sleutel, Mike, and Alexander E. S. Van Driessche. 2014. "Role of Clusters in Nonclassical Nucleation and Growth of Protein Crystals." *Proceedings of the National Academy of Sciences* 111 (5): E546–53. https://doi.org/10.1073/pnas.1309320111.

Smejkal, Benjamin, Bernhard Helk, Jean Michel Rondeau, Sabine Anton, Angelika Wilke, Peter Scheyerer, Jacqueline Fries, Dariusch Hekmat, and Dirk Weuster-Botz. 2013. "Protein Crystallization in Stirred Systems-Scale-up via the Maximum Local Energy Dissipation." *Biotechnology and Bioengineering* 110 (7): 1956–63. https://doi.org/10.1002/bit.24845.

References

Stellato, Francesco, Dominik Oberthür, Mengning Liang, Richard Bean, Cornelius Gati, Oleksandr Yefanov, Anton Barty, et al. 2014. "Room-Temperature Macromolecular Serial Crystallography Using Synchrotron Radiation." *IUCrJ* 1 (4): 204–12. https://doi.org/10.1107/S2052252514010070.

Stevenson, H. P., A. M. Makhov, M. Calero, A. L. Edwards, O. B. Zeldin, I. I. Mathews, G. Lin, et al. 2014. "Use of Transmission Electron Microscopy to Identify Nanocrystals of Challenging Protein Targets." *Proceedings of the National Academy of Sciences* 111 (23): 8470–75. https://doi.org/10.1073/pnas.1400240111.

Stevenson, H P, D P Deponte, A M Makhov, James F Conway, O B Zeldin, S Boutet, and G Calero. 2014. "Transmission Electron Microscopy as a Tool for Nanocrystal Characterization Pre- and Post-Injector Transmission Electron Microscopy as a Tool for Nanocrystal Characterization Pre- And." *Phil. Trans. R. Soc. B* 369: 20130322.

Streets M. Aaron Quake R. Stephen. 2010. "Ostwald Ripening of Clusters during Protein Crystallization." *Physical Review Letter* 104 (17). https://doi.org/10.1103/PhysRevLett.104.178102 TI.

Su, Xiao-Dong, Heng Zhang, Thomas C Terwilliger, Anders Liljas, Junyu Xiao, and Yuhui Dong. 2015. "Protein Crystallography from the Perspective of Technology Developments." *Crystallography Reviews* 21 (1–2): 122–53. https://doi.org/10.1080/0889311X.2014.973868.

Sumner, James Batcheller. 1948. "The Chemical Nature of Enzymes." *Journal of the Washington Academy of Sciences* 38 (4): 113–17. http://www.jstor.org/stable/24530953.

Svergun, D I, and M H J Koch. 2003. "Small-Angle Scattering Studies of Biological Macromolecules in Solution." *Reports on Progress in Physics* 66 (10): 1735–82. https://doi.org/10.1088/0034-4885/66/10/R05.

Taylor, Garry L. 2010. "Introduction to Phasing." *Acta Crystallographica Section D: Biological Crystallography* 66 (4): 325–38. https://doi.org/10.1107/S0907444910006694.

Tenboer, Jason, Shibom Basu, Nadia Zatsepin, Kanupriya Pande, Despina Milathianaki, Matthias Frank, Mark Hunter, et al. 2014. "Time-Resolved Serial Crystallography Captures High-Resolution Intermediates of Photoactive Yellow Protein." *Science* 346 (6214): 1242–46. https://doi.org/10.1126/science.1259357.

Vekilov, Peter G. 2010. "The Two-Step Mechanism of Nucleation of Crystals in Solution." *Nanoscale* 2 (11): 2346–57. https://doi.org/10.1039/c0nr00628a.

Weierstall, Uwe, Daniel James, Chong Wang, Thomas A White, Dingjie Wang, Wei Liu, John C H Spence, et al. 2014. "Lipidic Cubic Phase Injector Facilitates Membrane Protein Serial Femtosecond Crystallography." *Nature Communications* 5 (3309). https://doi.org/10.1038/ncomms4309.

Weinert, Tobias, Natacha Olieric, Robert Cheng, Steffen Brünle, Daniel James, Dmitry Ozerov, Dardan Gashi, et al. 2017. "Serial Millisecond Crystallography for Routine Room-Temperature Structure Determination at Synchrotrons." *Nature Communications* 8 (1). https://doi.org/10.1038/s41467-017-00630-4.

Weiss, Manfred S. 2001. "Global Indicators of X-Ray Data Quality." *Journal of Applied Crystallography* 34 (2): 130–35. https://doi.org/10.1107/S0021889800018227.

References

White, Thomas A., Richard A. Kirian, Andrew V. Martin, Andrew Aquila, Karol Nass, Anton Barty, and Henry N. Chapman. 2012a. "CrystFEL: A Software Suite for Snapshot Serial Crystallography." *Journal of Applied Crystallography* 45 (2): 335–41. https://doi.org/10.1107/S0021889812002312.

White, Thomas A, Richard A Kirian, Andrew V Martin, Andrew Aquila, Karol Nass, Anton Barty, and Henry N Chapman. 2012b. "{\it CrystFEL}: A Software Suite for Snapshot Serial Crystallography." *Journal of Applied Crystallography* 45 (2): 335–41. https://doi.org/10.1107/S0021889812002312.

Wilson, W. William. 1990. "Monitoring Crystallization Experiments Using Dynamic Light Scattering: Assaying and Monitoring Protein Crystallization in Solution." *Methods* 1 (1): 110–17. https://doi.org/10.1016/S1046-2023(05)80154-9.

Wilson, W William. 2003. "Light Scattering as a Diagnostic for Protein Crystal Growth—A Practical Approach." *Journal of Structural Biology* 142 (1): 56–65. https://doi.org/https://doi.org/10.1016/S1047-8477(03)00038-8.

Wilson, William W., and Lawrence J. Delucas. 2014. "Applications of the Second Virial Coefficient: Protein Crystallization and Solubility." *Acta Crystallographica Section F:Structural Biology Communications* 70 (5): 543–54. https://doi.org/10.1107/S2053230X1400867X.

Winn, Martyn D., Charles C. Ballard, Kevin D. Cowtan, Eleanor J. Dodson, Paul Emsley, Phil R. Evans, Ronan M. Keegan, et al. 2011. "Overview of the CCP4 Suite and Current Developments." *Acta Crystallographica Section D: Biological Crystallography* 67 (4): 235–42. https://doi.org/10.1107/S0907444910045749.

Wlodawer, Alexander, Zbigniew Dauter, and Mariusz Jaskolski. 2017. *Protein Crystallography: Methods and Protocols*. Vol. 1607. Humana Press. https://doi.org/10.1007/978-1-4939-7000-1.

Yamashita, Keitaro, Naoyuki Kuwabara, Takanori Nakane, Tomohiro Murai, Eiichi Mizohata, Michihiro Sugahara, Dongqing Pan, et al. 2017. "Experimental Phase Determination with Selenomethionine or Mercury-Derivatization in Serial Femtosecond Crystallography." *IUCrJ* 4 (5): 639–47. https://doi.org/10.1107/S2052252517008557.

Yamashita, Keitaro, Dongqing Pan, Tomohiko Okuda, Michihiro Sugahara, Atsushi Kodan, Tomohiro Yamaguchi, Tomohiro Murai, et al. 2015. "An Isomorphous Replacement Method for Efficient de Novo Phasing for Serial Femtosecond Crystallography." *Scientific Reports* 5 (14017). https://doi.org/10.1038/srep14017.

Yaoi, Mari, Hiroaki Adachi, Kazufumi Takano, Hiroyoshi Matsumura, Tsuyoshi Inoue, Yusuke Mori, and Takatomo Sasaki. 2004a. "Effect of Stirring Method on Protein Crystallization." *Japanese Journal of Applied Physics, Part 2: Letters* 43 (10 A): 1318–20. https://doi.org/10.1143/JJAP.43.L1318.

———. 2004b. "Effects of Solution Stirring on Protein Crystal Growth." *Japanese Journal of Applied Physics* 43 (No. 5B): L686–88. https://doi.org/10.1143/JJAP.43.L686.

Zander, Ulrich, Gleb Bourenkov, Alexander N. Popov, Daniele De Sanctis, Olof Svensson, Andrew A. Mccarthy, Ekaterina Round, Valentin Gordeliy, Christoph Mueller-Dieckmann, and Gordon A. Leonard. 2015. "MeshAndCollect: An Automated Multi-Crystal Data-Collection Workflow for Synchrotron Macromolecular Crystallography Beamlines." *Acta Crystallographica Section D: Biological Crystallography* 71: 2328–43. https://doi.org/10.1107/S1399004715017927.

References

Zhang, Fajun, Felix Roosen-Runge, Andrea Sauter, Roland Roth, Maximilian W. A. Skoda, Robert M. J. Jacobs, Michael Sztucki, and Frank Schreiber. 2012. "The Role of Cluster Formation and Metastable Liquid—liquid Phase Separation in Protein Crystallization." *Faraday Discussions* 159: 313. https://doi.org/10.1039/c2fd20021j.

Zhang, Yang, and Jeffrey Skolnick. 2005. "TM-Align: A Protein Structure Alignment Algorithm Based on the TM-Score." *Nucleic Acids Research* 33 (7): 2302–9. https://doi.org/10.1093/nar/gki524.

Zulauf, Martin, and Allan D'Arcy. 1992. "Light Scattering of Proteins as a Criterion for Crystallization." *Journal of Crystal Growth* 122 (1): 102–6. https://doi.org/https://doi.org/10.1016/0022-0248(92)90232-8.

Acknowledgements

I would like to firstly thank my academic supervisor, Prof. Dr. Christian Betzel, for his coordination and support throughout all stages of my doctorate program. The completion of my project would not have been possible without his academic support and his kind generosity for working in his laboratory and providing the necessary items involved in my scientific work. Equally, I would like to thank my co-supervisor PD Dr. Markus Perbandt for his scientific input and challenging conversations during my stay in this group. Thank you, Petra Belda, for always helping us with all the necessary things needed for conducting our work in the labs. Furthermore, I would like to thank Prof. Henning Tidow for the evaluation of my thesis and to the examiners committee of my defense presentation.

I would like to express my gratitude towards the company Xtal Concepts GmbH for choosing me for the Early Stage Researcher position via the EU Horizon 2020 Research and Innovation program under the Marie Sktodowska-Curie action "X-Probe". In particular, I would like to thank Dr. Arne Meyer, Dr. Karsten Dierks and Dr. Annette Eckhardt for all the constructive scientific discussions and input throughout my work. Furthermore, I would like to thank them for providing the company's equipment that was highly significant for my scientific research. At the same time, I would like to thank Saskia Flindt for all the administrative work and to Katarzyna Siewinska and Vinay Jois for being wonderful colleagues.

A special and warm thanks to Dr. Sabine Botha for all the advice and scientific discussions that contributed to the quality of my work. Your help during data collection and afterwards was highly significant to my final results. Thank you, Dr. Robin Schubert for introducing me to the topic of my thesis and providing help when needed. Especially, I would like to thank you both as well as to Tina Schmidt and Theresa Nuguid for all the kind support and friendship that lifted me throughout the hard times.

I would like to thank Prof. Dr. Jan Pieter Abrahams for offering me the possibility to investigate my samples with the cryo – electron microscopy technique