BATCH CRYSTALLISATION OF LYSOZYME AND CATALASE WITH THE AID OF SILICA NUCLEANTS

BY KEI KWAN CLARENCE CHUM

A dissertation submitted to Imperial College London for the degree of Doctor of Philosophy

SUPERVISORS: DR JERRY HENG (ACADEMIC)

DR DARYL WILLIAMS (ACADEMIC)

DR TIM HO (INDUSTRIAL)

DEPARTMENT OF CHEMICAL ENGINEERING, IMPERIAL COLLEGE LONDON, SOUTH KENSINGTON, LONDON, SW7 2AZ

Declaration

The work in this thesis is that of the author, Kei Kwan Clarence Chum, and was carried out at Imperial College London between the dates of March 2013 and August 2016, and at GlaxoSmithKline R&D, Stevenage between the dates of September 2016 and Feb 2017. Except where otherwise acknowledged and referenced, the material shown in this thesis is original work. No part of this thesis has been submitted for a degree at any other university.

Statement of Collaboration

The subject and theme of this thesis are a collaboration between the Department of Chemical Engineering and Technology at Imperial College London and the Particle Sciences and Engineering research group at GlaxoSmithKline (Hertfordshire, UK), and are referred to in this agreement as 'the Work'.

Copyright Declaration

Acknowledgement

The past four years have been a challenging experience for me. I would like to thank everyone who has offered me advice and guidance and have directly contributed to my project: Dr Jerry Heng, Dr Daryl Williams and Dr Tim Ho for your supervision, ideas input and advice on the project.

This project is funded by BBSRC and GSK. It has been a great honour for me to receive the CASE studentship for this work and I am very grateful of this opportunity. Also, I would like to thank the Particle Sciences and Engineering team at GSK R&D, Stevenage for the generosity and patience in granting me the opportunity to use their facilities. Not only did it have a positive impact on my research, it has also assured my career choice in the future.

I would also like to thank members of the Surfaces and Particles Engineering Laboratory (SPEL) who have contributed to my research directly: I would like to thank Umang for guiding my experimental work when I first started, and Efty for offering ideas, and Mark and Sabiyah for our exchange of experiences regarding protein crystallisation. I'd like to thank my industrial supervisor, Tim, for helping me arrange the placement at GSK, for always being available for discussions and for checking that I was 'winning' every day.

And also those who provided me emotional support throughout these years. They include the friends I've made in SPEL and in the department who made me feel like I belong, I cannot ask for better company to go through this roller coaster ride with, especially Naima, you have seen the best and worst of me. And also Miao, for being the best ever role model any aspiring researcher can have, your strength, positive outlook and determination is simply inspiring. I'd like to thank my friend Gary for always being there for me through almost 10 years of friendship, for telling me off when necessary, and also because you acknowledged me in your thesis.

Lastly I would like to thank my family, who have such great patience for me while I spent these four years of being away from home so often, and for tolerating my temper during the writing up process. I would like to thank my parents for nurturing me for all these years, and the sacrifices you made for my education that I am in this position today. I would also like to thank my brother for taking up Dentistry so I don't have to.

Abstract

There is a great interest in the use of heterogeneous seeds on the crystallisation of proteins, with most studies on the ability to obtain diffractive quality crystals in vapour diffusion. Reports by Shah et al (2011) developed a relationship between surface pore diameters of these seeds and the size of proteins for effective protein crystallisation. When there is a match between these properties, a significant reduction in induction time can be achieved with vapour diffusion experiments. Heterogeneous seeds of such size match are named 'nanotemplates' by Shah and co-workers.

This thesis aims to further understand the effects of the porous properties on the nucleation behaviour of proteins. Specifically, the effects in batch crystallisation will be studied, in which both the thermodynamics and kinetic pathways are expected to be significantly different. In order to achieve this, the induction time was used as a property to represent the nucleation process for proteins and was monitored using both UV-vis spectroscopy and turbidimetry.

Hen Egg White Lysozyme (HEWL) and Bovine Liver Catalase (BLC) are used as model proteins for the studies. Mesoporous heterogeneous seeds used are of pore diameters 5.5 and 9.0, and non-porous seeds are also used for comparison. These seeds would be represented as NT40, NT120 and NP respectively throughout the thesis. The protein hydrodynamic diameter and the pore diameters were characterised by Dynamic Light Scattering and Nitrogen Adsorption techniques respectively prior to crystallisation studies.

For studies involving HEWL, batch crystallisation was applied directly due to the vast details of thermodynamic behaviour available. Studies with lysozyme also serve as a purpose to establish a basic understanding of different parameters (seed concentration, seed type, supersaturation, stirring), which can be applied to the crystallisation of BLC. However, studies regarding BLC required the determination of solubility prior to the studies involving the use of seeds, and experiments were demonstrated at different volumes. In these studies, different seed concentrations were investigated to a) identify a suitable seed concentration range for heterogeneous seeding and b) investigate the varying amount of surface properties (pore diameter, surface area). Different protein concentrations (13.5 − 17.5 mg/mL) was also compared due as heterogeneous surfaces are expected to have different extent of contribution.

For the crystallisation of HEWL, the use of seed reduced the induction time compared to unseeded experiments. Unexpected results were obtained where NT40, the seed with pores

of closest match to the protein, resulted in the longest induction time across all seed concentration levels tested (0.05 -2.00 mg/mL), and NP resulted in the shortest induction time, especially at low seed concentrations (0.05 and 0.1 mg/mL). NT120 gave an intermediate result between the two.

For all seed types used, it appears that the induction time is constant at certain ranges of seed concentration, and the values for these ranges varies for different seed types. In other words, the induction time dependence on amount of seeds (hence amount of surface properties and pore volumes) is not linear for the seed concentration ranges tested. As expected, the effects of heterogeneous seeds are more significant at reduced protein concentrations. At 13.5 mg/mL, the induction time for lysozyme was over 10 hours when unseeded, the use of nanotemplates (1.25 mg/mL) reduced the induction time by ~75%, and the use of other seeds at the same seed concentration reduces the induction time to by ~80 %.

Conditions in which crystallisation of BLC occurs was identified. As there are limited information available in literature on the solubility of catalase, this property was determined for the protein at 20 °C, in solution conditions of precipitant concentration 8-10% PEG 4000, 5 % MPD in KPO_4 (50 mM) buffer at pH 7, in which crystallisation experiments were conducted for this protein, and was found to be ~ 10.5 mg/mL. Based on this, crystallisation was conducted at 13.4 mg/mL and 17.5 mg/mL at 8 % PEG (m/v), which corresponded to supersaturation ratios 1.28 and 1.67 respectively. Stirred batch crystallisation experiment was conducted at 0.5 mL and 20 mL scale. At 0.5 mL, even at low stirring, nucleation occurred instantaneously at a seed concentration of 1 mg/mL. At 20 mL working volume and a stirring rate of 100 rpm, there was a slight increase in induction time compared to the 0.5 mL experiments.

At this scale, experiments were conducted at BLC concentrations at 17.5 mg/mL and 13.4 mg/mL, and a range of seed concentrations were investigated. Non-porous material results in a slight decrease in induction time at 17.5 mg/mL, while at 13.4 mg/mL, the effects of NT 120 were more prominent at high seed concentration (>2.0 mg/mL). Demonstrating the different surface property contributes differently to the influences on crystallisation induction time of protein at different supersaturations.

It is shown that batch crystallisation of proteins at batch scale (~20 mL) is achievable, and that the use of heterogeneous nucleants assisted in the nucleation/crystallisation of proteins. Results presented in this thesis highlights the relationship between nucleant concentrations

used and its effect on induction time, and also the supersaturation dependency of the effects of nanopores compared to non-porous surfaces.

Table of Content

List of Figures

List of Tables

List of Symbols

S	Solubility
S_0	Solubility of protein in the absence of precipitants
K_s	Salting out constant
m	Concentration of precipitant
ΔG	Free energy change of nucleation
ΔG_s	Surface free energy of nuclei formation
ΔG_v	Volume free energy of nucleation
r	Radius of cluster
r^*	Critical nuclei radius
A	Surface area of nucleation cluster
γ	Molar surface free energy per unit area
μ	Chemical potential
T	Temperature
σ	Supersaturation ratio
v	Volume
ΔG^*	Nucleation energy barrier
J	Nucleation Rate
κ_0	Pre-exponential factor for the calculation of nucleation rate
t_{nuc}	Nucleation time
t_{growth}	Time for crystal nuclei to grow into a detectable size
t_{ind}	Induction time
a	Length of a crystal nucleus treated as a cube
Ω	Volume of a nucleus (as opposed to cluster) used in two-step nucleation
ϕ	Electrostatic potential of substrate surface
θ	Contact angle between nuclei clusters and substrate surface
ΔG^*_{hetero}	Energy barrier to heterogeneous nucleation
β	Geometrical parameter of a pore as nucleation site for crystallisation
ΔG_{pore}	Energy barrier to nucleation within pore
A_c	Surface area of interface between nuclei cluster and bulk phase
Af	Surface area of interface between nuclei cluster and porous substrate
dM/dT	Mass deposition rate for crystal growth
ρ_c	Crystal density
f_v	Volume shape factor for crystal growth
L	Length of crystal
k_d	Diffusion rate constant for crystal growth
k_r	Surface reaction rate constant in crystal growth
k_g	Overall rate constant
c_i	Concentration of solute at interface between crystal and bulk

ω	Surface roughness factor
A_{280}	UV-vis absorbance at 280 nm
τ	Turbidity
N	Concentration of solid particles
Q_{ext}	Mie Scattering extinction coefficient

List of Abbreviations

HEWL	Hen Egg White Lysozyme
BLC	Bovine Liver Catalase
CNT	Classical Nucleation Theory
BET	Brunauer-Emmett-Teller (surface area determination model)
BJH	Barrett-Joyner-Halender (surface pore volume determination model)
PEG	Polyethylene Glycol
MPD	2-Methyl-2,4-pentanediol

1. Introduction

1.1 Overview

Biopharmaceuticals are high molecular weight, complex biomolecules with a structure difficult to define that are derived from living organisms and generally cannot be produced by chemical synthesis,[1] and are often suffering from limited stability.[2] Biopharmaceutical development currently represents the largest and fastest growing portion of the pharmaceutical market globally.[3,4] Such growth would ultimately rely on the minimisation of production costs for manufacturing, which is dependent on both upstream and downstream processing.[5] Upstream processing demonstrated a huge improve over time, attributing to efficient fermentation and cell culture technologies.[6] On the other hand, downstream processing still remains the bottleneck of biopharmaceuticals manufacturing. Chromatographic techniques that are currently used can contribute up to 80% of the manufacturing cost.[5] Costs associated with the technique include the use of material, cleaning and validation and the time needed to be invested.[7]

Separation techniques alternative to chromatography have therefore regained attention from biopharmaceutical manufacturers. Among these techniques, the potential of crystallisation has been highly recognised (Figure 1).[8] Crystallisation is a commonly used technique for the separation and purification in different industries, such as pharmaceutical, food and petrochemical industries.[9,10] In addition to the reduction of energy and time consumption compared to chromatographic techniques, the advantage of this technique is associated with many attractive properties of crystalline products that also allows for the integration of formulation to the processing step.[2,11] In fact, protein crystallisation was used as an extraction technique to isolate proteins from animal tissues and plants before it found its application in crystallography.[12,13] These include the enhanced chemical and physical stability of the product form.[14,15] While the reason for the enhanced stability of protein within a crystal is uncertain, it was suggested to result from the defined positions of molecules within crystal lattices.[14,15] The process of protein crystallisation itself has been demonstrated to have a slight preference (40 kJmol^{-1}) over the native structure of proteins compared to their denatured state.[14] Crystallisation also results in high product purity, allowing for the high dosage, low viscosity purity.[16] This compares favourably with other delivery forms such as soluble and amorphous formulations, which results in product form that or of poor stability.[2,15,16]

To date, the main application of protein crystallisation is protein crystallography, in which the methods and techniques available aim to achieve crystals of sufficient size and

minimal defects.[17–19] From an industrial perspective, high product yield produced in short duration is favourable for manufacturing processes. However, a low concentration of biopharmaceutical compounds is typically acquired in steps prior to downstream process; such conditions would result in difficulties in achieving crystallisation at large scale.[2] Currently, insulin remains as the only proteins crystallised commercially and can be administered in crystalline form, owing to its small size, physical stability and its ability to refold upon exposure to organic solvents.[20] Limited studies have been conducted to develop process controls for all other proteins, as the main application of protein crystallisation to date is crystallography. In addition to this, the additives used to promote crystallisation would have to be non-toxic and inexpensive in large scale.[2,21] Should the protein crystals be resuspended in environment with pharmaceutically accepted excipients, these products with have to maintain both protein and crystalline stability.[2] Numerous complexities and limited knowledge are also associated with difficulties in protein crystallisation in general, which are discussed in more details throughout this report in Chapter 2. It is clear that there is a need to develop a greater understanding in protein crystallisation processes.
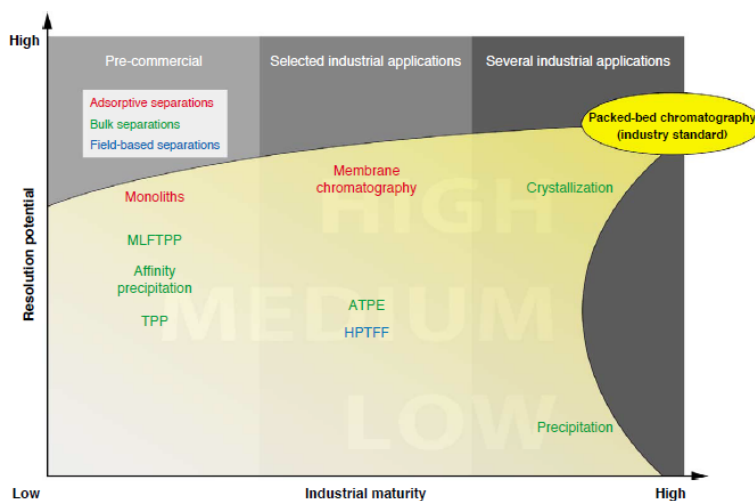
Figure 1 Graph showing the resolution potentials and industrial maturity of bioseparation processes.[8]

Despite the difference in requirement of crystals between the application of crystallography and industrial crystallisation of protein, knowledge can be obtained from the development of crystallisation process from crystallography: screening and optimisation is still required for the in the initial stage to determine the crystallisation conditions.[20] Also, various process designs have been proposed in order to simulate the thermodynamic pathways commonly adopted which would be discussed extensively in Chapter 2. Alternatively, heterogeneous seeding is employed as a technique to improve the likelihood of obtaining

protein crystals and was used commonly in the screening and optimisation of protein crystallisation.[22,23] The examples of possible materials to be used as heterogeneous seeds for proteins as well as the seeding strategies was discussed also in Chapter 2. In particular, the use of mesoporous materials has gained recognition in protein crystallisation.

Mesoporousity represents materials with accessible pores of a diameter of 2-50 nm.[24] In particular, silica has been used as a common material, owing to its chemical inertness and biocompatibility.[25] The use of this technique was first demonstrated by Chayen and co-workers (2001), attributing this effects to the disorderliness of the pores, making the materials attractive as a universal nucleant.[26] The need for disordered pores to achieve protein crystallisation has inspired theoretical studies on the heterogeneous nucleation within pores, which will be highlighted in Section 2. The efficiency of these materials was due to the presence of an 'optimal pore diameter' for any protein of interest.[27]

Seeing the effects of pore size on protein crystallisation as a potential method of selective crystallisation of proteins, Shah and co-workers (2011) developed a range of mesoporous silica of different specifically designed pore sizes.[28]The possibility of using a soft templating approach involving the formation of hexagonal tubular micelles using tri-block copolymers allowed for a a narrow pore size distribution to be achieved.[29] Using these materials, Shah and co-workers have developed a relationship between the hydrodynamic diameter of the protein of interest and the pore diameter. Mesoporous silica of pore diameters matching the diameter of proteins are referred to as 'nanotemplates', and were demonstrated to be most efficient in achieving protein crystallisation. This was further extended to incorporate the effects of surface chemistry, further enhancing the effect.[30]

1.2 Research Hypothesis

The pore diameter in relation to diameter of protein is an important factor in influencing the nucleation behaviour of protein in batch crystallisation process. With the use of silica based nucleants of varying pore diameter in batch crystallisation systems (~20 mL) of proteins, the nucleation behaviour of protein would vary. It is also expected that the crystallisation behaviour of protein would vary from what was understood from Shah et al (2011) due to the different thermodynamics and kinetics pathway involved in batch scale.

1.3 Research Objectives

This project aims to further investigate the role of nanotemplates in batch crystallisation of proteins, specifically on the nucleation process. Shah et al (2011) demonstrated these effects

with vapour diffusion, which follows different thermodynamic and kinetic pathways. To demonstrate the nucleation process, the project focuses on using induction time to quantify the behaviour of heterogeneous seeds used.

In the experiments conducted, the effects of the presence of silica particles with different pore diameters were compared against non-porous seeds. Due to the huge difference in surface properties (e.g. pore diameter, surface area) expected, in order for this to proceed, we first investigated the effects of different seed concentrations on porous and non-porous seeds, and the performance of these materials are compared across a wide range of seed concentrations. These results also identify the seed parameters for batch crystallisation of proteins using porous nucleants. The effects of these seeds are also varied under different protein concentrations, as heterogeneous surfaces are expected to have different contributions at different supersaturations.

The specific objectives are:

i) Compare the effects of different seed types to observe any difference in the induction time resulted from the crystallisation processes. These seed types different in porosity. Nanotemplates were compared against the performance of other mesoporous silica as well as non-porous silica.

ii) Demonstrate the effects of seed concentration to identify a suitable range at which these seeds are effective, and also to compare the effects of variation of surface area vs pore volume.

iii) Demonstrate the effects of seeds at different protein concentration to compare the performance of the seeds at conditions in which heterogeneous surfaces contribute differently to the reduction of crystallisation energy barrier. This can determine whether there are other factors such as protein supersaturation dependencies on the role of these surfaces.

2. Background

2.1   The Crystalline Form

Crystals are solids that consist of ordered three dimensional structures of molecular assemblies uniform in all direction within a lattice, and is defined by a repeating unit known as unit cells. Ideal crystals would display and infinite and perfect arrays of these repeating units. However, real crystals often contain dislocations, causing the arrangement to deviate from this. Solids that do not display these crystalline properties are amorphous materials.[17,31,32]

Each unit cell has its own dimensions and symmetry. Symmetry elements can be reflection, rotation, inversion and/or rotational inversion. It has been established by Bravais that there can be no more than 14 point lattice that can be constructed, each with a characteristic shape and symmetry elements. The internal order in crystal lattices results in defined shapes for crystals. Miller indices are used to define the planes as one goes across a crystal lattice depending on the point of intersection of the plane at a set of axis along the lattice.[32]

2.2   Proteins

2.2.1 Structure of Protein

Amino acids are the building blocks of protein molecules. These are chiral molecules of a side group, a carboxylic acid and an amine group. Amino acids are defined by their side chains, of which its physicochemical properties are classified into different types: acidic, basic, aromatic and hydrophobic. Amino acids are linked via peptide bonds, which is formed from the condensation reaction between the carboxylic acid and amine group. There are 20 known amino acids, giving endless possible combinations for proteins.[33]

Proteins are defined by its sequence (primary structure) of amino acids linked via peptide bonds and the higher order structures (secondary, tertiary, quarternary) that is held together by a balance of intramolecular interactions across the large chains of sequences. The secondary structure is mainly dominated by the effects of hydrogen bonding of the peptide hydrogens, forming either helical or sheet like structures depending on the degree of rotation of the peptide bonds. A combination of intramolecular interactions (electrostatic, hydrophilic/hydrophobic, Van der Waal, and hydrogen bonding) contributes to the folding of a polypeptide chain, i.e. the tertiary structure. These polypeptides can exist on their own as a protein monomer, or as a subunit of a functional oligomeric protein, the assembly of these polypeptides corresponds to the quarterary structure. The overall 3D structure contributes to the functionality of the proteins for protein- substrate interactions. When a protein is folded

into the structure that enables its biological activity, it is said to be in its 'native' or 'active' state.[34,35]

Protein molecules are divided into three main types according to their structures, fibrous, globular and membrane proteins. Most crystal structures published in the PDB are globular proteins as they perform most of the chemical functions in life forms. These proteins are more compact, and as the name suggests, spherical. Unlike the other two classes of proteins, the active globular proteins are soluble in water (the 'native solvent'): In the native state, the exposed surface of the protein consists mainly of hydrophilic amino acid residues, buried in the core of proteins are dominantly hydrophobic amino acid residues. Fibrous proteins are relatively simple, elongated structure and not widely explored by crystallisation scientists. Membrane proteins, which have received a large attention from protein crystallisation scientists in recent years, make up to ~25% of all proteins. They are usually found in the lipid bilayers of cell membranes, consisting of both hydrophilic and hydrophobic environments. Therefore, both hydrophobic and hydrophilic components of the protein are exposed in different ways, depending on the location of the protein with respect to the bilayer, giving poor salvation behaviour in polar or non-polar solvents.[33]

Due to the presence of charged amino acid side chains, a protein molecule is typically represented with an overall charge across the exposed surface of a protein molecule. The overall charge is subjected to the pH of the environment, in which the proteins may protonate/deprotonate.

2.2.2 Protein Stability and Aggregation

While the instability of small molecules only attributed to its susceptibility of chemical changes, handling of protein also requires the consideration of physical instability.[36] The dependence of protein structures on these weak non-covalent interactions is the cause of the difficulties in maintaining the stability of these proteins. Jeong (2012) summarised the contribution to the stabilisation energy relative to the entropic driving force to unfolding, and that the net free energy of the folded state is ~5-20 kcal/mol. This small amount of energy is sufficient in causing the protein structures to deviate from the active state; this includes both chemical and physical instability.[35] While it is important to select conditions where proteins are chemically stable, for the interest of crystallisation and its application to the biopharmaceutical formulations, the physical instabilities would be the focus of the following discussion. When physically destabilised, proteins can undergo denaturation and aggregation.
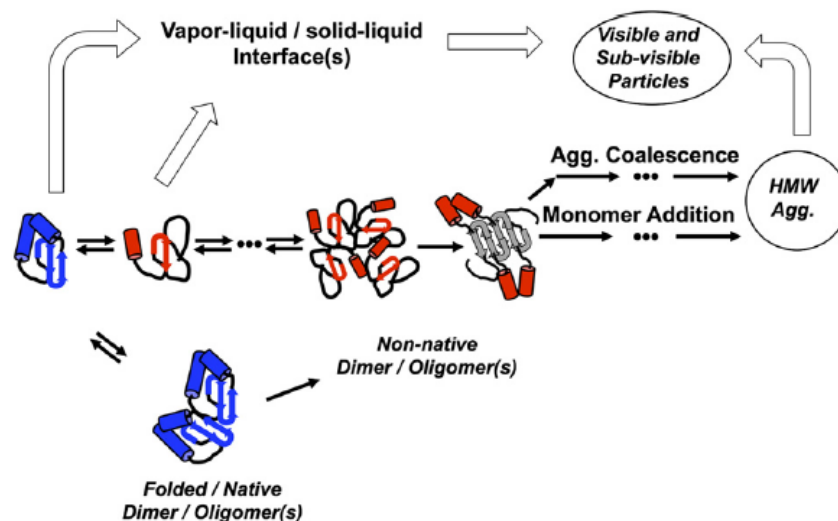
Figure 2 Possible routes of Protein Aggregation (image obtained from [[37]])

The possible aggregation pathway can be summarised in Figure 2. As evident, it can undergo both native and non-native aggregation. The presence of interfaces (container wall, water/air interface, presence of air bubbles) can cause the disruption of structures by encouraging the exposure of the hydrophobic group. The preparation of high protein concentration is recommended for the storage of protein solution, as the protein-protein interaction would be more significant. However, when the concentration exceeds a certain extent, native aggregation can occur. These aggregations can be reversible or irreversible, irreversible aggregations result in the misfolding of protein.[35,36,38] The recommended storage concentration is 5-25 mg/mL for the purpose of protein crystallisation as it promotes sample stability and homogeneity.[39]

Intramolecular forces can be affected by temperature, pH, other solution environment condition e.g. ionic strength of components, which defines the working range of this variable on all protein studies including crystallisation.[36,38]

2.2.3 Properties of Protein Crystals and Characteristisation

The non-specificity and the low strength of intermolecular interactions compared to weight of these macromolecules, as well as the shape irregularity of protein molecules is the cause of the difficulties for protein to self assemble into an ordered 3D lattice structure, which causes the network in a protein crystal to be weak.[40–45] This contributes to the physical and mechanical properties that are highlighted below.

Protein crystals are typically smaller, typically 0.1 mm for preliminary structural studies.[17] As the typical distance between bound atoms are 1.5 – 2 Å apart,[46] in order to solve

the protein structure, a resolution of 3.5 Å is typically required,[17] and this usually requires crystals of 0.3 – 0.5 mm in length.[47] However, obtaining diffraction quality crystals for the study of novel protein remains a challenge.

Water is reported to contribute significantly to the mass of protein crystals. It is reported that water occupies 50-70% of protein crystals.[40,48–50] The presence of water molecules as part of the protein crystal structure has been observed in crystallography, demonstrating that water molecule is incorporated in crystalline lattice.[51] In addition to this, solvent channels are common within protein crystals, allowing the occupancy of solvent molecules in these interstitial spacings. In these conditions, water molecules are not only adsorbed, but present as liquid phase.[17,18] The presence of these channels reduce greatly the crystal density.[17] However, the water present within crystals also contribute to the crystal structural integrity, as removal of water causes the collapse of crystal.[49,50,52]

Due to the weaker interactions between molecules within lattice, protein crystals typically exhibit poorer mechanical properties, they are soft and easily crushed. The presence of water also contributes to the softness.[40,44,49]

Characterisation

The size of proteins causes the crystal unit cells to be large. This, in addition to the typical size of crystals makes lab scale X-ray Powder Diffraction an unsuitable technique for structural characterisation. Structural studies rely on single crystallography with a synchrotron beam. Careful preservation of crystals is therefore required.[47]

Prior to such characterisation, most protein crystallisation scientists typically rely on qualitative assessments of crystal properties. The assessment typically is to determine whether solids observed are crystalline, and whether the crystalline material observed are proteins. This include mechanical crust test as protein crystals are known to be much softer than crystals of small molecules.[50,52] Dyes are also available to mix with crystallisation droplets. The solvent channels within protein crystals allow the accommodation of the dye molecules, causing the staining of crystals, and therefore distinguishing a protein crystal from salt crystals.[47,49] Publications are also available as guidance for the visual examination of crystals.[50]

2.3 Phase Behaviour in Crystallisation

2.3.1 Solubility and Supersaturation

Supersaturation is the driving force of crystallisation.[31,32,53–55] When a solute is being dissolved, a solvent can hold a certain concentration of the chemical as a solute while existing as a homogeneous phase, and this concentration is the solubility.[32] When a solution reaches the concentration, it is in equilibrium. Above the solubility, the solution is supersaturated. As it is not in equilibrium, the system expels excess solute molecules as a new solid phase, re-establishing the solution equilibrium.[32,40,53,56] The supersaturation of a solution is commonly given as a ratio of the concentration of solute in solution to the solubility.[32]

Solubility is affected by different factors: In general, the solubility is higher with increased temperature.[53] The ability of a solvent to solvate a molecule also depends on the interaction between solute and solvent molecules, which in turns is affected by the intrinsic properties of the solute and the solvent.[31] For example, in the case of small organic molecules, where the crystallisation environment can be either water or organic solvents, or a combination of both, solubility varies in different solvent or different mixtures.[32] However, in the case of proteins, in which water is the native solvent,[51] additives are used to tune the solvent/solute interaction to drive crystallisation, and are sometimes called precipitants, variation in precipitant and concentrations used affects the ionic strength, dielectric constant and polarity of solvents. The role of these precipitants would be explained in more detail in 2.3.2.[17,18]

To control the crystallisation process, phase diagrams are crucial in illustrating the thermodynamic behaviour of protein in a given set of condition. In these diagrams, most commonly presented in two dimensions, the state of a material is represented as a function of its solute concentration and one other parameter representing the crystallisation environment.[32] In the case of protein crystallisation, the phase diagrams represent phases in terms of protein and precipitant concentration.[56,57]

2.3.2 Phase Diagram

As evident from the example phase diagram, four different zones representing four different phases are present, separated by three curves. The solid line represents the solubility curve, indicates the equilibrium conditions. Below which is the undersaturated zone, where protein remains dissolved in solution. The supersaturated region lies above the solubility curve, and is further divided into three zones that represent denser phases than the undersaturated protein solution phases; the metastable, nucleation and precipitation zone.[18,32,56] In the metastable zone, the concentration of protein is sufficiently high to result in a growth in volume

of the solid phase, but low enough that the possibility of nucleation is infinitely small, even upon agitation[58,59] This zone was first introducd by Ostwald[59]. Nucleation is possible only at a higher supersaturation level, which occurs in conditions represented in the nucleation zone (sometimes represented as the labile zone) marked by the supersolubility (dash line) and precipitation curve (dotted line). The energy of systems within these conditions are sufficiently high that a new phase is created as it changes to a lower energy state to achieve stability.[60] When supersaturation exceeds the precipitation curve, uncontrolled nucleation occurs and amorphous precipitates are obtained.[53,56]
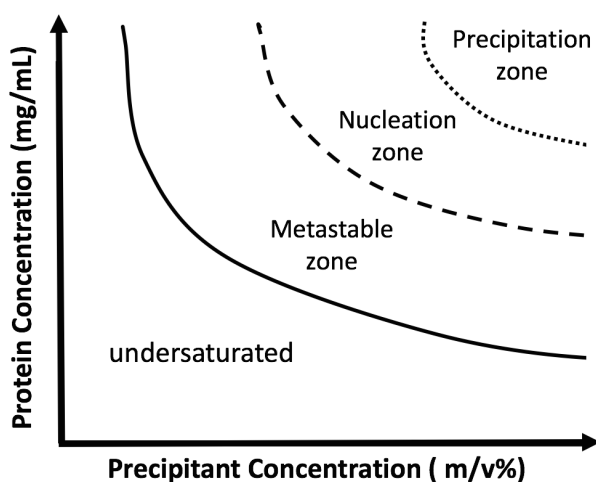
Figure 3 Typical phase diagram for protein crystallisation systems[56]

2.3.2   Factors Affecting Protein Solubility

To drive the protein crystallisation condition to supersaturation, crystallising agents are typically added to the native solvent of protein, i.e. water. These crystallisation agents, also referred to as precipitants, are typically either salts, polymers and non-volatile organics and are commonly used to drive crystallisation.[17,18,40] These precipitants alter the protein solubility in water via different mechanism, and has been suggested that the type of precipitant used for specific protein crystallisation is not interchangeable with another.[61] Currently, there exist no rule that determines the choice of precipitant to be used for a species of protein, predicting the type and amount of precipitant required to successfully drive the crystallisation of proteins remains difficult. A typical protein crystallisation trial (discussed in more detail) would therefore consist of an exhaustive screening of conditions of additives combination at different temperature, concentrations and pH.[49,62]

Ionic salts are commonly added to crystallisation to alter the ionic strength of water, and this method is known as the *salting-out effect*.[12,40,63–66] The ability of the ions within the

salts to influence protein solubility depends on how kosmotropic the ions within the salts are, and is represented by the Hofmeister series.[67] In the series the ions are ordered depending on their ability to make or break structures of water molecules. The interaction between water and kosmotropic ions are stronger than water-water interactions, these ions can therefore disrupt the hydrogen bonding network in the hydration shell of proteins. However, certain ions have the capability to denature proteins; examples are I^- and SCN^-.[68] Dumetz et al (2007) reported that ammonium sulfate is one of the best salting out agents by considering the effects of these ions on protein-protein interaction, and this salt was in fact used most widely in the early days of protein precipitation processes.[69] However, sulfate ions have a strong tendency to bind to lysozyme, therefore inhibiting crystallisation for the case of this protein. It was later demonstrated that these salts do not have an effect on bulk water but directly the water molecules in hydration shells. However, there are cases in which the increase in salt concentration also increases the protein solubility, this is known as the *salting-in effect*.[65] This occurs when the ionic strength of solution is sufficiently weak (low salt concentration or presence of chaotropic ions) such that binding between ions and proteins occur. It has been reported to cause the denaturation of protein sometimes.[70]

Water soluble polymers typically occupy a large space within the solution, therefore forcing the protein out of solution phase via a volume exclusion effect.[18] It is also known to reduce the solvent polarity. Typically, polyethylene glycols (PEGs) are used as they are hydrophilic and non-toxic. While Polson et al (1964) demonstrated an increasing effect in PEGs with increasing molecular weight,[71] McPherson (1976) demonstrated that lower PEGs concentration is required for lower weight polymers to drive crystallisation of various model proteins for PEGs weighing between 1000-6000 g/mol. The sizes of these proteins also have an added advantage of being excluded in solvent channels and also not incorporated into the crystal lattice. A well supported trend suggested that larger proteins require less PEG to crystallise.[72] Atha and Ingham (1981) rationalised this as a volume exclusion effect as the proteins are sterically hindered from the solvent molecules interacting with the PEGs, and observed a linear correlation between protein solubility and PEG concentration. The higher the concentration and the larger the polymer chain, the lower the protein solubility.[73]

Organic solvent changes the dielectric constant of water, causing the native solvent to solubilise proteins more poorly.[17,18]

The relationship between protein solubility and precipitants are related by Equation 1, where S is the solubility of protein at a given precipitant concentration, S_0 is the solubility of protein in the absence of precipitant, K_s is the salting out constant and m is the concentration of precipitant.[12] Variations in this equation has been quoted where the solubility is given as a function of the ionic strength of solution instead of the precipitant concentration.[63] This relationship was first explained by Cohn[12] based on the reported solubility of ovalbumin in the presence of ammonium sulfate at concentrations corresponding to the salting out region, but this has also been used to describe the effects of polymer precipitants as well.[74]

$$\log S = \log S_0 - K_s m \qquad (1)$$

A protein molecule is represented by an overall charge depending on the pH of solution. As both acidic and basic residues are found in protein structures, the alteration of pH would cause protonation and deprotonation of amino acids, altering the ionic strength of the protein itself.[12]

Theoretically, the solubility of a protein (pI) is lowest at the isoelectric point of protein, where there is no net charge on the molecule and the protein-protein repulsion is minimal.[12] The isoelectric point is not necessarily the optimum point to conduct crystallisation in, a correlation has been shown between the pI of the protein molecules that exist in the PDB and the reported pH at which crystals were obtained by Kantardjieff and Rupp (2004)[75] whereas negative charge is said to decrease the solubility of a protein.[74] The knowledge of the pI, which can be predicted from the sequence, however defines the crystallisation screening range of the proteins.

While temperature also has an effect on the protein solubility, there are challenges associated with conducting crystallisation at varying temperature as in the crystallisation of small molecules. A protein is associated with a temperature of denaturation, sometimes called melting temperature, and crystallisation below 0 °C in aqueous medium is also not possible. In typical crystallisation method such as vapour diffusion (discussed in more detail below), most crystallisation experiment is conducted in an incubator kept at isothermal conditions.

In most crystallisation experiments, different combinations of precipitants are used. Phase diagrams for protein crystallisation therefore consist of multiple parameters.

2.4   Nucleation

Mechanisms involved in crystallisation can be summarised as nucleation and growth. Nucleation is a process in which an embryo or a 'nucleus' of a new phase is created out of a homogeneous bulk phase, and is still regarded as a complicated process and is still not well understood nowadays as this involves a transition from a disordered to an ordered phase.[76–78] The nucleation process can be primary or secondary.[31] Primary nucleation occurs when nuclei are formed from a bulk phase spontaneously and can be homogeneous (in the absence of a surface) and heterogeneous (formed on a foreign surface), while secondary nucleation are nucleation events that depends on the prior presence of crystals of the solute.[32,77]

The probability of nucleation occurring is dependent on the free energy change of crystallisation. Two main theories are often cited to explain the nucleation phenomenon, these are the Classical Nucleation Theory and the Two Step Nucleation Theory.

2.4.1   Classical Nucleation Theory
The Energy Barrier

To quantify the concept of nucleation, the Classical Nucleation Theory (CNT) is one of the most important models, developed in the late 19th Century by Gibbs.[79,80] For phase transformation to occur, the free energy change of nucleation would have to be favourable ($\Delta G < 0$). This free energy can be separated into two components, a surface component (ΔG_s), which is unfavourable, and a volume component (ΔG_v), which is favourable, given the condition is in supersaturation.[54] This model states that when the thermodynamic condition is above saturation, clusters of solute molecules of radius r form and redissolve, and the free energy contributions can be expressed as a function of the nucleation radii, r, and that nucleation occurs only at a critical nuclei radius, r^*.
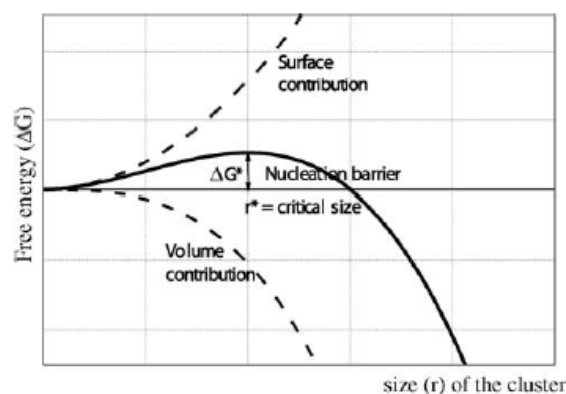
Figure 4 Free Energy curev highlighting the nucleation barrier at critical nuclei radius, r^*[81]

Like all phase transformation processes, the creation of a new phase requires the formation of an interface is associated with an energy cost (ΔG_s) which is related to the surface area of the cluster of solute molecules, *A* and the molar surface free energy per unit area, γ by Equation 2.[54]

$$\Delta G_s = A\gamma \qquad (2)$$

The formation is also associated with a volume free energy gain (ΔG_v) (Equation (2)) when the thermodynamic condition in which the formation of a new phase is favourable. This is dependent on the chemical potentials of both the solution and liquid phases, μ_α and μ_β respectively, which can be expressed in terms of their activities, *a* by Equation (4a) and (4b).[76]

$$\Delta G_v = -n(\mu_\alpha - \mu_\beta) \quad (3)$$

$$\mu_\alpha = kT\ln a_\alpha \quad (4a) \qquad\qquad \mu_\beta = kT\ln a_\beta \quad (4b)$$

ΔG_v is related to the supersaturation ratio, σ, given by a_α/a, of *n* moles of solute by Equation (5).

$$\Delta G_v = -n[kT\ln\sigma] \qquad (5)$$

Assuming a spherical geometry of these clusters with a radius of *r*, the resultant molar Gibbs' free energy of nucleation can be given by Equation (6).

$$\Delta G = -\frac{\frac{4}{3}\pi r^3}{v}kT\ln\sigma + 4\pi r^2\gamma \quad (6)$$

As evident from Figure 4, crystallisation occurs when the nucleation barrier at *r=r** is crossed at a given thermodynamic condition. At this point, the surface component is equal to the volume component of free energy. When *r<r**, it is more energetically favourable for the cluster to redissolve than to grow. When *r>r**, further incorporation of solute molecules would result in a decrease in free energy, and the new phase can therefore sustain and continue to grow into a macroscopic crystal. The critical nuclei radius is obtained as Equation (7) when the first derivative of Equation (5) is 0, from this the nucleation barrier is given as Equation (8).[81]

$$r* = \frac{2v\gamma}{kT\ln\sigma} \qquad (7)$$

$$\Delta G* = \frac{16\pi v^2\gamma^3}{3[kT\ln\sigma]^2} \qquad (8)$$

As evident in Equation (5), the other variable for a given solute would be the supersaturation, and the dependence of ΔG on this variable is shown in Figure 5. When the supersaturation is sufficiently high such that there is no nucleation barrier, nucleation is only mass transport limited.[82]

Based on the free energy for nucleation, one should be able to determine the nucleation rate, J via Equation (10), adopting the form of Arrhenius Equation (9) and can be visualised in Figure 6. κ_0 is a pre-exponential factor associated with the kinetics.[55,76,83] As evident in Figure 6, the nucleation frequency remains unchanged at a supersaturation ratio $\sigma < \sigma^*$. The time for nucleation to occur, τ is therefore an inverse of the nucleation rate, which can be linearised to Equation 11, where there is a linear relationalship between $\log\tau$ and $(\log^2\sigma)^{-1}$.[81]

$$J = \kappa_0 \exp\left(-\frac{\Delta G*}{kT}\right) \quad (9)$$

$$J = \kappa_0 \exp\left(-\frac{-16\pi\gamma^3 v^2}{3k^3 T^3 \ln(\sigma)}\right) \quad (10)$$
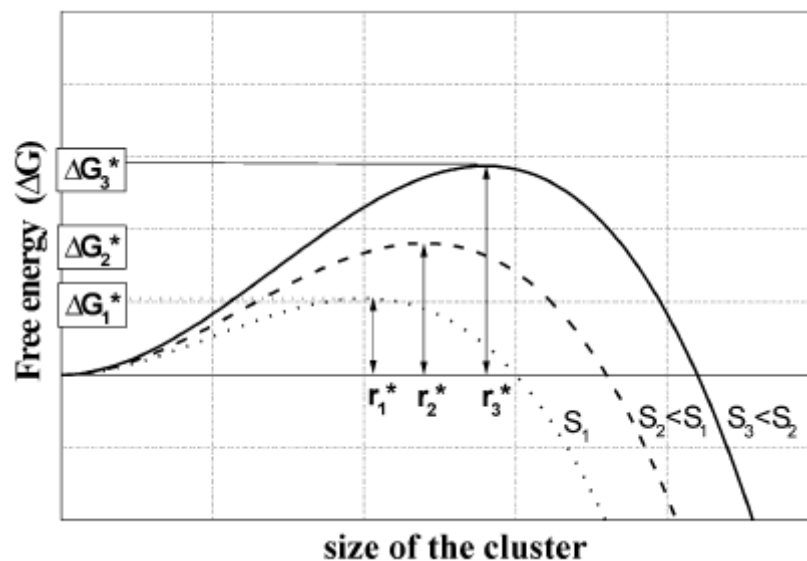
Figure 5 Gibbs free energy curve at different supersaturation[81]
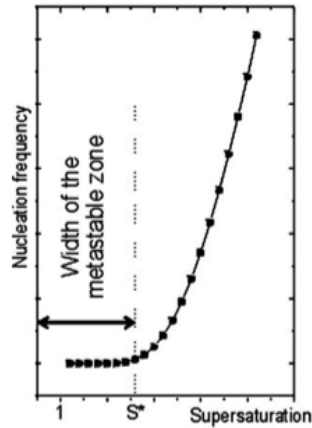
Figure 6 Equation 9 represented visually - the relationship between nucleation frequency and supersaturation.[81]

$$\log t_{nuc} = \kappa_0 + \left[\frac{16\pi\gamma^3 v^2}{3k^3 T^3}\right]/\sigma \quad (11)$$

Limitations to the Classical Nucleation Theory

Classical nucleation theory remains as the main framework to the understanding of the nucleation phenomenon and have been demonstrated to give good predictions for nucleation barrier,[84] theoretical prediction of crystallisation behaviour based on this theory deviates significantly from observed experiments due to the assumptions involved. The theory states that only one nucleation barrier is necessary to be overcome for the formation of new phase, which does not always apply.[85,86]It assumes clusters are spherical, which is the formation that results in the lowest surface energy to volume ratio.[82] This theory also assumes a discontinuous change of density from a diluted bulk phase to a nuclei cluster with density of the new (crystalline) phase, and that the interfacial boundary is distinct. In the case of crystallisation, it also suggests that these clusters formed are ordered.[87] Other factors that this theory has not accounted for include the kinetic and molecular rearrangement, and were covered in a review by Sear (2012),[88] in which various phase change process that does not display classical nucleation behaviour was discussed.

One of the behaviours that shows deviation from prediction would be the nucleation rate. In the case of protein crystallisation, Galkin and Vekilov (2000)[89] investigated the dependence of nucleation rate on supersaturation and precipitant concentration and the experimental results were compared with theoretical predictions: At high supersaturations, the exponential relationship outlined in Figure 6 discontinues. While this exponential dependency was indeed observed at low supersaturations, the nucleation rate determined from experiment is 10 times lower than predictions from classical nucleation theory. A maximum nucleation rate was observed upon decreasing temperature before the rate decreases again.[90]

16

As CNT assumes that the nuclei cluster formed are structurally different from the liquid to a large extent, it is evident at high supersaturation level that distinction between nuclei formed and the densely packed bulk liquid state becomes increasingly difficult. With similar energy states between a highly supersaturated solution and a dense cluster, the energy barrier becomes negligible. This is when CNT alone cannot give us a full representation of the processes occurring in the phenomenon of crystallisation.[91]

2.4.2 Two-Step Nucleation

Another model to explain the processes occurring in crystallisation would be the Two-Step Nucleation Model. This model was inspired by the study by ten Wolde and Frenkel (1997) as part of their studies on the nucleation of a model protein.[92] It accounted for the fact that the crystallisation mechanism depends not only of the strength of solute-solute interaction but also on the range of interaction as well, in these conditions crystals nucleate within the clusters of high solute concentration. The two step nucleation theory was then put forward by Galkin and Vekilov (2000)[89] and was named so as it suggests the two energy barriers needed to be overcome, which are associated with the liquid-liquid and liquid-solid phase separation (Figure 7).[90]

The co-existence of two liquid phases in protein solutions has been observed by light scattering techniques before the two-step nucleation model. Muschol and Rosenberger (1997) demonstrated the impact of this phenomenon on protein crystallisation:[93,94] Initially proposed for the crystallisation of proteins, multiple studies on the crystallisation of lysozyme has reported good correlation to this model, and direct visualisation of the formation of the crystals within clusters of dense liquid was reported. It was reported that the lifetime of clusters are sufficiently long such that these are not just a fluctuation in protein concentration.[95] Observation of the dense liquid phase formation in inorganic or small organic molecules system is more uncommon due to the short lifetime of the metastable phase in these systems[88], but not non-existent,[96] demonstrating the applicability of the two step nucleation theory to a wide range of systems. Liquid-liquid phase separations have been observed in organic pharmaceutical compounds,[97] and amorphous phases are commonly observed for the crystallisation of calcium carbonate.[88,98,99] Kawasaki and Tanaka (2010) also modelled the crystallisation systems and found certain degree of medium range ordering within a supercooled solution that is otherwise 'homogenous'.[100]

In the two-step nucleation model, the crystal nuclei formed were assumed to be of cubic geometry, the free energy of formation of a nuclei of length a consisting of n molecules is therefore given as Equation (12), and the energy barrier and the number of molecules in a critical nucleus is given as Equation (13) and (14) respectively, where Ω is the volume of the nucleus.[101]

$$\Delta G(n) = -n\Delta\mu + 6a^2 n^{2/3}\alpha \quad (12)$$

$$n* = \frac{64v^2\alpha^3}{\Delta\mu^3} \quad (13) \quad \Delta G* = \frac{32\Omega^2\alpha^3}{\Delta\mu^2} \quad (14)$$

The energetics involved in this model is given in Figure 7 for homogeneous crystallisation. The top curve represents the most common condition, where the dense liquid state is less stable than the bulk liquid phase. Both the formation and the growth are thermodynamically unstable. Crystallisation is only possible if the dense droplet formed can persist for sufficient time until the crystal phase is formed. Later work by Vekilov for the study of lysozyme crystallisation reported a long lifetime (10 ms) and large size ($\sim10^6$ as oppose to 10-100 molecules) of the metastable clusters, exceeding the expectations as predicted based on the energy barrier[88,102]
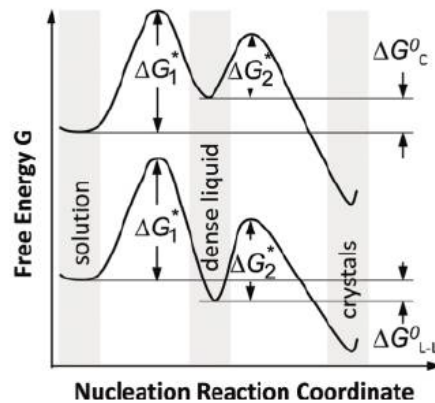
Figure 7 Free energy curve of the crystallisation pathway when the metastable phase is less stable than the vapour phase (above) and vice versa (below) as described in two-step nucleation theory.[90]

2.4.3 Heterogeneous Nucleation

Heterogeneous nucleation occurs on surfaces in which the solute molecules are adsorbed and form clusters. Solid-state impurities have long been observed to increase the chances of obtaining crystals.[22,77,85,103,104] The presence of these foreign bodies reduces the energy barrier for nucleation by paying the interfacial energy cost for the formation of a solid phase nuclei. This causes heterogeneous nucleation to be more favourable than homogeneous nucleation, and is therefore hard to avoid completely due to the presence of vessel walls and

the difficulties in eliminating all impurities in system.[32,53,76,88] However it is believed that homogeneous nucleation still dominates at high supersaturation (Figure 8).[105]

The extent in which the nucleation barrier can be lowered by surfaces of course depends on the interaction properties between the surface and the solute molecules.[106] The relationship between free energy of homogeneous and heterogeneous nucleation (of a flat surface) are related by Equation (15), where ϕ is the electrostatic potential of the surface, which is given by the Volmer Equation (Equation 16), where θ is the contact angle of a solvent droplet on the surface ($0° < \theta < 180°$),[32,53,76] which is the result of the interfacial energies between the solution and the surface as summarised by the Young's Equation (Equation 17),[107] in which γ_{ns}, γ_{bs}, and γ_{nb} are the interfacial energies between nuclei/surfaces, bulk/surface and nuclei/bulk faces respectively present in a heterogeneous system represented by Figure 9.
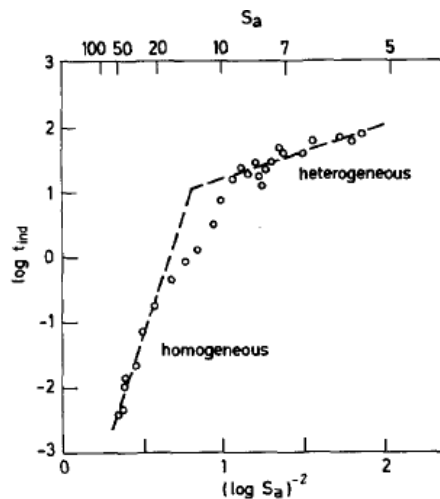
Figure 8 Work by Sohnel et al (1982) to determine the relationship between supersaturation and induction time. Two different gradient was found, which was correlated to homo- and heterogeneous nucleation as labelled.[105]
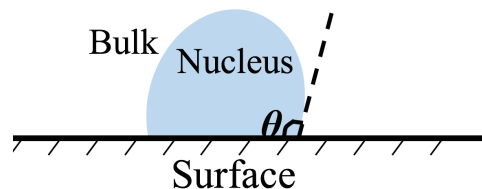
Figure 9 A nucleus (dense phase) formed from a bulk solution wets a surface at the contact angle $\boldsymbol{\theta}$.

When the surface is perfectly wetted ($\theta = 0°$), $\cos\theta = 1$, $\phi = 0$, the nucleation barrier for heterogeneous nucleation would be 0. It is however, unrealistic to say there are no barriers to nucleation in the said condition. When there is no wetting occurring (nucleus is detached from

the surface), $\theta = 180\circ$, $\phi = 1$, the energy barrier for heterogeneous nucleation is equal to that of homogeneous nucleation.[108]

$$\Delta G^{*}_{hetero} = \phi \Delta G^{*}_{homo} \qquad (15)$$

$$\phi = \frac{1}{4}(2 + \cos\theta)(1 - cos\theta)^2 \qquad (16)$$

$$\gamma_{bs} - \gamma_{ns} = \gamma_{nb}\cos(\theta) \qquad (17)$$

In terms of two-step nucleation, surfaces that concentrate solute molecules locally via wetting or capillary condensation promote the formation of dense clusters, even when this phase is not thermodynamically stable in the dense phase. As reported by Galkin and Vekilov,[89] the rate of nucleation was at maximum just before the metastable phase is reached in a homogeneous system. Sear however attributed such observation to heterogeneous nucleation. Such claims have been supported by the theoretical studies[85] in which a surface has promoted the nucleation of crystals just before the formation of intermediate phase in bulk. The intermediate phase has appeared to become stabilised, giving a free energy curve similar to the bottom curve in Figure 7.

Also for the case of two step nucleation, it was reported by van Meel et al (2008)[109] that the metastable clusters formed has to exceed a certain size such that a newly formed crystal is surrounded by a monolayer of solute molecules in the cluster. Further on from that, van Meel and Frenkel (2015)[110] reported that with large wetting, the clusters might adopt a 'pancake shaped' geometry, causing a small thickness (a few layers) for cluster size of <200 molecules in the case of their simulation, which hinders the process of crystal nucleation from the metastable clusters.

2.4.4 Nucleation within Pores

Fluids adopt different behaviour from the bulk phase under the influence of a curved meniscus via the effects of surface tension, contact angle and pressure differences, this is known as the capillary effect.[111] The pressure difference due to the capillary effect can be quantified by the Young-Laplace Equation, which expresses the pressure difference in terms of the curvature of a pore. This effect is responsible for influencing the properties and behaviour of fluids within confined spaces, which is significantly different from their bulk phase.

While the effects of an ideal flat surface were discussed in previous chapters, real surfaces are associated with roughness, pores and pits on surfaces, enhancing the contacts

between the cluster and these heterogeneous surfaces.[106] As will be discussed in this section, geometry of pores contributes largely to the heterogeneous nucleation energy barrier as the surface area of contacts is enhanced due to the curvature. It was found that the larger curvature towards the centre of nuclei (the greater concavity), the lower the nucleation barrier.[88]

While the heterogeneous nucleation barrier approaches the homogeneous nucleation barrier as the contact angle approaches 180◦, the case may differ for porous surfaces. The internal angle of the pore as illustrated in Figure 10 can be given as β, such that heterogeneous nucleation within pores can be expressed as Equation (18).[112,86] The Volmer function ϕ applied to flat surfaces would be replaced with the function $f_w(\theta, \beta)$, suggesting that the energy barrier would depend also on the pore angle in addition to the contact angle of cluster. The function can be expressed as Equation 19, where r is the radius of curvature, A_c and A_f are the surface area of the cluster/bulk and the cluster/substrate interface respectively, which in turn is a function of θ and β.[114] This function was defined such that when $\beta=180$, $f_w(\theta, \alpha) = \phi$.[112]

$$\Delta G_{pore}^* = \Delta G_{homo}^* f(\vartheta, \beta) \qquad (18)$$

$$f(\theta, \beta) = \frac{1}{\pi r^2}(A_c - 2A_f cos\theta) \qquad (19)$$
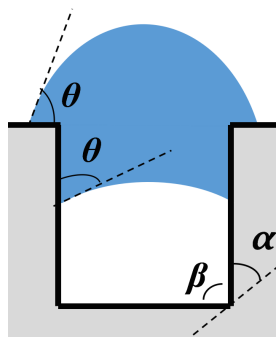
Figure 10 A droplet interacting with a pore of internal angle, β, also represented as 180◦-2α, in which a droplet wets the surfaces at a contact angle θ.

The mechanism occurring within pores are actually more complex than purely geometrical effects and the availability of surface area, which accounts for the reduced energy barrier according to the classical nucleation theory. For nucleation systems following the two-step mechanism, the initial capillary condensation step, where a large uptake of molecules is observed within the pores is of great importance. Nucleation has been reported by van Meel et al (2010) to occur rapidly inside the pores, in which a dense liquid phase is also observed while the bulk solution is still in the vapour phase.[109] The formation of the dense phase via a capillary

effect rapidly decreased the nucleation barrier. Similar trend was observed for flat surface but with a smaller crystalline nucleus.

The idea of an 'optimal pore size' is demonstrated in various publications. Sliwinska-Bartkowiat (2001) demonstrated for the case of the freezing of nitrobenzene that strain free crystals of nitrobenzene can only be obtained when the pore diameter is >20 times the diameter of the molecule,[115] whereas no nucleation was observed in pores <15 time the diameter of molecule. Whereas the study by Page and Sear (2006), an optimal pore size (12-13 lattice sites) is also observed for high nucleation rate in crystallisation.[116] While the pores have to be sufficiently large for a critical nucleus to form, they should also be narrow enough for the capillary condensation to take place. As reported by van Meel and Frenkel (2015),[110] small pore sizes are associated with a tendency of a 'pore break-out', in which homogenous nucleation occurs outside of the pore, limiting the nucleation rate. Whereas above a certain pore size, the liquid within is no longer metastable. At the optimum pore size, the nucleation barrier is the lowest.

Lattice strain is also an important factor to the contribution of pores to the nucleation energy barrier. Page and Sear (2009) investigated the effect of different wedge angle on the nucleation rate of crystals. It was predicted that the plane most densely populated with atoms would have the highest energy.[113] Nucleation barrier would be lowest with these faces in contact with a surface. The nucleation rate was found to be lowest at a wedge angle that corresponds to the coincidence of the two planes, as these walls of wedges reduce the lattice strain upon the formation of crystals, thereby providing a low energy pathway for nucleation. Looking at the crystallisation of aspirin, Diao et al (2011) proposed the mechanism of Directed Angle Nucleation, where nucleation is assisted by the topological features that would give rise to the molecular orientation of the lowest strain.[117] Pores with flat walls are demonstrated to promote an ordered arrangement of crystals, whereas competition of molecule layering on pore bottom vs curved wall exists in cylindrical pores.[109]
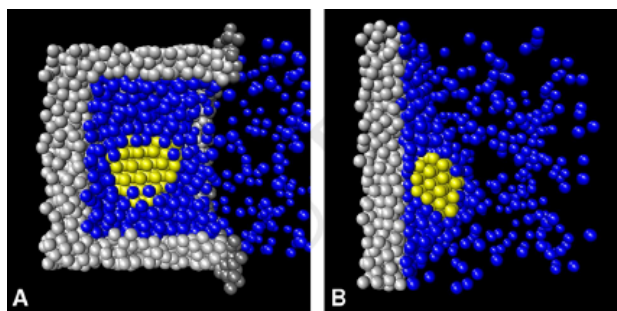
Figure 11 Image highlighting the formation of a crystalline nucleus (yellow) within a pore (A) and a flat surface (B). The influence of the surface causes the liquid phase (blue) within to be denser than outside of the pore.
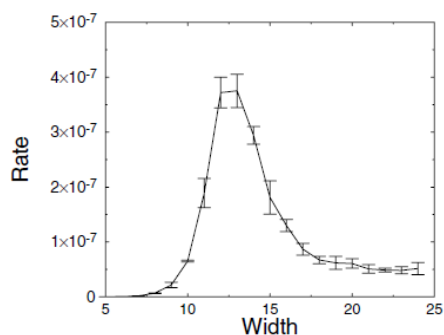
Figure 12 Nucleation rate of molecules in pores of different width. The rate is given in per cycle and the width is measured in terms of lattice sites.

2.4.5 Secondary Nucleation

It has been observed that in the presence of crystals in a supersaturated solution, further nucleation is catalysed by these parent crystals. These processes are termed secondary nucleation, and have been demonstrated to occur via different mechanisms. However, the exact mechanism of secondary nucleation is not certain.[32,53,118]

Secondary nucleation can arise from the nature of the parent crystal (initial breeding or dendritic breeding) or the hydrodynamics (contact nucleation or fluid shear) within the reaction system. Initial breeding is caused by the crystalline fragments of the surface of a seed that may act as nuclei in the crystallisation process and is only of importance in the presence of homogeneous seeds. Dendritic breeding is caused by the fragmentation of the dendritic structure of the crystals that are formed as a result of high supersaturation, which is not a typical condition, applied in batch crystallisation processes.[32,118]

In an agitated system, fluid shear and contact nucleation are both important factors towards secondary nucleation. Contact nucleation occurs as a result collision between a crystal and another solid entity, which could be the reactor wall, impeller or with another crystal, causing fragmentation. This was suggested to be an important factor, while the energy of collision is small between two crystals; the frequency of encounter is high. For the case of

crystal-impeller or crystal-vessel wall collision, the energy of collision is high. Both these two motion results in a significant contribution from contact nucleation. Fluid shear refers to the crystal fragment produced as a result of the shear flow within the fluid and was proposed to be one of the mechanisms of secondary nucleation by Melia and Moffitt (1964).[119] Such fragmentation can result in a new crystal nuclei being formed from the parent crystal. As reported in the same study, the occurrence of secondary nucleation events increases with supersaturation as well as the degree of agitation within the crystalliser as both of these factors increases the degree of attrition.

2.5   Crystal Growth

While the understanding in nucleation is far from complete, the mechanisms for crystal growth are quite well understood. Crystal growth is a mass transport process that occurs after nucleation, in which solute molecules become incorporated into the crystal lattice. In terms of formulation engineering, the control in the crystal growth process is critical in controlling the crystal sizes and crystal size distribution (CSD).[3253]

Mechanisms of crystal growth can be summarised into two main steps: (1) The diffusion of molecules onto the crystal/solute interface, where they become adsorbed, and (2) the incorporation of these adsorbed molecules into lattice. The second step can be considered as more complex. In terms of molecular behaviour, this step would involve the partial release of the solvent shell before the incorporation onto the lattice, and the remaining solvation shell of the molecule would lastly be released as well.[82] It was also reported that the surfaces of crystals are not static even at equilibrium, in which the rate of attachment and detachment are equal.

Crystal growth rate can be expressed as the increase in a dimension along perpendicular to a crystal face over time (linear growth rate), dL/dt, which relates to the mass deposition rate dM/dt, by Equation (20), where ρ is the crystal density and f_v is the volume shape factor. This shows that the growth rate depends largely of the geometry of the crystal.

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \frac{\mathrm{d}(\rho_c f_v L^3)}{\mathrm{d}t} = 3\rho_c f_v L^2 \frac{\mathrm{d}L}{\mathrm{d}t} \ (20)$$

To explain the phenomenon occurring close to the crystal-solution interface, Diffusion-Reaction Model can be applied, and is relevant in industrial crystallisation.[32] The mechanism proposed consists of two steps, the diffusion of solute molecules to regions close to the crystal-solution interface (as explained by Equation (21), and the incorporation of molecules into

crystals (Equation 22), where k_d and k_r are the diffusion and surface reaction rate constants, r is between 1 and 2, and c, c_i and c^* are the concentration of solute in bulk, at interface, and equilibrium saturation concentration respectively. These equations can be combined to give the overall rate, which is given by Equation (23), where k_g is the overall rate constant. The rate constants involved can be related by Equation (24). A concentration gradient established between the bulk and the boundary near the interface as a result of the solute molecules leaving the solution at such region allows for the diffusion from bulk to the interface region to occur.[53]

$$\frac{\mathrm{d}M}{\mathrm{d}t} = k_d A(c - c_i) \quad (21) \qquad \frac{\mathrm{d}M}{\mathrm{d}t} = k_r A(c_i - c^*)^r \quad (22)$$

$$\frac{\mathrm{d}M}{\mathrm{d}t} = k_g A(c - c*)^g \quad (23)$$

$$\frac{1}{k_g} = \frac{1}{k_d} + \frac{1}{k_r} \quad (24)$$

Other models to explain crystal growth are the Continuous Growth, Screw Dislocation and the Birth and Spread models, these models are mainly proposed to explain the possible integration processes of solute into the crystal lattice. Most of these models place a strong emphasis on the layer-by-layer growth on the crystal faces. The components of a surface can be defined as steps, kinks, terraces and islands (Figure 13).[32] As a solute molecule becomes adsorbed onto a crystal surface, it can either diffuse across the surface until it becomes incorporated into the crystal structure by forming bonds with one of these lattice sites or it can dissolve into the solvent. It is hypothesised that the kinks are most probable site at which solute molecules attach or detach, due to the larger number of bonds the solute can form with these sites and the interaction is the strongest. The growth rate is therefore largely dependent on the number of kinks on the surface.[82] However, on a perfect crystal, the formation of islands (2D nucleation) would have to precede the formation of steps and kinks. Like the nucleation of a new phase, the formation of these islands, or 2D nuclei would have to overcome an energy barrier, which is also supersaturation dependent. Another possibility to achieve kink sites would be the presence of dislocation, which can lead to spiral growth. It has been observed with AFM that more islands are observed in crystals grown in higher supersaturation, and spirals at lower supersaturation.[120]
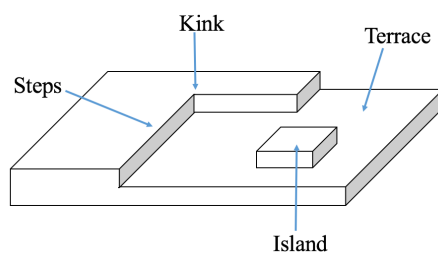
Figure 13 Possible sites for the adsorption of solute molecules on crystal surfaces[82]

The mechanism of protein crystal growth is considered to be similar to inorganic systems, with low kinetic coefficients as its distinctive features. Chernov reported the kinetic coefficient of conavalin, which was two orders of magnitude smaller than conventional crystal growth, and have attributed this to the low probability of incoming molecule to adopt the right orientation for the incorporation of individual molecules. This slow phenomenon was rationalised with Bond Selection Mechanism (BSM),[121] suggesting that only the surfaces of protein molecules are involved in the crystalline bond forming process, and that protein molecules have to be oriented such that lattice contact is possible. This theory was also used to account for the resultant shape of protein crystallisation formed. Nanev and co-workers demonstrated this with the crystallisation of both apo- and holo-ferritin, proteins of the same accessible surface with differing core. Crystallisation behaviour (e.g. crystal shapes) were identical for both of these proteins in the same crystallisation condition,[122] and that single crystals of alternating layers of both proteins were obtained, which has not been observed in any other cases.[123]

Protein crystal growth cessation was reported by Durbin and Feher (1986),[124] even when the protein concentration in the crystallising solution was raised. The possibility proposed were due to shearing of solution that caused the detatchment of proteins from crystals or denaturation of protein. Grant and Saville (1991) demonstrated that the shear forces are orders of magnitude weaker than both the weekest bond interactions that holds the protein structure and the weakest bond that holds a protein to the crystal.[125] The accumulation of crystal defects (Feher and Kam, 1985) was also suggested.[64] This was rationalised by Hirschler (1997)[126] and Vekilov and Rosenberger (1998)[127] as a result of poisoning of growth sites by impurities. Alternatively, Rodriguez (2000)[128] attributed this to the weak lattice force.

2.6 Difficulties Associated with Protein Crystallisation

While the crystallisation mechanisms and processes of small molecules are well understood fundamentally, obtaining protein crystals is still a challenge at all scales. Currently insulin remains the only protein that is crystallised and marketed.[2,13]

In protein crystallisation, huge difficulties arise from the structural complexity and the large size of the proteins. Crystallisation of these macromolecules typically require a supersaturation 2-3 orders higher than inorganic systems, with the nucleation time also being 2-3 orders longer.[129] In the crystallisation of small molecules, each molecule possess a spherical interaction field, and crystallisation occurs in a supersaturated system regardless of the molecular orientation. The heterogeneity of protein molecule causes the molecular orientation for specific protein-protein contact to be important.[130] Whilst applying thermodynamic understandings is crucial to maximising the probability of obtaining crystals, the kinetic requirements also have to be satisfied for crystallisation to happen. For example, the nucleation barrier for lysozyme and ferritin to be of the same order of magnitude as for small molecules, the nucleation rate remained low and was attributed to the kinetic factors that is represented by κ_0, the kinetic pre-exponential factor for the prediction of nucleation rate.[83] While high supersaturation can result in higher nucleation rate, these conditions would not be ideal in controlling the crystal growth to obtain the desired crystal property.

Crystal lattice packing is often cited as the complexity with protein crystallisation. There are a few studies comparing the protein-protein contact within crystal packing with physiological protein-protein interfaces (e.g. processes that drive protein complex formation or binding interaction), and described the area of interaction as surface 'patches'.[131] While the number of pairs of interacting atoms per monomers is similar between the two cases, the patch areas are quite small compared to physiological protein-protein interfaces. The ability of proteins to form such crystal contact depends on a wide variety of different factors and has been investigated in detail in [[132]], such information can be applied to the engineering of protein to obtain a crystallisable protein.[133]

In the perspective of thermodynamics of crystallisation, protein as a long polypeptide chain gives system great flexibility, high entropic component as a free molecule. The overall free energy of protein crystallisation is reported to be orders of magnitude lower than inorganic salts (-10 – 100 kJ mol^{-1}). The enthalpy change gives an indication of the intermolecular bonds formed between protein crystals in a lattice, which was reported to be -70 kJmol^{-1}.[134] In addition

to the unfavourable entropy loss due to formation of an ordered state in crystallisation, there is an additional entropy gain associated with the immobilisation of protein chains of dynamic motion within the crystal lattice, with respect to the rearrangement of protein, the entropy effect is highly unfavourable. However, crystallisation does occur. Protein crystallisation also involves the rearrangement of solvent molecules (water) bound to protein and the release of molecules from hydration shell as a solute, the net gain in entropy can therefore be possible.[51]

In addition to the thermodynamics and kinetics considerations discussed above, difficulties are associated with the practical aspects of protein handling. The physical stability was discussed in Chapter 2.2 and imposes restrictions to the possible conditions that can be used to drive crystallisation, particularly for pH and temperature.[17] At the same time, the influence on protein solubility is multiparameteric, affected by conditions such as temperature, pH, and the concentrations of additives.[49] The next chapter highlights the recent development of various aspects of protein crystallisation.

3. Literature Review

3.1 Protein Crystallisation

3.1.1 Progress in Crystallisation as an Application to Structural Genomics

The possibility to obtain protein crystals is known since 1840, when haemoglobin from earth worm blood was accidentally crystallised by Hunefeld.[13,45,49] Since then, Funke has demonstrated more controlled process of achieving protein crystallisation with the dilution of human haemoglobin prior to the slow evaporation of the mixture.[13,49] Later on, Sumner has investigated the purification of different proteins from jack beans via selective crystallisation.[135]

The main purpose of protein crystallisation to date is to serve as a tool for the study of the 3D structures of protein to atomic resolution via X-ray crystallography, therefore gaining insights into the interaction between protein and metabolites, nucleic acids and other proteins.[136,137] The knowledge of these mechanisms is crucial to the design of drugs and therapeutics. With the vast amount of protein sequence currently in the Uniprot database (over 80 millions),[138] just over 120 000 protein structures are solved and deposited in the Protein Data Bank (PDB).[139] The low success rate is mainly attributed to the complexity of protein, and the multiparameteric nature of the conditions that can affect the phase behaviours of these macromolecules.[49] The protein crystallography processes begin with cloning, followed by the overexpression, purification process before crystallisation. The crystallisation process remains as the bottle-neck in limiting the progress in crystallography.[6,13,19,140] Nevertheless, the last 40 years saw a large increase in the structure solved by crystallography,[141] owing to progresses in five main areas as summarised by McPherson (2004): the understanding in protein crystallisation processes, development of crystallisation approaches, automation and minimisation of experiments with the use of robotics, protein engineering to improve crystallisability and development of screening conditions based on database available.

The first diffraction pattern for protein crystals was obtained by Bernal and Crowfoot (1934) for crystalline pepsin,[142] one of the earlier molecules that have been reported to crystallise reproducibly. The early protein structures that were solved involved the molecules with robust crystallisation protocol, available in large quantities[13]. Once these structures are solved, new proteins of interest were crystallised with limited success.[13] Subsequently, various publications shared crystallisation strategies. Different studies have since attempted in understanding the roles of different common additives to protein crystallisation. Some insights were gained relating to the roles of different additives, methods were also explored to determine

the solubility of different proteins. With the limited amount of crystallised proteins available in the early days, limited information can be extracted from these studies. (ref for e.g. effects of PEGSs, salts) For this reason, protein crystallisation remains an empirical process.

Crystallisation therefore requires an exhaustive screen involving different conditions such as varying pH, temperature and precipitants and additives used. Carter and Carter (1979)[143] devised an incomplete factorial approach, aiming to scatter all screening conditions evenly across a multidimensional phase diagram, where Jancarik and Kim's (1991) sparse matrix screening protocol aimed to narrow the search based on the previous successful crystallisation attempts shared in publications.[144] Other screens were also reported based on random approaches, which Segelke (2001) compared and suggested the effectiveness of completely random screening approaches.[145] McPherson and Cudney (2006) observed the stabilising effects of small molecules on proteins, and also the incorporation of these molecules within protein crystal lattices which indicated the ability of these molecules to promote contacts between protein molecules within a lattice.[146] The use of these additives has been demonstrated to enhance the crystallisability of different proteins that were not crystallised before. However, the conditions necessary to drive crystallisation are protein-specific, and cannot be predicted based on the successful attempts in similar structures.[41–43]

High throughput techniques in protein crystallisation contribute hugely to the exponential increase in the number of crystal structures solved despite the existing difficulties with protein crystallisation.[2,6,13,49,140,147–150] The challenges highlighted in the issues above include the long processes required to set up experiments as well as the scarce amount of proteins obtained. Robotics that enables the automated dispensing of nanolitre volume of crystallisation solution (reviewed in details by Stevens (2000))[149], coupled with the rapid analysis of droplets (discussed in detail in Pusey (2005)) allows for this empirical search for crystallisation condition to be enhanced significantly.[150] While the progress in the development of such technology has greatly increased the efficiency, the amount of samples that yielded a structure remained at ~20 % for over a decade.[151]

It was reported that ~80% of the crystallisable proteins can crystallise in a wide range of different conditions, and this property is attributed to the inherent properties of the proteins.[41] Protein engineering, which involves the modification of proteins to reduce the conformational entropy, making crystallisation more favourable, also enhancing the contact points between

protein crystals.[19,42,45,48,49,137,148,150] This aspect of protein crystallography is also enhanced by the existing information in the PDB.

Significant amount of studies has been published around 1980s and 1990s on the effects of different crystallisation variables, which was discussed in Chapter 2, these include the systematic studies on effects of different precipitants used, methods in determining the solubility. Methods adopted to achieve diffractive crystals was discussed in Chapter 3.2.1. Another method in exercising certain degree of control would be the use of seeds. While the use of protein seeds has demonstrated some success in optimisation, obtaining seeds for a novel protein is a challenge. In the perspective of batch crystallisation of proteins, protein crystalline materials exhibit poor mechanical properties and are difficult to handle as seeds. An alternative would be the use of heterogeneous seeds. Chayen et al (2001) first demonstrated the use of disordered mesoporous silica as a nucleant for protein crystallisation, inspiring the development of other mesoporous material as well as theoretical studies. However, limited experimental data has been published to demonstrate the relationship between the pore size and the protein of interest, until Shah et al (2011) using vapour diffusion experiment.

3.2   Methodologies in Protein Crystallisation

3.2.1   Driving conditions to supersaturation

While the aim of protein crystallisation is to grow large protein crystals, most of the methods described below are developed to generate supersaturation slowly, and is achieved across different route through the phase diagram (Figure 14).
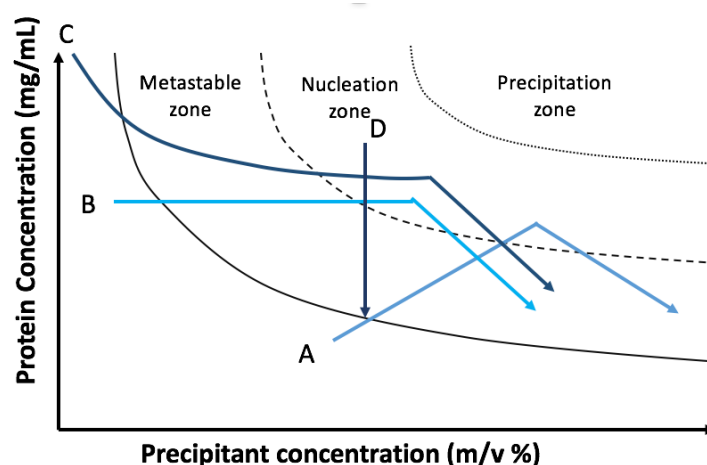
Figure 14 Crystallisation pathways of different crystallisation method: A) Vapour diffusion, B) Dialysis C) Counter Diffusion, D) Microbatch

Vapour Diffusion:

This technique was first used in 1968 by Hampel and co-workers for the crystallisation of tRNA,[152] and currently remains the most widely used method for protein crystallisation. As vapour diffusion requires small volume of protein per condition, it is extensively used in screening processes, a trial and error process to find crystallisation condition of proteins from a range of different combinations of conditions. Crystals produced with this method typically ranges from 10-100 μm in diameter.[147]

This method involves an undersaturated droplet consisting protein (<5 μL) and a 'reservoir solution' in a closed system. The reservoir would be a buffered precipitant solution, and the droplet would consist of a 50:50 mixture of protein stock solution to precipitant solution. The resultant concentration difference in precipitant would therefore drive the equilibration of process, concentrating both the protein and precipitant within the droplet, driving the condition to supersaturation. Vapour diffusion can be conducted in two main methods, hanging drop vapour diffusion and sitting drop vapour diffusion, while sitting drop is used in automated screening systems. These techniques are illustrated in Figure 15 a and b respectively.[18]

The gradual increase in concentration means that multiple conditions can be explored from one experiment. This method is however not without its limitations. The equilibration rate is very difficult to be controlled. Different methods have been reported, such as controlling droplet size[153] or the variation in the distance between droplet and reservoir.[154] The use of oil as a barrier placed on the reservoir solution was introduced by Chayen (1997) to limit the diffusion rate of water.[155] However, even with these variations possible, it is very difficult to pin point the concentration of both the protein and the precipitants at which the crystals are formed.[153]
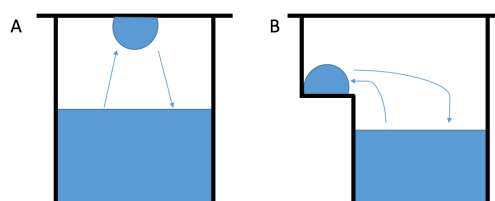
Figure 15 Scheme of (A) hanging drop and (B) sitting drop vapour diffusion method, adopted from [62]

Counter diffusion

Like vapour diffusion, counter diffusion is reported as a technique that relies on the diffusion of water molecules. However, the protein and precipitant solutions are kept separated by a medium such as air, oil or gel within the same capillary tube. For the interest of crystallographers, where the crystal size is important.[156] This technique is attractive as it does not only give a temporal change in solution concentration, a concentration gradient of both the protein and precipitant varies along the thickness of the protein droplet (Figure 16),[157] and the diffusion behaviour can be simulated.[158]
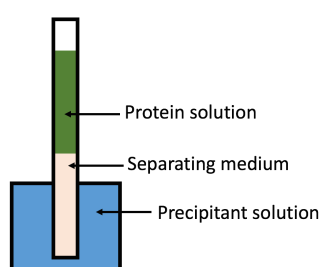
Figure 16 Scheme of counter diffusion set up, adapted from [159]

Dialysis

In this technique, protein solution of ~1 mL volume is kept within the capillary tube to equilibrate with a precipitant solution separated by a dialysis membrane. This method was in fact in early days used quite commonly to grow large crystals; however, due to the large volume needed, it was not widely explored by researchers.[18]

Other methods:

Microbatch

While the above methods mentioned are good approaches in obtaining large crystals, batch crystallisation is useful in determining the concentration in which crystallisation occurs, and there are even claims that better crystal quality than from vapour diffusion can be grown.[153] While in a sealed system, no evaporation is expected,[137] Chayen has investigated extensively the use of oil above the crystallisation droplet as it was discovered that certain paraffin and silicone oil gives rise to the slow evaporation of water.[160]

3.2.2   Phase Properties Determination:

Knowledge on the solubility of a compound is critical for the design of crystallisation process for small molecules such as active pharmaceutical ingredients (API), and the methods

of determining solubility of these species of molecules are quite well established and summarised within [161]. There have been several reports on the methods of determining the solubility of the protein. Ataka (1993) summarised these methods in three main approaches: crystallisation until equilibrium is reached, the dissolution of crystals, and the monitoring of crystal size via microscopy, the early methods for solubility determination was also summarised within.[162] In the case of proteins, the determination of solubility is not as straight forward. Most protein solubility determination literature described below was produced during late 1980s to 1990s. Most of these methods demonstrated involve proteins that have been reproducibly crystallised before. Crystallising a newly developed protein therefore remains a challenge. The equilibration time for proteins is also very slow (~1-5 months),[163,164] there is still a need for more time and experiment efficient techniques.[165]

Boistelle and Astier (1992) measured the solubility of porcine pancreatic alpha amylase by monitoring the concentration of both a supersaturated solution that undergone crystallisation and an undersaturated crystalline suspension that undergone dissolution process for equilibrium. After one month of equilibration there was still a 5-10% difference in the concentration between the residual concentration of these two suspension, which was attributed to growth cessation at low supersaturation of protein although impurity was not apparent. The crystal growth process shown in the study was particularly slow compared to the dissolution process,[164] which was not unsual as discussed in Chapter 2.5, the crystal growth rate of protein was orders of magnitude lower than crystals of small molecules.

Rosenberger (1993) has developed a set-up to determine the solubility of lysozyme and two different types of albumin at different temperatures.[166] This technique is akin to the solubility determination of small molecules, where crystallisation occurred at low temperature (<10 ∘C), and were redissolved gradually upon small temperature increments. Maosoongnern et al (2012) extended this further to investigate the nucleation point as well with parallel experimentation systems. Devices are also available for similar studies and have been used by Christopher et al (1997) for the solubility determination of 30 model proteins at various temperatures, visual examinations was used rather than light scattering signals.[167]

Column based methods have been investigated by Pusey and Gernert (1988) extensively also for the determination of lysozyme solubility.[168] This method involves a supersaturated and an undersaturated solution running through two separate columns backed with small lysozyme crystals, speeding up the equilibration process to 12 h by enhancing the

interfacial area between the solution and crystals. Extensive protein crystals are required as a result. Further studies were also published on reducing the amount of materials required. More studies were published by the same group to reduce the proteins used and reducing the equilibration time.[169]

Since then, several automated techniques were developed for the purpose of generating a 'working phase diagram', in which the supersolubility curve was determined and not the solubility curve and was reviewed extensively in Chayen (2005),[165] these techniques relies heavily on imaging. Baumgartner et al (2015) demonstrated the use of imaging based software to assess the solids formed from different experiment, of whether gels, crystals or precipitates are formed, defining most phases within the phase diagram.[170]

In recent years Talreja et al (2010) mathematically related the evaporate rates of water in ambient conditions with the cross sectional area of a protein crystallisation droplet, and have since developed instrumentation to control the size of this property, coupling this with imaging technique, they have determined the supersolubility curve.[171] The instrument also allowed for the introduction and replacement of reservoir solutions of lower precipitant concentration, allowing the rehydration of droplet condition and subsequently the dissolution of the crystal that was grown firstly from the supersolubility curve estimation, similar to the principles used by Saridakis and Chayen (2003) to improve crystal quality.[172] It was claimed within the literature that this method can also apply to membrane proteins as well.

While dissolution techniques serves as a robust method to determine the solubility of crystals due to its success with small molecule, this technique relies on protein crystals to be achieved reproducibly at certain conditions and the abundance of these protein crystals. Minaturisation of experiments reduces the amount of crystals needed but relies heavily on imaging techniques to monitor crystal formation and dissolution, which compromises in sensitivity as solids formed are typically only detected visually above 5 μm, and specific instrumentation is needed.[171] Given the availability of instrumentation, this technique is useful as a rapid tool to determine the crystallising conditions. Micro column based method results in the fastest equilibration time and was acknowledged as a high accuracy technique, however the quantity of crystals are needed. Most of the techniques discussed here were developed based on the solubility of lysozyme and other proteins reproducibly crystallised.[165]

3.2.3 Batch Crystallisation of Proteins

The first case of successful macromolecular crystallisation was reported in 1957 for the crystallisation of insulin. However, insulin is still the only protein available in crystalline form in the market to date. Yang et al (2003) crystallised three approved antibodies with a yield of >90% and demonstrated the retention of bioactivity and structural integrity. A full length IgG1 antibody has been crystallised in batch scale, reporting a yield of >90% with a crystallisation onset of 3 minutes when the conditions are optimised at 10˚C in a concentration of 25 g/L. However, the details of the crystallisation process were not discussed.[173]

While process controls are frequently applied to the crystallisation of API to allow for the possibility of the controlled drug release in bodies,[174] this is less frequently employed in the protein crystallisation of proteins. Several publications report on the modelling and control of protein crystal size and shape, which are again, mostly based on studies with lysozyme, which are discussed later in this chapter. Cooling crystallisation, a common technique in the batch crystallisation of API, has been demonstrated for lysozyme at <1 mL,[175] while batch crystallisation of lysozyme at 1mL scale with a spatial temperature gradient was demonstrated for the same protein.[176,177] As protein phase behaviour is commonly described in varying precipitant concentrations, the application of alternative process designs (summarised below) can be beneficial.

Ryu et al (2010) applied a solvent freeze out to the batch crystallisation of lysozyme, in which a 'cold finger' is incorporated into the batch crystallisation system. Supersaturation was generated by the freezing of water onto the surface of the cold finger, therefore concentrating both the proteins and the precipitants.[178] This technique was later developed further for the control of temperature gradient across a system to control the nucleation process[179] and also found its application in the purification of lysozyme from a lysozyme/ovalbumin mixture.[180,181]

Gross and Kind (2016) developed an evaporative technique for protein crystallisation as large as 3L by the use of vacuum in elevated temperature.[182] The activity was demonstrated to retain even at reduced pressure. Like the solvent freeze out method, the motivation was to reduce the salt usage at large scales.

Huettman and co-workers reported the batch crystallisation of interferon-gamma with the goal of obtaining high yield. Ammonium sulphate was used as the precipitant due to the steep solubility curve, which results in low solubility and therefore high yield in the

crystallisation of proteins. They have also reported on the controlled crystallisation of a single chain antibody in a two phase system consisting of sodium sulfate and PEG 2000 rich solutions (2015).[183] The sodium sulfate solution was described to play a role in the nucleation of the antibody chain described while the latter was suggested to control the crystal growth process. In stirred systems from 10 mL to 220 mL, crystals reported to have nucleated at the interface between the two phases before growing in the PEG 2000 rich droplets.[21]

With the difficulties present in transferring from the small scale crystallisation experiments to large scale batch crystallisation, research has been conducted to investigate the scaling up kinetics. In most scale-up studies, screening for optimal conditions at small scale is usually conducted in the first stage. Hekmat and co-workers reported in 2007 the effects of different degrees of agitation on the qualities of crystals produced and the effects on phase diagram. In the study lysozyme was first crystallised in microbatch scale to determine the phase diagram. The supersolubility curve was shifted to the left upon either shaking or stirring, this was also found to be frequency dependent. Stirring was found to narrow the crystal size distribution as crystallisation progressed, which the authors attributed this to mechanical shearing.[184] Further investigation has also been done on the scaling up parameter on crystallisation. Using lysozyme as a model protein, Smejkal and co-workers (2013) investigated the scalability of different parameters in a stirred system. The maximum local energy dissipation (ε_{max}) was found to be more suitable than the conventional parameters such as mean power input and impeller tip speed.[185]

The first case of protein purification by bulk crystallisation is reported by Judge et al (1995) by seeding at metastable conditions, in which crystals of high ovalbumin purity (99.4%) was achieved as a result of the lattice match. Ovalbumin was crystallised from a solution containing also other smaller (lysozyme) and larger (conalbumin) proteins. The growth rate was also studied using the combination of both UV-vis spectroscopy and crystal size monitoring technique. The presence of protein impurities in solution shows no effect on the crystal growth rate of ovalbumin.[186] Carbone and co-workers used this method and proposed a growth model.[187] Since then there are a few other reports of using crystallisation as a purification process in isolating each protein found in egg white.[188]

Jacobson et al (1998) purified lipase from a concentrated fermentation broth. SDS-PAGE results of the crystalline product shows the retention of the target protein (35 kDA) where the contaminant proteins of 25 and 40 kDa were not apparent in the results. Using a

combination of particle size determination technique, the crystal nucleation and growth rate were studied. A large crystal size distribution (5-65 μm) and small mean crystal size (16.2 μm) was observed, the values compared favourably with the CSD reported by Judge et al (1995). The authors have attributed the difficulties due to the strong dependency on supersaturation compared to inorganic or small molecules.[189] Giffard et al (2011) studied the effectiveness of using crystallisation as part of the downstream processing of urate oxidase from fermentation broth. In here, crystallisation was conducted at different stages of purification (concentration, ultrafiltration and chromatography). Implementing crystallisation at earlier stage was found to result in the least amount of impurity. In term of quality of crystals produced, aggregates are difficult to be removed when crystallisation is conducted at the final stage.[190]

3.3   Seeding

Homogeneous seeding is commonly employed in the industrial crystallisation of small molecules as it plays a role in the control of crystal forms and sizes.[191] Due to the difficulties in controlling the nucleation process in the nucleation zone, the addition of seeds obtained external to the system can bypass the nucleation process by allowing the growth of crystals from these seeds in the metastable zone.[32] With protein crystallisation, this is not as straight forward. In batch crystallisation, obtaining crystalline protein solids of sufficient quality to be used as seed is a challenge to begin with.[176] And as discussed previously, the crystals are also known to be more fragile in comparison, and breakages are likely upon stirring/ handling. In the field of batch crystallisation of proteins, homogeneous crystallisation is not commonly explored. Homogeneous seeding however, has found some use in the optimisation of crystallisation condition.

Heterogeneous seeding, however, have been reported for the crystallisation of protein to achieve diffraction quality crystals. This involves the use of solids that are not the crystal of the solute molecules.[192] This would be discussed further in Chapter 3.3.1.

3.3.1   Homogeneous seeding in protein crystallisation

In protein crystallisation, homogeneous seeding is used to estimate the metastable conditions of the protein, and is used in the optimisation process after screening.[140] Also, crystals that are too small or contains defect can be placed in the optimisation condition.[193] Once this information is determined, seeding can be used to as a strategy to grow diffractive quality crystals. This is obtained by placing 'seed stocks' into the screening droplets. Homogeneous seeding can be 'macroseeded', 'microseeded' or 'streak-seeded'. Macroseeding

involves the transfer of a whole crystal (5-50 µm) into the reservoir solution of the seeding well, whereas microseeding involves the use of crystals of submicroscale, crushed or pulverised from larger crystals.[23] Streak seeding is used when the protein based material of certain degree of crystalline order is formed as a result of screening, these material can serve as a seed for further screening and optimisation purposes.[22,23,192] The use of these methods has proven great success in the optimisation process in obtaining diffraction quality crystals. However, due to the poor mechanical properties of protein crystals, there are always difficulties associated with handling.[190]

3.3.2   Heterogeneous seeding in protein crystallisation

As heterogeneous nucleation is recognised as a suitable seeding technique, different materials have been proposed as potential nucleants for protein crystallisation.[22,193] The effects of heterogeneous nucleation on surfaces and pores have been discussed in Chapter 2. As our project aims to promote crystallisation by manipulating the nucleant-protein interactions, this part of the review would focus on the use of different surface properties (roughness, porosity and surface chemistry).

Epitaxy

This method involves the deposition of solute molecules onto a surface of matching lattice. The use of epitaxy to induce crystallisation of proteins was first demonstrated by McPherson and Shlichta (1988). In this experiment 50 different minerals were used as surfaces to promote crystallisation of four different model proteins. However, only limited matches between the epitaxial surfaces and proteins were reported, demonstrating the specificity of the lattice match required.[194] Edwards et al (1994) reported the use of lipid layers as an epitaxial surface for the crystallisation of streptavidin. In which both the specific and electrostatic interactions between the protein molecules and polar headgroups have contributed to the increase in local concentration on the surface to 500-1000 mg/mL in a solution of 100 mL. Crystallisation of the protein has shown to occur at a lower precipitant concentration in the presence of the epitaxial layer.[195] In 2008, a zeolite was used to promote epitaxial crystal growth by Sugahara et al. Most of the proteins used in the study were obtained in a shorter time frame in the presence of a zeolitic species. The main advantage of using these microporous species was the enhancement of crystal diffraction quality for crystallographic studies. However, it has been shown that the slightest mismatch between the two different lattices imposes lattice strains onto the crystal,[196] increasing the crystallisation barrier. This technique

was not pursued further due to the inapplicability to the crystallisation of novel proteins in which lattice information are absent.

Surface Curvature

The effect of surface curvature was demonstrated for the batch crystallisation of lysozyme by Weichsel and co-workers. The mass of seed used was adjusted such that the surface area present can be controlled for these differently sized particles (10-200 nm). In the experiments the phase diagrams in the presence of different seeds were generated and the effect on supersolubility curve was not very clear in the time scale of 22 h. However, a significantly larger number of conditions has yielded crystals within 19 minutes for particles between 10 nm and 100 nm, and between 100 nm and 200 nm. The authors have attributed the larger uptake of the proteins at a given surface area due to the lower curvature at large particle sizes.[197]

Surface Roughness

The surface roughness can be quantified as the ratio of the actual area to projected area, known as the roughness factor, r,[198] and can be expressed as Equation (25) with an assumed geometry represented in Figure 17, where h, l and s are geometrical parameters represented in the diagram, and m is the number of dents present on the surface.[199]
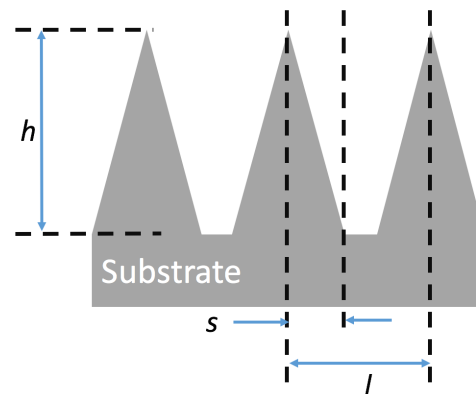
Figure 17 Diagram of rough surfaces of geometries defined by parameters *s, l,* and *h*, adopted from [199]

$$\omega = \frac{(\pi l^2 - m\pi s^2) + m\pi s\sqrt{(h^2 + s^2)}}{\pi l^2} \quad (25)$$

Curcio and co-workers modelled the effects of surface roughness of polymeric membranes (roughness factor: $1<r<1.6$) on the energetics of protein crystallisation. While a perfectly smooth surface ($r=1$) results in a reduction in energy barrier by 50% compared to homogenous nucleation, increase of surface roughness to just $r=1.6$ results in an energy barrier less that 30% of the homogeneous nucleation barrier, depending on the model used, which caused the critical nuclei radius to be reduced by a similar factor.[199] Ghatak and Ghatak (2013)

produced substrates by stretching polydimethylsiloxane to length scale of ~125 nm and plasma oxidised the material to generate a silica surface, giving a resultant depth of ~20 nm. The topology in this case was represented by curvatures ranging from -10 (concave) to 10 (convex). The disordered geometry allowed for proteins to self select at a suitable position, and that proteins were observed to self-select a suitable site corresponding to these geometries for nucleation. The roughness of these materials were sufficiently large that thaumatin was reported to crystallise in the absence of precipitants.[200]

Surface Functionalised Materials

Fermani et al (2001) reported the effects of surface chemistry on the crystallisation of Concavalin A (ConA) using various polypeptide-coated polymeric matrices as well as sulfonated polystyrene. In these experiments polystyrene films were sulphonated to different extent, and the nucleation time of the model protein was found to vary inversely with the length of time the films were in the sulfonation solution. Using the hanging drop vapour diffusion method at 20˚C, the nucleation time reported for unfunctionalised polystyrene was 10 days, and was reduced to 2 days when the nucleant was fully sulfonated. Gelatin and silk fibroin coated or entrapped with poly-L-Lysine or poly-L-aspartate have also been used in the study for the crystallisation of ConA and lysozyme, and the ability of these materials to promote crystallisation is related to the relative charge of the residue and on the protein surface.[201]

The use of silica also allows for the attachment of chemical functionalities onto the surface with the use of silanes. In early researches, chemically modified mica sheets were used. As these materials are exposed at the (001) surface, the flatness of these materials enables the study of surface chemistry by minimising the morphological effects.[202] Crystallisation of various model proteins (lysozyme, Concavalin A, thaumatin) on partially amine-functionalised mica surfaces have been conducted (Falini et al, 2002). [203] All of the studied proteins displayed a decrease in nucleation time with the increase in extent of functionalisation. This trend becomes more prominent at lower concentration level of proteins. However, in terms of the nucleation density and the size of crystals, the trend varies depending on the characteristics of proteins. Crystallisation of large biomolecules on chemically modified mica surfaces is then further explored by Tosi et al (2006) in which insulin and ribonuclease A were the molecules of interest. Mica surfaces functionalised with ionisable groups (sulfonyl and amine) were demonstrated to be effective at concentration below 10mg/ mL.[202] As with results reported by Falini et al, the lower the concentration, the more dependent the proteins are on surface chemistry. The effects of surface chemistry have also been demonstrated by Tsekova et al

(2012) to affect the crystal habit and sizes.[204] In this study, glass slides were silanised with chemical functionalities of different polarity to achieve crystallisation for various model proteins. These groups have shown to influence the resulting crystal habit due to the specific interactions with the protein molecules to promote crystal growth on certain planes. In the case of both lysozyme and catalase, rhombic crystals are obtained on surfaces with alkyl groups. This habit is also obtained in high supersaturation in homogeneous nucleation, suggesting the ability of these groups to promote high supersaturation. In the case of Thaumatin where low protein concentration was used, the dependency on the surface chemistry becomes more apparent.

Natural Polymers

With an observation that crystals appear spontaneously on fibril material, D'arcy et al (2003) reported the use of horse hair to achieve crystallisation of lysozyme.[192] The cuticles on the hair were suggested as a desirable microstructural property to trap fragments of crystals. Other naturally occurring materials such as whiskers have also been attempted. Horse hair, however have the advantages of abundance of cuticles and uniform diameter across the whole length. To produce a homogenous sample of horse hair, these materials were crushed into crude powder before they are used as seeds. It was reported that the concentration of seed has a profound effect on the crystal density. Whilst for a given condition, only liquid-liquid phase separation is observed in the absence of seed materials, the addition of horse hair can drive the condition to obtain crystals. The disadvantage of using this method is the inclusion of material within the seed.

Porous Material

Chayen et al (2001) pioneered the use of porous materials to achieve protein crystallisation, and early works have shown that a large distribution of pore size resulted in the most successful protein crystallisation for multiple proteins. The pore size present in these materials ranged from 2-10 nm, comparable with the scale of the diameter of protein investigated.[26] Striving for a 'universal nucleant', materials of similar pore properties were developed; examples include Bioglass, carbon nanotubes or porous gold nanoparticles, each with their own advantages such as stability or ease of design and synthesis.[22]

The work by Chayen and co-workers triggered the interest in the theoretical understanding in the roles of the pores on crystallisation, which were summarised previously in Chapter 2. The effects of capillary condensation were extensively discussed in these studies. Additional considerations need to be taken with protein crystallisation. Stolyarova et al (2006)

reported local supersaturation within porous silica can be achieved for the crystallisation of proteins,[205] which Nanev (2017) attributed this to the ease of re-adsorption or trapping of protein molecules within pores despite the low desorption energy. Protein-protein interaction would have to be sufficient and specific in order for crystallisation to occur, otherwise a deposition of molecule on surface is obtained.[121]

The development of molecularly imprinted polymers (MIP) followed as the next candidate of materials by Chayen and co-workers. The surface topology of these materials are formed by the imprint of protein of interest onto the polymer, causing a close match in size and shape to the protein. It was however, found to be difficult to remove the protein, and was suggested to not be very practical for crystallisation.[206]

While studies regarding crystallisation in confined pores relates the optimal pore sizes with the critical nuclei radius (discussed in Chapter 2), the examples of porous seeds above suggest that pores that fit one protein molecule is sufficient in promoting protein crystallisation. Shah et al (2011) rationalised this by the pore confinement effect, in which the pores can stabilise the folded structure of proteins, and the local concentration of folded protein is increased.[28]

It is known that the protein folding that occurs spontaneously in cells is influenced by the crowded environment within, thereby forming the native state that contributes towards their activity. This crowding effect was first investigated by Minton (1992). It was suggested that the folding can occur via the volume exclusion effect from either surrounding macromolecules (crowding) or confined boundaries (confinement).[207] Protein confinement is associated with a high entropy cost when the allowed protein conformation is reduced from all possible conformation that can occur in bulk by the restricted volume,[208] also the equilibrium between the folded and unfolded state of protein is shifted,[207] therefore influencing the stability of the folded protein. While the energy cost associated with crowding relates to the volume of confined space, the energy associated with confinement also depends on the shape of the confined area. Mittal and Best (2008) however weighted the importance of the size of confinement over the shapes studied.[209]

Using the protein confinement effect to explain the effect of pore sizes on protein, Shah and co-workers (2011) synthesised silica particles of narrow pore size distribution using a soft templating approach. In fact, a close match between the diameter of protein and the diameter of the pore has been demonstrated to result in the shortest induction time compared to other

surfaces for a range of model proteins tested, with hydrodynamic diameter tested ranging from 4-20 nm. When there is such a size match, the research group termed these mesoporous silica material 'nanotemplates'.[28] This finding was significant as it demonstrates a potential for bioseparation, as these materials have been demonstrate to selectively crystallise one protein from a crude mixture based on this relationship.[210]

Materials with combined effects of chemistry and porosity

Delmas et al (2012) reported the combined use of surface porosity and surface chemistry with the use of functionalised silica nanoparticles deposited on glass cover slips (Figure 18). In addition to the chemical interactions between surface functional group A and protein side-chains B as well as the microporosity within template, the intrapartiular mesopores help promoting nucleation. In this study, both surface chemistry and porosity were investigated separately prior to the use of silanised particles for the combined effects on the crystallisation of lysozyme. The induction time for the process increases with increasing hydrophobicity but an opposite trend was observed for the crystal density, demonstrating the lack of control possible with the use of surface chemistry alone. No apparent trend was observed for induction time nor crystal density with increasing particle size, suggesting a more dominant factor than surface area, which has been attributed to the inter- and intraparticle porosity.[211]
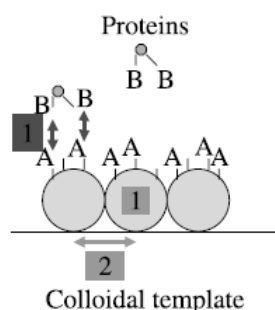
Figure 18 Schematic diagram illustrating the possible interaction mechanism between the colloidal template and proteins. Possible interactions include chemcial interactions (A-B), microporosity within template and interparticle distances.[211]

Based on the effects observed in the study described above, Shah and co-workers (2012) has applied this and silanised the nanotemplates with functional groups of different hydrophilicity. As a result, crystallisation was possible at concentrations even lower than reported in literature prior to publication date.[30]

4. Experimental Design

4.1   General experimental design and scaling up routes

Crystallisation of most proteins begin with vapour diffusion experiment, which cannot be easily transferred to technical scale.[184] Different process designs adapted for protein crystallisation have been reported in literature and was summarised in Chapter 3.2.3, with the aim of gradually increase both the protein and precipitant concentrations, and therefore mimicking the thermodynamic pathways in vapour diffusion. While scaling up parameters are important in crystallisation kinetics in order to maintain batch-to-batch reproducibility in industrial crystallisation,[212] limited information was reported on the detailed investigation in this area for the crystallisation of proteins, and will not be the main focus in this thesis. The objective is to demonstrate the effects of nanotemplates at conditions that are used in crystallisation vessels, stirred experiments are performed.

Both HEWL and BLC were chosen as model proteins for this study based on the feasibility of obtaining quantity sufficient for batch scale (20 mL) crystallisation of proteins. While the crystallisation conditions of model proteins used within this thesis was deduced with the use of vapour diffusion and are given in literature, the ability of the protein to crystallise in bulk liquid in the absence of evaporation was studied for BLC prior to studies on nanotemplates. Therefore initial experiments involve a working volume of 0.5 mL was used to demonstrate the crystallisability of catalase in batch scale prior to the experiments at 20 mL. For HEWL, which is known to crystallise at a large range of conditions, crystallisation in 25 mL stirred system was studied directly.

In the volume of 0.5 mL, the stirring rate used within was kept at < 400 rpm. At this scale, a screening was conducted to identify conditions in which nucleation is relatively slow, which would allow for the comparison using nanotemplates and other heterogeneous seeds. Upon scaling up, the vessel design was changed from a magnetically stirred system of the volume described above to a 20 mL system with overhead 2-blade pitch stirrers, and the kinetics was not easily compared.

In our studies, the requirement of our operating conditions is to suspend the heterogeneous seeds in the crystallisation system but without causing protein denaturation. Few cases of batch crystallisation of proteins were documented in literature. Crystallisation of various proteins at large scale (100-500 mL) reported the use of stirring rates between 200 to 700 rpm[189,213], and small scale (6 mL) at up to 300 rpm for HEWL.[185] In crystallisation, the

increased stirring rate was demonstrated to decrease the MSZW[214] or reduce induction time[185], suggesting that nucleation happens more readily at higher degree of agitation. The effects of crystallisation on scaling up in the lab scale range (30 – 450 mL) was reported by Yi and Myerson (2006). It was demonstrated that the induction time increases with working volume, the effects of the longer circulation time, and extent of mixing and the reduced energy dissipation rates on induction times as a result of larger volumes used were also discussed [212].

4.2. Use of Heterogeneous Seeds and Seeding Parameters

Combining with cooling crystallisation, seeding is often cited as a common technique to control the crystal size distribution (CSD) in batch crystallisation of small molecules, and also to control the purity of the product from downstream processing. Due to the difficulties in controlling the nucleation process, the addition of seeds obtained external to the system can bypass the nucleation process by allowing the growth of crystals from these seeds in the metastable zone.[32] While homogeneous seeding is commonly used in most crystallisation process, difficulties are associated with the handling of protein crystal seeds.[176] Heterogeneous seeding can serve as an alternative in protein crystallisation.

Chadwick and co-workers reported the direct crystallisation of acetaminophen directly onto a crystalline excipient to achieve a stable polymorph form via epitaxy,[215] leading to the possibility of heterogeneous nucleation of API onto excipients without the need for the removal of heterogeneous nucleants on which crystallisation occurs. While the above epitaxial based methods were suitable for certain API crystals, obtaining lattice match with macromolecular crystals are problematic as highlighted in Chapter 3.3.1. Mesoporous silica can be a possible alternative as it was demonstrated to be non-toxic in biological environment that it has been considered as delivery vehicles of bioimaging, biocatalysis and biosensing agents.[216] Mesoporous silica was used in the studies described in the next two chapters as the design processes of these materials allows for the specific design of pore size and the ability to achieve narrow pore distribution.[29] Also, due to the amorphous nature of the material, reduces the lattice strain of the crystal formed that destabilises the crystallisation process as a result of lattice mismatch.[196] In addition to proteins, the use of silica based material as a heterogeneous surface to influence nucleation processes has been reported in various literature for other materials. Rengarajan and co-workers (2007) reported the crystallisation of an unstable polymorph of acetaminophen with the use of mesopores of <10 nm which is usually inaccessible in its bulk solution form.[217] Dwyer and co-workers (2015) used controlled pore glass to confine the crystallisation of the poorly soluble drug, fenofibrate to form nanocrystals which showed

melting point depression and an enhanced dissolution rate.[218] Lapidot and Heng reported the use of surface functionalised mesoporous silica to influence the nucleation process of calcium sulfate as a method to address the problem of fouling in heat exchangers, and established a relationship between the particle loading of silica of different surface chemistry on the induction time.[219] To the author's knowledge, there has been no systematic investigation of the effects of pore diameters or the seed concentrations of mesoporous silica on the batch crystallisation of proteins under agitation.

Seed surface properties

Different types of heterogeneous seeds were described in Chapter 3.3.2. While surface chemistry and roughness were reported to influence nucleation rate, this report focuses on the effects of pore diameter.

While the work by Chayen and co-workers (2001) believes in the use of disordered pores and Shah and co-workers (2011) argued on the importance of confinement effects of specifically designed pore sizes, the nucleants used by both researchers contain pore dimensions comparable to the typical diameter of the proteins crystallised in the reports. The confinement effects are also supported by the effectiveness of the use of MIPs. To demonstrate the effects of the specifically designed pores on nanotemplates, crystallisation in the presence of these materials are compared with the use of other silica based heterogeneous seeds. In this case, in addition to the nanotemplate, mesoporous silica of alternative pore size and non-porous silica are also used as heterogeneous seeding materials.

The seeds used in this PhD are purchased and are used as received. Mesoporous silica are SBA-15 type particles purchased from Glantreo Ltd (Cork, Ireland) and the non-porous silica from Sigma-Aldrich (≥99.5 % trace metal basis, 637238 Silicon Dioxide). The mesoporous silica NT40 and NT120 are named according to the pore sizes in Angstroms the supplier claimed to adopt, and were chosen based on the measured pore sizes, which are comparable with the size of the protein measured.

Nitrogen adsorption studies were conducted to investigate the surface properties for a larger degree of confidence. These materials were degassed for 24 h at 100 °C, and afterwards analysed by nitrogen adsorption studies at 77 K using instrument Micromeritics Tristar 3000 (Micromeretics, Norcross, USA). Surface properties of all seeds used (Table 1) were calculated by the software based on the adsorption and desorption isotherms obtained (Figure 19). Figure 20 gives an indication of the pore size distribution based on BJH calculations using the

desorption isotherms. As evident from microscopy (Figure 21), the mesoporous silica are quite irregular in shape and sizes.
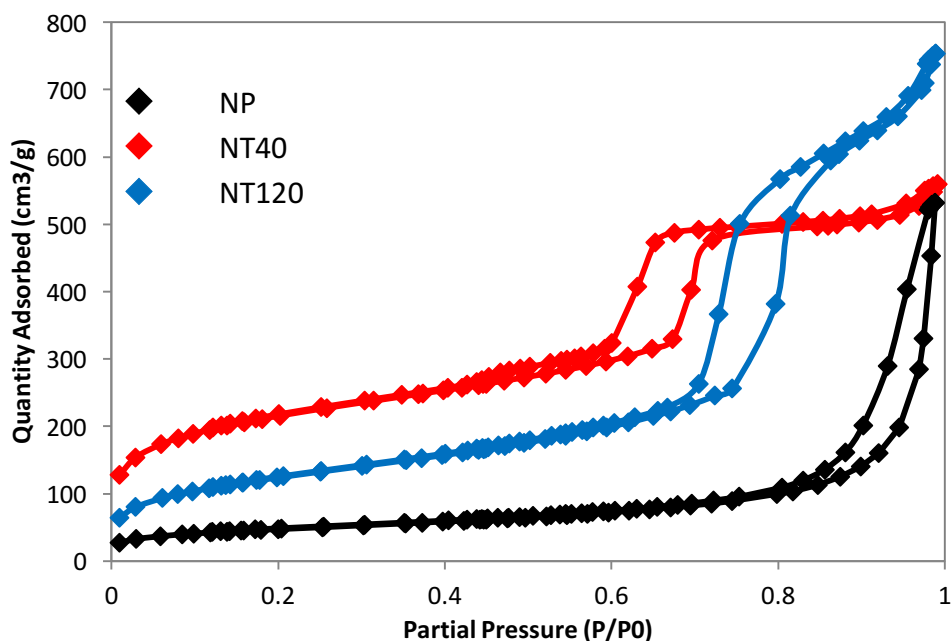
Figure 19 Nitrogen adsorption and desorption isotherm for NT40 (Red), NT120 (Blue) and NP (Black) seeds

As expected, the isotherm of both NT40 and NT120 displays a Type IV isotherm behaviour, suggesting the mesoporosity. Type IV isotherms are characterised by a 'shoulder', commonly identified as 'Point B' (shown in Figure 19), corresponding to the full coverage of monolayer, and the hysteresis indicates the condensation process taken place within the pore. The hysteresis of these isotherms can be assigned as H1 types, in which the loop is parallel, which indicates the pore shape regularity. Pores can be cylindrical, or can adopt ink bottle or slit shaped pores. H1 isotherms indicate the pores are cylindrical.

Table 1 Summary of surface properties of the silica based heterogeneous seed

	Pore Diameter (nm)	BET Specific Surface Area (m^2/g)	BJH Specific Pore volume (cm^3/g)
NP	-	171.0	-
NT40	5.6	769.6	0.81
NT120	8.9	451.8	1.08
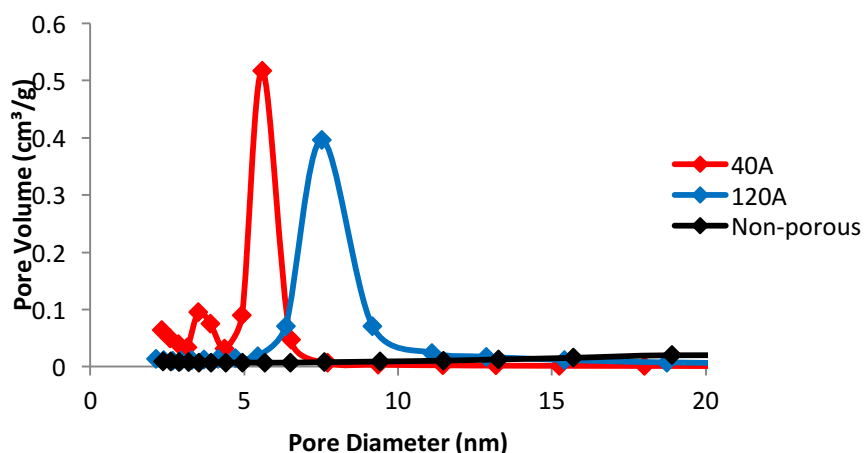
Figure 20 Pore size distribution based on BJH calculation
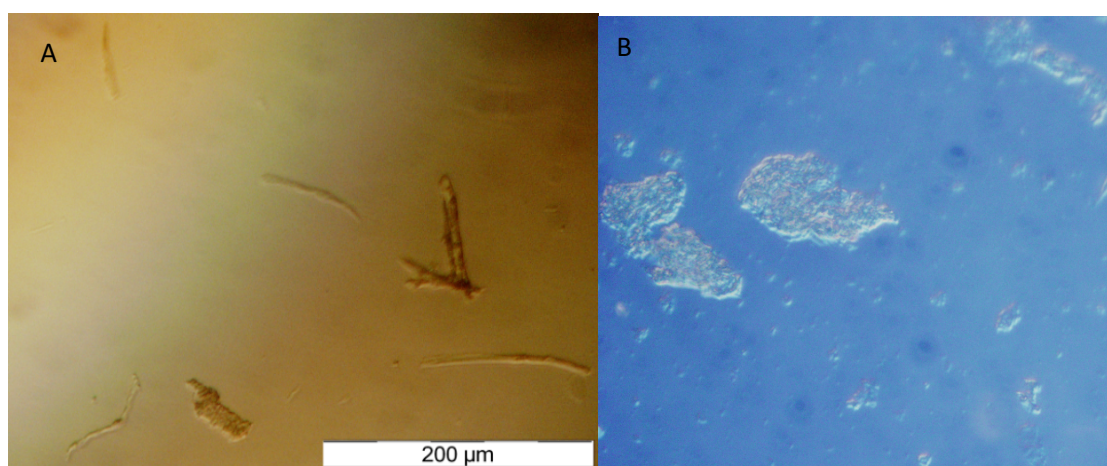
Figure 21 Microscopy image of (A) NT120 and (B) NT40

Seed Concentration

As discussed previously, seeds play a role in the stabilisation of the batch crystallisation process in the crystallisation of small molecules (Kubota, 2001). The quantity of seed used is typically expressed as the seed loading. This is defined as the ratio of the seed used (by mass) to the maximum theoretical yield of crystals as determined from solubility.[220] Typical seed loading used in literature was reported to be 0.5 – 10 wt %, which was adopted for the interest of obtaining a narrow CSD.[10] Limited research has reported on the use of heterogeneous seed to promote the nucleation of crystals, seed loading used in the experiments reported in later section are also used in this range. However, in the experiments discussed in the next two chapters, the mass of seed was described in terms of the working volume, as this does not fit the definition of 'seed loading', also, it is quantified here as 'seed concentration'.

Due to the differences in specific surface properties (see Table 1), a fair comparison cannot be drawn by sthe control of seed used for each seed type by mass. The seed

concentration is therefore varied, so the effects of varying surface pore volumes and surface area can be compared.

4.3 Induction time measurements and interpretation

The relationship between nucleation time with respect to the thermodynamics of nucleation was discussed in Chapter 2.4. In a practical perspective, where different analytical technique has a different detection limit, the induction time, t_{ind} is used, which accounts for both the nucleation time, τ, and the time required for crystal nuclei to grow into a detectable size, t_{growth} (Equation 26).[221]

$$t_{ind} = t_{nuc} + t_{growth} \ (26)$$

A typical crystallisation profile can be represented as Figure 22, the concentration of a protein was monitored throughout a period of time, in which nucleation of the crystals are progressing through. Solution remained metastable until a sharp decrease in protein concentration was observed. As expected, the protein concentration does not decrease linearly. An S-shaped curve is typically associated with the phase change processes as described by the Johnson-Mehl-Avrami-Kolmogorov (JMAK) Equation.[222] In all experiments described in the next two chapters, the protein concentration was only monitored when an observable decrease in concentration (~4 mg/mL) was reached after the induction time. From this curve, tangents of the inflection point at the desupersaturation can be drawn to intercept with the protein initial concentration in the manner of the blue lines in Figure 22, and the interception can be taken as the induction time.
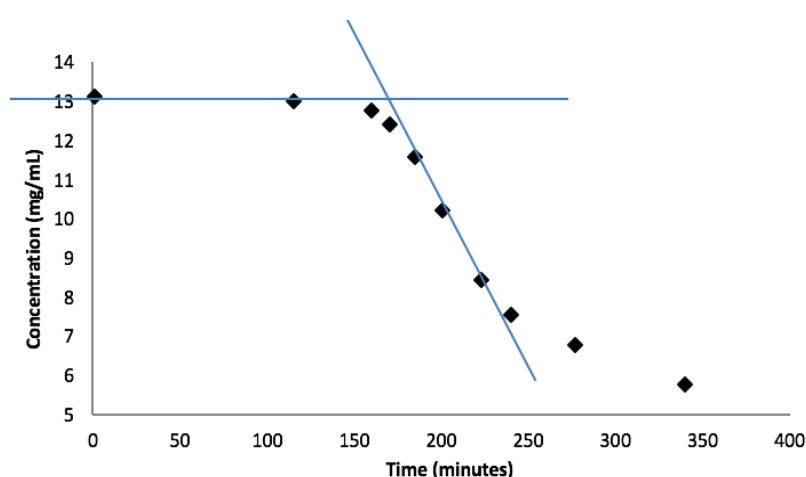
Figure 22 A typical crystallisation curve corresponding to changes in protein concentration over time obtained by offline-sampling for UV-vis measurement.

UV-vis Spectroscopy

UV-vis spectroscopy is a spectrophotometric technique relating the concentration of a component in solution to the electronic absorbance in the UV/vis region via the Beer Lambert Law.[223] This technique is applicable to solutes of delocalised electronic systems in their molecular structures, with each compound giving its own characteristic spectrum. In the case of proteins, a characteristic peak at the near UV region (240 – 300 nm) peaking at 280 nm is typically observed corresponding to the aromatic amino acids such as tryptophan, tyrosine and phenylalanine, the absorbance at 280 nm is therefore typically used as a standard wavelength for measurements.[224] Typically, the relevant concentrations for protein crystallisation (5-30 mg/mL) result in absorbance values that deviate from the linear Beer-Lambert relationship in a standard spectrophotometer where the pathlength would be 1 cm for ease of calculation. To overcome this problem, dilution of protein solution of interest is often required for conventional spectrophotometers.[225] However, for the purpose of this work, where the concentration is monitored throughout the whole process, the error resulting from this dilution process affects the quality of the crystallisation profile. Alternatively, spectrophotometers of narrow and variable pathlengths can be used to preserve the linearity suggested by the Beer Lambert Law at relevant concentrations. The use of such specific equipment requires the offline sampling of solution. While this technique can confirm the decrease in protein concentration in the solution phase, additional analysis would have to confirm the formation of protein crystals.

Turbidimetry

Turbidity is a measurement of the reduction in transparency in liquid caused by undissolved matter.[226] Quantification of this property is either based microscopic particle count or the quantification of light obscuration resulted from undissolved particles.[227] In the study relating to the crystallisation of catalase, the latter technique was used, and turbidity was therefore given as a function of light reaching a detector position through a suspension with respect to a beam source.[38] This serves as an automated technique to quantify the solid phase forming process and have been used in the experiments on the crystallisation of catalase.

As this is a technique that involves monitoring light transmission, the turbidity, τ , depends on both the particle sizes (represented by the diameter of particle, r) and the concentration of solid particles, N, and can be related by Equation 27 for a monodispersed system. The light transmission, T_r through a sample of a defined wavelength, l, as a result as the turbidity is represented by Equation 28. Q_{ext} represents the Mie scattering coefficient, which

is a function of the wavelength of the light source, λ_0, particles size, and the refractive indicies of both the solution and the solid particles. [228]

$$\tau(\lambda_0) = N\pi r^2 Q_{ext} \quad (27)$$

$$T_r = \exp{(-\tau l)} \quad\quad (28)$$

These equations are used to illustrate the idea that the increase in turbidity measurement may arise from either the formation of particles (nucleation) or crystal growth. In the system described in Chapter 6, the crystallisation processes were not strictly controlled to achieve a monodispersed suspension, and polydispersity is expected. The presence of the irregularly shaped mesoporous silica of size ~100 μm would also contribute to the polydispersity of the equation. Therefore, the relationship cannot be directly used to model the processes in the experiments.

5. Crystallisation of Hen Egg White Lysozyme

This chapter describes the experimental set-up and conditions, methods and results of batch crystallisation of Hen Egg White Lysozyme. Crystallisation curves were obtained by monitoring the concentration offline throughout each experiment. The induction time deduced from these curves was used to represent the effects of different conditions on the nucleation process.

5.1   Materials
Hen Egg White Lysozyme (HEWL)

Lysozyme is described as a bacteriolytic agent known to be present in different species of animals, and was found to be the most concentrated in hen egg white.[229] The size of HEWL is approximately 14 kDa, consisting of 129 amino acid residues, with an isoelectric point at pH 11. The dimension as reported be 45 x 30 x 30 Å according to crystallographic studies at 2 Å resolution.[230] Its crystallisation properties are well studied and is frequently used as a model protein. Its solubility properties at different conditions (salt concentration, temperature and pH) have been extensively reported, and as a model protein it has been used in thermodynamics studies.[231–233] Because of its well understood thermodynamic properties, it was chosen as our main model protein for the experimentation described in this section.

HEWL was first crystallised by Alderton and co-workers (1945), in which the protein was first isolated from egg white by ammonium sulphate precipitation and redissolved in acetic acid, followed by the crystallisation with the addition of 5 % sodium chloride. Crystals were observed in pH as low as pH 3.5 up to its isoelectric point and the shape of crystal varies across the pH range.[234] Systematic investigation of the effects of pH showed moderate crystallisation from pH 4.8 to pH 7.2, in which crystals were octahedral; and copious crystallisation above this range, in which spherulitic structures were observed.[235] Most studies on lysozyme reported crystallisation condition between pH 4 to pH 7 with sodium acetate as stabilising buffer (25 mM) and sodium chloride (2-5 % m/v) as precipitant.[168,169,178,179,182,184] Numerous methods of solubility measurements were demonstrated using lysozyme. Selecting a crystallisation experiment within this condition range would therefore be suitable.

At the working conditions described, the hydrodynamic diameter was measured to be 4.5 nm in the crystallising conditions used with the use of Dynamic Light Scattering, which does not deviate significantly from the dimensions obtained from crystallographic studies reported in literature.

5.2   Method

Experimental Set Up

The effects of different silica seeds were investigated through a batch crystalliser. Figure 23 depicts the setup of the crystalliser and its dimensions. The temperature was controlled at 5±0.01 °C by circulating a 20 % (v/v) ethylene glycol solution through the vessel jacket and a water bath (make). The working volume was 25 mL, such that the thermocouple is submerged in the crystallisation solution, and that withdrawal of solutions throughout experiment (~20 samples, in aliquots of <200 µL) would cause a negligible change to the geometry of the crystallisation volume. The stirring rate was kept constant, three stirring rates were compared: 50 rpm, 100 rpm, 200 rpm.
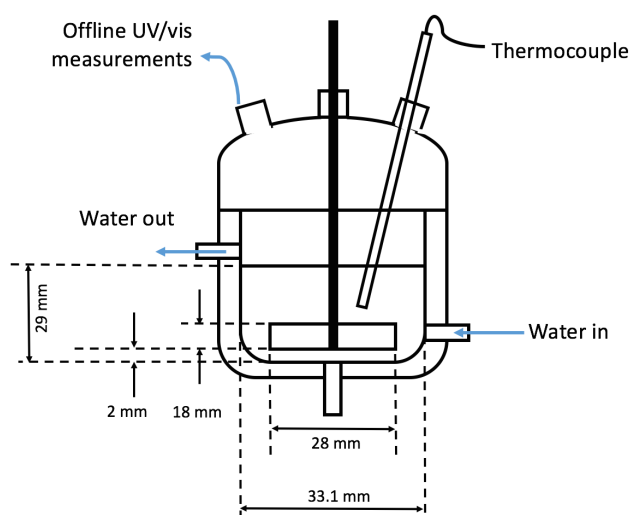
Figure 23 Crystallisation vessel set-up

Concentration Calibration

The concentration of HEWL was monitored using Nanodrop ND2000c, Thermo Scientific. This set up consist of a variable pathlength (0.05 – 1 mm) and gives linearity up to high absorbance level (A = 300, 10 mm equivalent) as specified by supplier, which was well above the maximum absorbance level of the protein at the concentrations used within this thesis.

Solutions were prepared by weighing HEWL crystalline powder (~70 000 U/mg, 62971 Sigma Aldrich) into large volumes (50 mL) of buffer solution in a volumetric flask to make up a concentration expected for the solution (x-axis in Figure 24), such that the human error associated with weighing and volume of solution was minimised. The measurements (y-axis in Figure 24) were taken with the sodium acetate buffer solution as a background reading. While the software for the spectrophotometer was equipped with a pre-calibrated relationship for the

concentration of lysozyme, there is a discrepancy between the concentration of lysozyme expected from solution preparation and the measured concentration given (Figure 24). The difference in the two values at high concentration was beyond human error. Concentrations are described in the results within this chapter as the measured concentrations (y-axis).
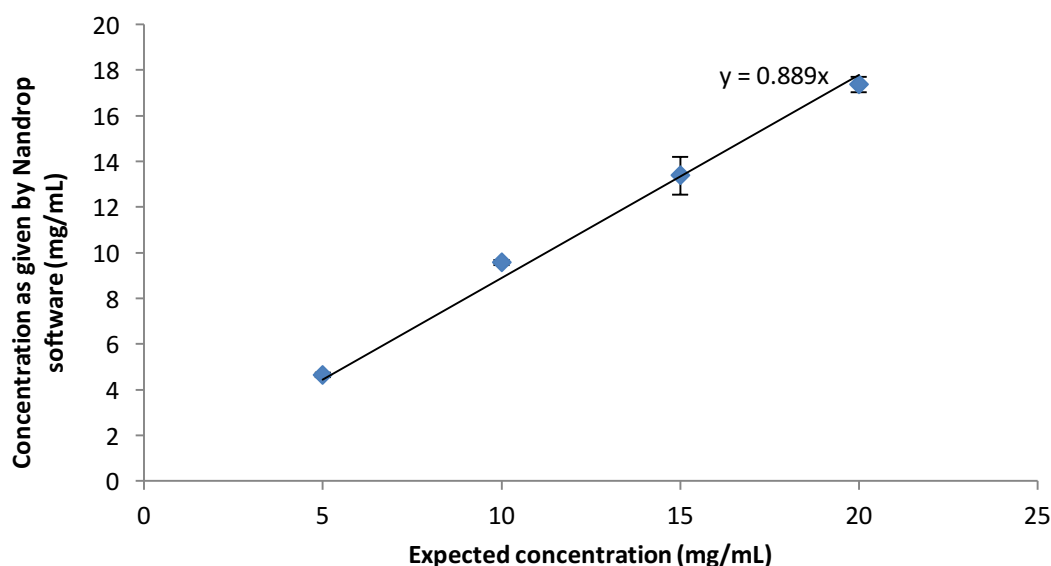
Figure 24 Calibration plot for lysozyme concentration measurement

Pre-experimental preparation

Prior to protein solution preparation, stocks of buffer solutions of sodium acetate, NaAc (25 mM, pH 4.8± 0.01) were prepared using crystalline sodium acetate trihydrate (≥99.5% purity, S7670, Sigma Aldrich). The pH was adjusted to pH 4.8 using diluted acetic acid and sodium hydroxide solutions. Stocks of precipitant solutions (NaCl, 1M, NaAc, 25 mM, pH 4.8± 0.01) were prepared and the pH was adjusted in the same manner as the buffer solution but with the addition of crystalline sodium chloride (≥99.0% purity, S7653, Sigma Aldrich).

Prior to each experiments, solutions of HEWL (27-35 mg/mL) were prepared by dissolving crystalline HEWL powder (~70 000 U/mg, 62971 Sigma Aldrich) into the buffer solution prepared above (filtered with 0.22 μm Milipore syringe filters). The concentration was double checked with UV/vis spectroscopy at A_{280} (Nanodrop ND2000c, Thermo Scientific). This solution was incubated overnight at 5±2 °C until start of experiment. No solid formation was observed and the concentration was unchanged when kept at this condition (in the absence of precipitants) for > 14 days. The precipitant solution was also filtered with 0.22 μm Milipore syringe filters and kept in separated containers in the same incubator for the same amount of time.

For seeded experiments, seeds stocks were prepared by dispersing NT120 and NT40 (SBA15100120R and SBA1510040R respectively, both from Glantreo Ltd, Cork) and NP (≥99.5 % trace metal basis, 637238 Silicon Dioxide, Sigma Aldrich) into filtered precipitant solutions such that the seed concentration was twice the concentration used in experiments. These were again incubated to reach the desired temperature prior to experiments. Due to the need of the presence of these seeds, the suspensions cannot be filtered again and was not kept for over 24 h after filtration. Heterogeneous seeds were dispersed in the precipitant solution by stirring prior to start of experiment.

Experimental Conditions

Crystallisation experiments reported within were kept at 3% m/w NaCl concentration and isothermally at 5 °C. Most reported lysozyme solubility data was reported between 10-25 °C, in which the supersolubility was reported to be over 20 mg/mL.[181,236] To limit the amount of protein required to achieve supersaturation in the comparatively large working volume, a lower temperature was used in. The working temperature used in these experiment was 5 °C, further decrease in temperature results in minimum changes in solubility. The proximity of working temperature to the freezing point of water also render cooling crystallisation impractical.

Solubility Determination of Lysozyme

The solubility of lysozyme was obtained from literature. Cacioppo and Pusey (1991) studied extensively the pH from 4.0 to 5.4, and formulated the relationship between solubility and temperature at different pH levels with third order polynomial fits for these conditions, which gave a solubility of 2.3 mg/mL,[231] whereas the data reported by Forsythe and co-workers (1999) reported a solubility of 1.3 mg/mL.[233] Most other studies of lysozyme in literature discussed previously reported solubility at higher temperature (>10 °C). The two values available deviate quite significantly with each other. The solubility in the condition chosen was therefore investigated.

The solubility studies conducted in this experiment is based loosely on the solvent addition method described in [[161]]. A 13.1 mg/mL HEWL solution was left to crystallise and the residual concentration of the suspension was measured after 24 h after start of experiment, such that the rate of desupersaturation was sufficiently low and the concentration is close to equilibrium. Crystalline suspensions were withdrawn and were immediately diluted by known ratios (0.05 to 0.5) of blank solution (crystallisation solution at 0 mg/mL at 5±1°C) and was

allowed to equilibrate in this condition. A concentration measurement was taken once after 24 h and again after 48 h to ensure equilibrium concentration is reached.

Crystallisation Experiment

The reaction temperature was lowered to 5 °C prior to the addition of crystallisation solution. Both the protein solutions and the precipitant solutions were filtered again immediately before they were mixed into the vessel to reduce the possibility of bacterial growth. The precipitant solution (12.5 mL) was added into the vessel as the HEWL solution (12.5 mL) was stirred. The effective concentration of HEWL then became 13.5-17.5 mg/mL (error of ±0.5 mg/mL). Experiments were rejected if the initial concentration was beyond the error quoted. Time at t=0 was taken as the point at which both solutions were poured into the vessel. <200 µL was withdrawn at regular intervals for concentration determination with UV/vis spectroscopy at A_{280} (Nanodrop ND2000c, Thermo Scientific). The error of the time of the sampled data point would be ± 1 minute, corresponds to the maximum error of 10 % compared to the *shortest* sampling interval (10 minutes), and negligible compared to the variation in data. 3-5 measurements were taken per data point, depending on the aliquot sampled, to obtain a mean concentration of the crystallisation solution sampled. The induction time was determined based on the shape of the crystallisation curve plotted from these data points.

5.3   Results and Discussions
5.3.1 Solubility Determination of Lysozyme:

A crystallisation solution of 17.5 mg/mL was kept at 5 °C for 24 hours and the concentration was monitored at regular interval, the concentration reached ~3.5 mg/mL after 24 hours with negligible changes upon further measurement. Aliquots were withdrawn at this condition and was diluted by the 'blank solution' by the ratios shown in Figure 25. The black line through the origin represents the concentration that the mixture would have if there were no crystal growth/ dissolution processes occurring. Data points above this lines suggest dissolution of crystals occurred as a result of the dilution, data points below this line represents crystal growth to reach equilibrium, the latter case was expected to be slow. Crystals were still present for the suspensions of dilution factors 0.25, 0.4 and 0.5 when sampled for concentration measurements, and the first two suspensions were therefore placed again in the incubator to equilibrate for another 24 hours. Both mixture at 0.25 and 0.40 dilution factor equilibrated at ~1.3 mg/mL, which can be taken as the solubility, corresponding well with Forsythe et al (1999).
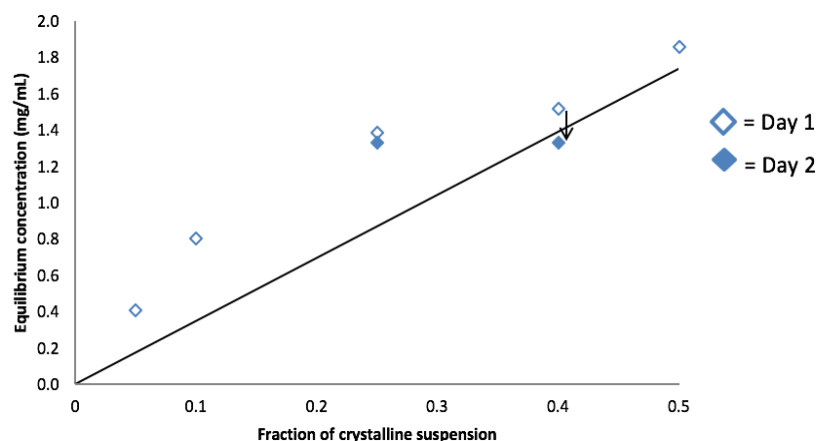
Figure 25 A plot of residual concentration of HEWL solution (y-axis) when a crystalline suspension (supernatant concentration of ~ 3.5 mg/mL) was withdrawn and diluted by the dilution factor represented in x-axis.

5.3.1 Unseeded experiments- Concentration Variation:

Unseeded experiments were conducted at the protein concentrations 17.5 mg/mL, 15.5 mg/mL and 13.5 mg/mL at a fixed stirring rate of 50 rpm at 5 ˚C, with the induction time represented in Figure 26. The crystallisation experiment was only monitored until a significant decrease in concentration was observed such that the induction time can be deduced. This set of experiments serves as a control to compare with all other seeded experiments. Multiple runs were shown for each condition to indicate the reproducibility of results. An increase in concentration by ~4 mg/mL results in a reduction in induction time by a factor of 4. As the induction time increases, the variation in data also increased. This can be attributed to the stochastic nature of the nucleation process due to kinetics element, despite the supersaturation being controlled.[89,237] However, results have shown that at different protein concentration investigated, the induction time is significantly different between each increment for comparison, which can be summarised in Figure 30.

The gradient of the curve after the induction time was significantly less steep at 13.5 mg/mL, the desupersaturation corresponded to a ~3.5 mg/mL drop in ~5 hours. This is expected as both nucleation and the growth rate are dependent on the supersaturation of the conditions. The relative contribution of the two processes to the desupersaturation of proteins was unclear in these experiments.
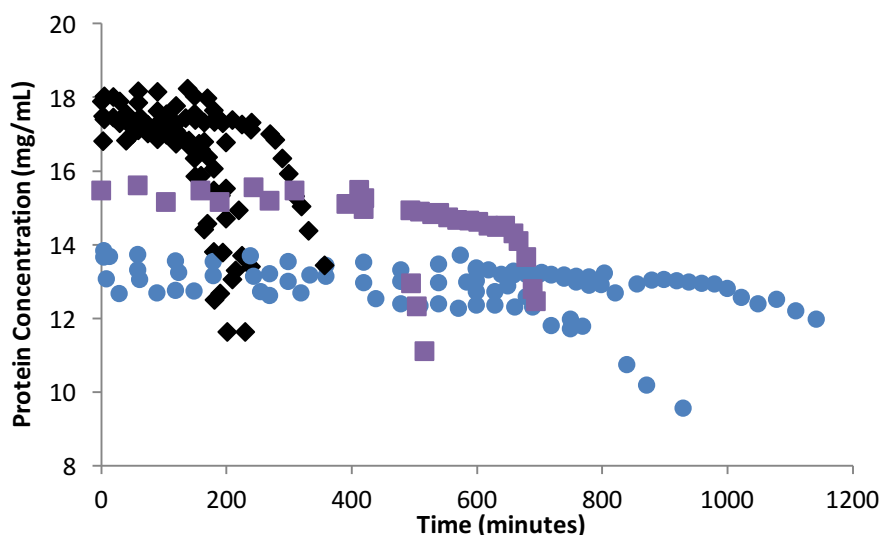
Figure 26. The crystallisation profile of lysozyme at various concentration (◆= 17.5 mg/mL, ■= 15.5 mg/mL, ●= 13.5 mg/mL) in the absence of seed, with stirring speed = 50 rpm.

Images of lysozyme crystals are obtained in Figure 27. No qualitative differences were observed between the appearance of crystals obtained at different supersaturations investigated. The crystals were less than 10 µm in length and needle-like. In agitated conditions, the edges of the crystals are less well defined compared to those obtained in quiescent mode (Figure A1 in appendix).
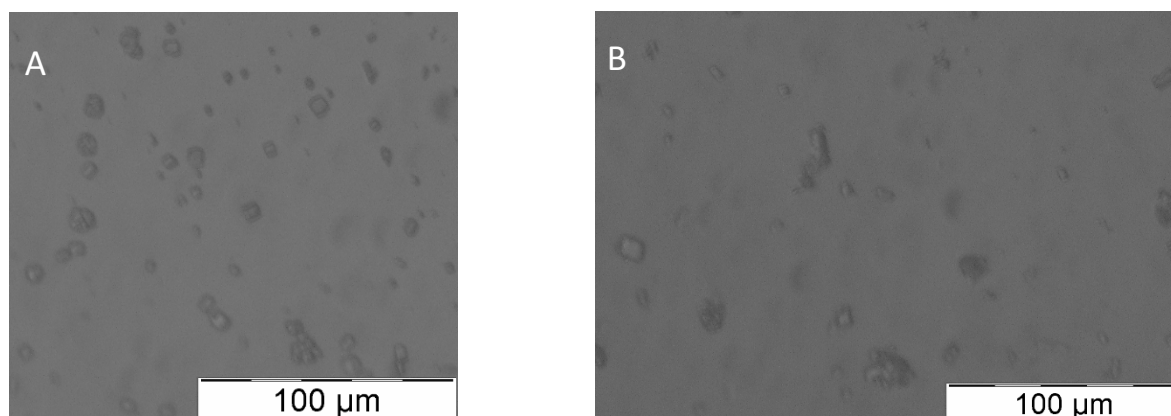
Figure 27 Appearance of crystals grown in (A) 17.5 mg/mL solutions and (B) 13.5 mg/mL solutions

5.3.2. Seed Type Variation

To compare the effects of the nanotemplates, other silica based seed types (NT120 and NP) in addition to NT40 were used. Initial seeding experiments were conducted at a protein concentration of 13.5 – 17.5 mg/mL and a seed concentration of 1.25 mg/mL. Sample were withdrawn from reaction vessel 24 hours after the start of experiment and was analysed by optical microscopy (Figure 28 a-d). Crystallisation profiles for these experiments is presented in Figure 29.

Needle shaped crystals was observed to have formed on the silica particles, supporting the effects of heterogeneous nucleation. Significant amount of individual crystals was also observed which are not grown on the surfaces of silica particles or other protein crystals. However, based on what was observed with microscopy, the shape and sizes bear no significant differences as those nucleated on surfaces, which may suggest the limited effects of silica particles on crystal growth. Figure 28D shows lysozyme grown in 13.5 mg/mL in the presence of NT120 seeds (similar condition as Figure 2A), but stained blue using Izit Crystal Dye (HR4-710, Hampton Research) to confirm the identity of the material was protein crystals. The alignment of the protein crystals with respect to the seeds (non-birefringent and non-transparent) was also visualised.
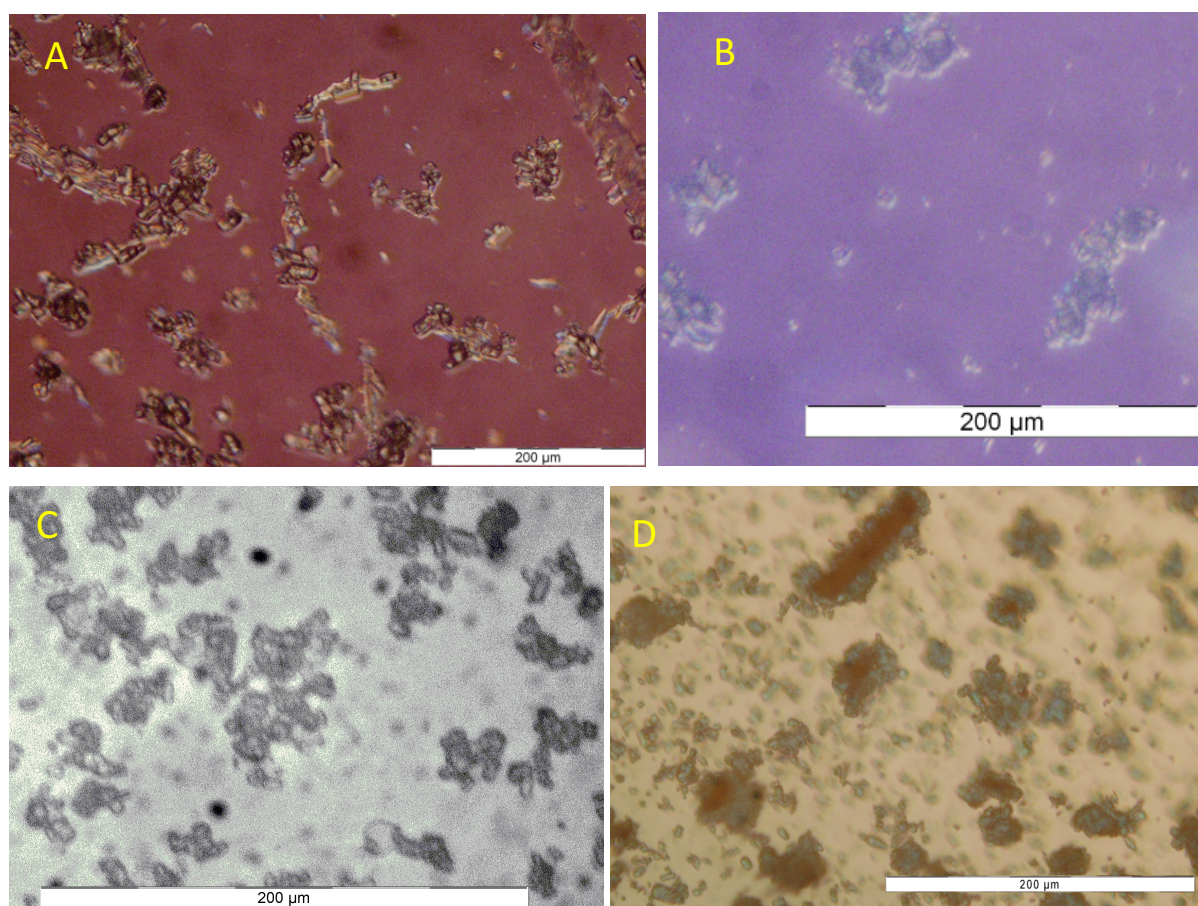
Figure 28 Appearance of HEWL crystals formed on A) NT120 and B) NT40 and C) NP seeds at 13.5 mg/mL. Crystals in (D) are grown in the same condition as (A), and stained with protein crystal dye.

As expected, the variation in induction time in these seeded conditions is much smaller than in the unseeded experiments, suggesting some controls on the nucleation of protein crystals are imposed by the seeds (Figure 29). At an unseeded condition, the variation in induction time was almost 2 hours when the mean induction time value was ~2 hours at 17.5

mg/mL. The variation in induction time was ~15 minutes (mean induction time ~ 80 minutes) when seeded with NT120 at the same concentration, and lower for the two other seeds used. Contrary to expectations, NT40, the seed with the highest surface area while and the closest protein diameter/ pore size match, resulted in a much longer induction time, which is just below the induction time of unseeded conditions. At 1.25 mg/mL seed concentration, both NP and NT120 seeds reduced the induction time compared to unseeded experiments quite significantly, with NP giving a slightly shorter induction time (~45 minutes). The induction time as the result of the same NT40 concentration was just less than half of the unseeded induction time.
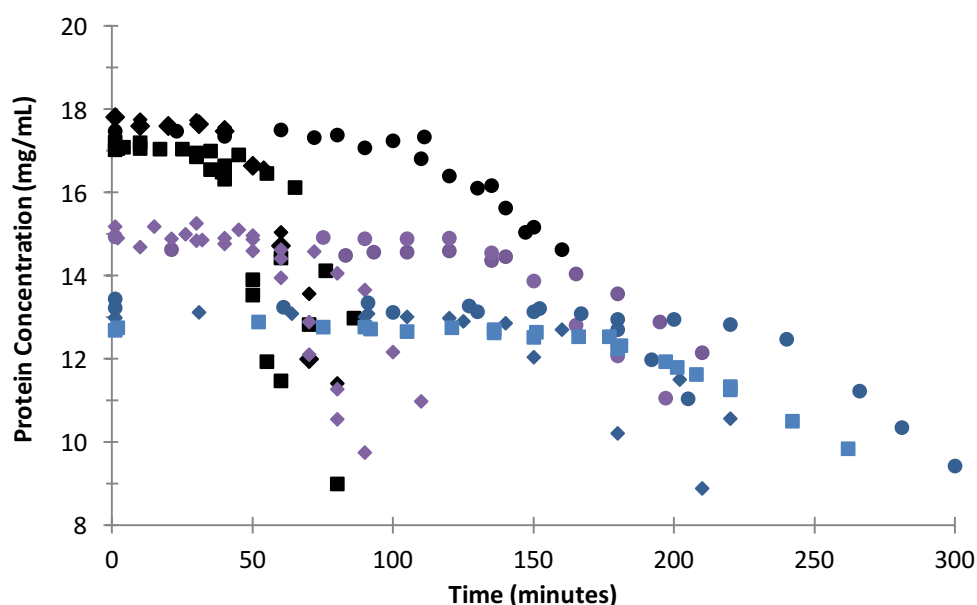
Figure 29. The crystallisation profile in the presence of different seed types (non-porous (◆), NT120 (■) and NT40 (●)) under different concentrations.

When the same mass of silica is used at a reduced protein concentration (13.5 mg/mL), the effects of seed addition are a lot clearer compared to unseeded experiments are a lot more significant. As shown in Figure 29, the induction times for all seeded experiment at 13.5 mg/mL are all under 5 hours, corresponding to ~20-25 % of the induction time at unseeded conditions. The distinction between the three seed types, however, are not as clear. These values are summarised in Figure 30 and was compared with the unseeded experiments at different supersaturation. It can be projected from the graph that further decrease in supersaturation would result in an even stronger effect on the nucleation time, whereas the induction time for unseeded experiment might be sufficiently short that the effects of seed would be unclear. Seeded crystallisation using NT40 was attempted at 8 mg/mL, however, concentration remained stable after 24 h at the operating condition (5 °C). This condition is

therefore assumed to be metastable. Wide MSZW was reported for lysozyme at this 5 °C, 3 % m/v NaCl elsewhere in the literature.[181] Concentrations explored were therefore limited to the concentrations discussed.
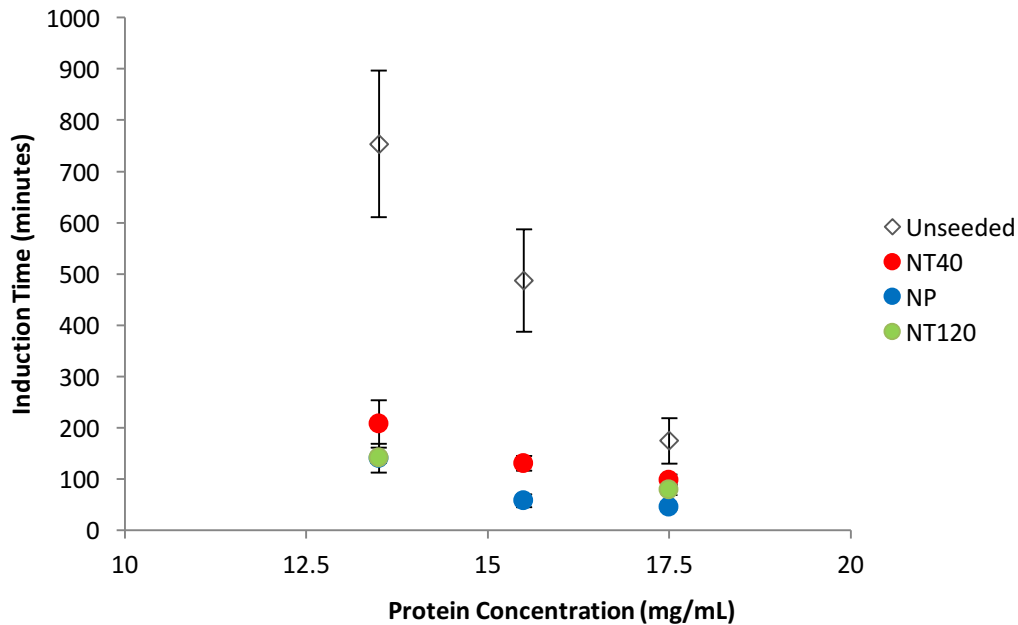
Figure 30 Summary of effects of different types on nucleation time at various protein concentration

Summarised to Figure 30, it also appears that as the protein concentration decreases, the relative effects of the different seed type vary. Table 2 represents the mean induction time of different seeded conditions, t_{ind}(seeded), as a fraction of the mean induction time of the unseeded experiment, t_{ind}(unseeded), at HEWL concentrations of 13.5 mg/mL and 17.5 mg/mL. By quantifying the effectiveness of seed using this fraction, the difference between the porous and non-porous materials can be observed. The non-porous seeds reduce the induction time to ~ 0.20±5 of the unseeded induction time at both 13.5 and 15.5 mg/mL. When both porous seeds were not as effective at high supersaturation (t_{ind}(unseeded) / t_{ind}(seeded) were 0.56 and 0.46 for NT40 and NT120 respectively), the values were compatible with the NP seeds at reduced protein concentration. While this cannot be further extrapolated to lower or higher supersaturation based on information provided in the graph, it is worth noting that the effectiveness of different seed types in reducing the induction time also has a dependency on the protein supersaturation. Experiments were not conducted at higher supersaturation, as discussed previously, the effects of seed are expected to be unclear as short induction time can be expected.

Table 2 The effects of seed (1.25 mg/mL) at varying supersaturation of proteins

Protein Concentration (mg/mL)	t_{ind} (seeded) / t_{ind} (unseeded)		
	NT40	NT120	NP
13.5	0.26	0.18	0.17
17.5	0.56	0.46	0.25

Differences in the slopes of crystallisation at the desupersaturation was apparent for 17.5 mg/mL. It can be clearly demonstrated that the NP seeds have a much steeper desupersaturation slope, compared to NT40. By approximating this region as linear, the rate of desupersaturation was estimated, this was described in per half an hour as the gradient changes through time, and from the data points obtained for 17.5 mg/mL the change in gradient one hour after t_{ind} is uncertain. At 17.5 mg/mL, NP seeds would correspond to a desupersaturation rate of 5.2 mg/mL per half an hour, while NT40 would correspond to a desupersaturation rate of 1.6 mg/mL per half an hour. Later figures (Figure 31-33) would show that the desupersaturation region of the porous materials was of a similar gradient as the unseeded experiments at 17.5 mg/mL.

Referring back to Figure 26, where the desupersaturation region was less steep for 13.5 mg/mL, there is a reduction in gradient at unseeded experiment. This is also true for NP seeds. At 13.5 mg/mL, the slope was reduced to ~1.6 mg/mL per half an hour. This is similar to the value for NT40 seeds, which varied by minimal extent at all protein concentrations studied. A correlation between the seed influence on induction time and the desupersaturation slope can be made qualitatively.

5.3.3 Seed Concentration Variation

Due to the uncontrolled size of these seed particles and the difference in surface areas and pore volume associated with it, comparing crystallisation experiments with different seed mass used may not be suitable. As evident in Table 1, where the surface properties (pore volume, surface area) differ per unit mass as a result of varying pore diameter, the effects of pore sizes of seeds on crystallisation time are not as straightforward. Therefore, crystallisation of lysozyme at 17.5 mg/mL was tested across a range of seed concentrations corresponding to seed loadings 0.05 – 10% depending on seed types used, with results represented in Figure 31-33. These plots can be summarised by Figure 34. By varying the seed concentration of these different seeds, we aim to investigate the impact of the change of surface area vs the change of pore volumes.

For each of the three seed types investigated, it was found that a seed concentration exists where above this point, the induction time remains unchanged upon further increase in seed concentration (Figure 34). For the purpose of this project, the seed concentration at which the induction time decreases sharply shall be regarded as the critical seed concentration. This value varies for different seed type, and has a correlation with the resultant induction time. The lower the induction time, the lower the critical seed concentration. Regardless of the amount of seeds used, the induction time remain at around ~45, ~90, and ~100 minutes for NP, NT120 and NT40 seeds respectively above the critical seed concentration (as summarised in Figure 34). While the increased seed concentration may often be associated with increased nucleation site, this has not resulted in a linear decrease in induction time. These regions of seed concentration gave rise to a smaller variation in induction time data.

Below the critical seed concentration, both porous seeds NT40 and NT120 had crystallisation curves coinciding with the unseeded experiments. The reduced seed concentration of NP seeds also resulted in a longer induction time, however this value remained below the induction time for unseeded experiments. The curves in Figure 34 do not intersect, showing that at a protein concentration of17.5 mg/mL and the range of seed concentrations studied, NP seeds are more effective in reducing the induction time, while NT40 is the least effective. The effects in the further decrease in NP seed concentration would be of interest but was not pursued due to the anticipated error in seed concentration values. The seed concentration of 0.05 mg/mL also is converted to just less than 0.5 wt % of seed loading, which corresponds to the lower limit of typical seed used.[10] The variation in induction time was also larger, demonstrating the reduced effects of the seeds at those levels.

For the purpose of the work, the critical seed concentration was estimated to be halfway between the two data points in which a difference in induction time is observed. This is used to deduce the concentration of surface areas and pore volume present at the critical seed concentration (Table 3). The surface area present at the critical seed concentration for NT40 was 13 times larger than in the critical seed concentration of NP seeds and 5 times that of NT120 seeds. This suggest that the effects of specifically designed pores, whether it is the promotion or hindrance of nucleation, demonstrates an effect at agitated environment, that is more than the availability of surface nucleation site or geometrically enhancing the nucleant/cluster interface. The critical seed concentration for NT40 also correspond to over 2 times the pore volume of the critical seed concentration of NT120.
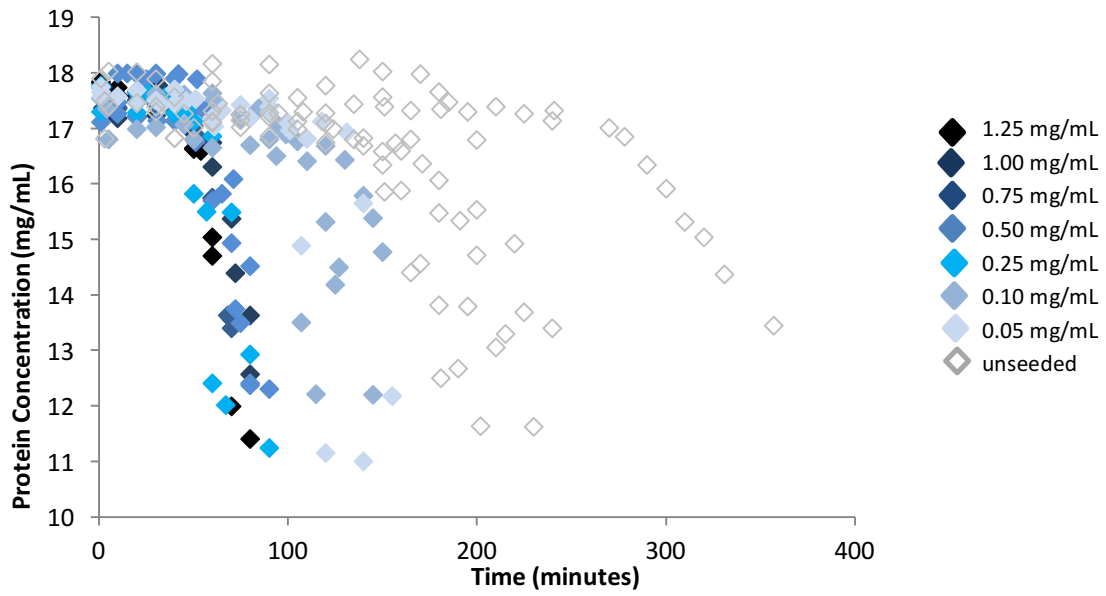
Figure 31 The crystallisation profile of lysozyme at 17.5 mg/mL under different concentrations of NP seeds.
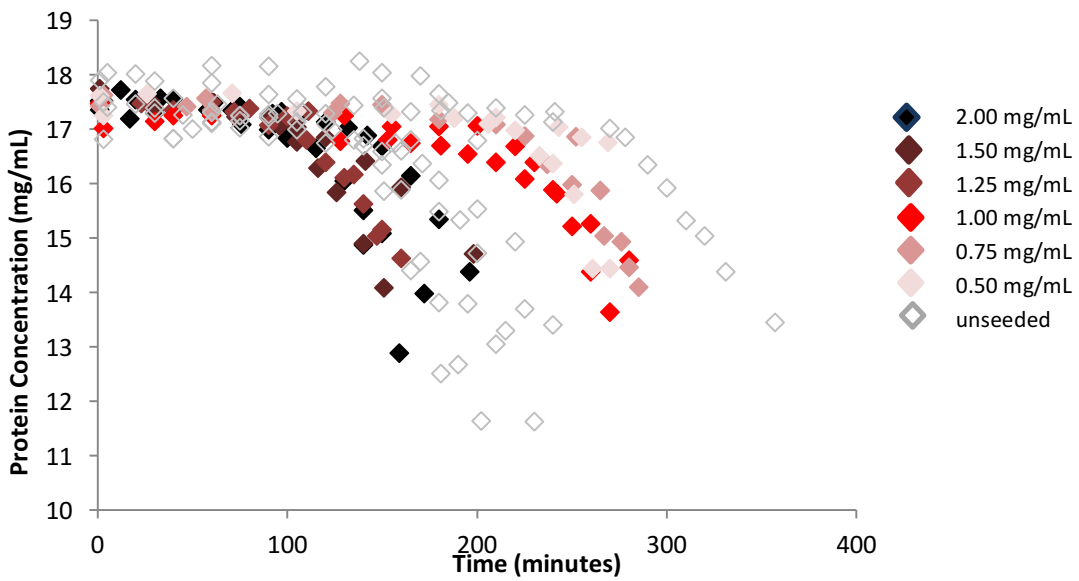
Figure 32 The crystallisation profile of lysozyme at 17.5 mg/mL under different concentrations of NT40 seeds.
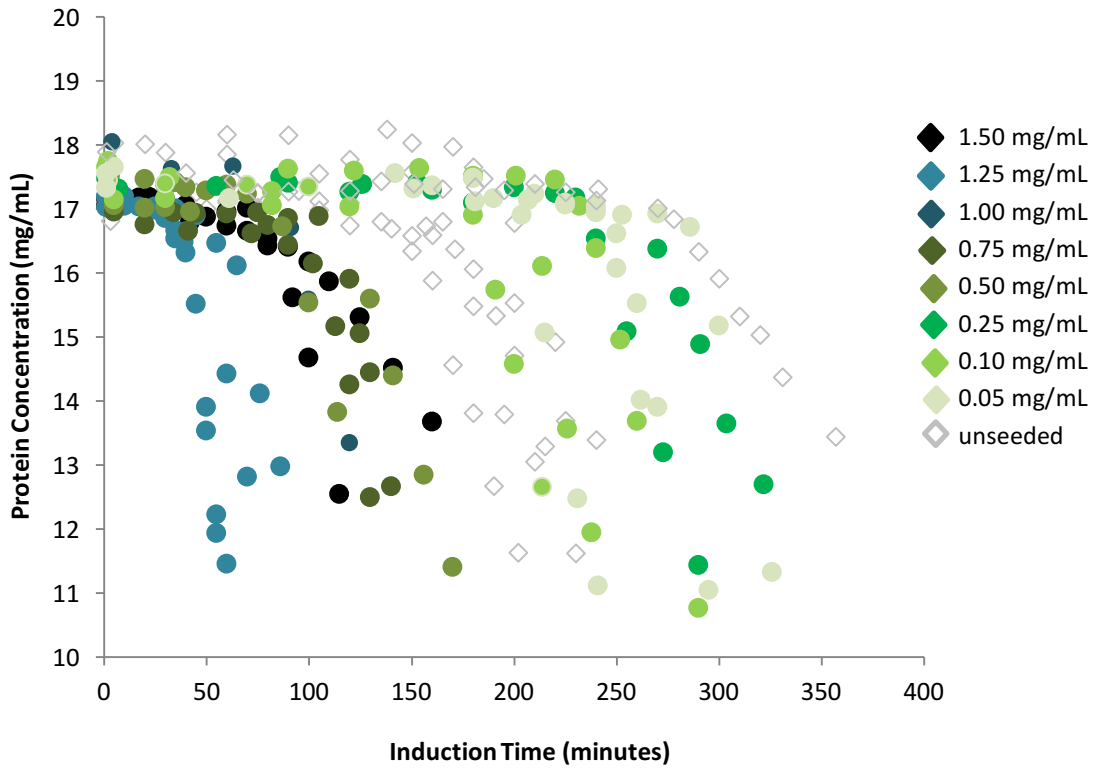
Figure 33 The crystallisation profile of lysozyme at 17.5 mg/mL under different concentrations of NT120 seeds.
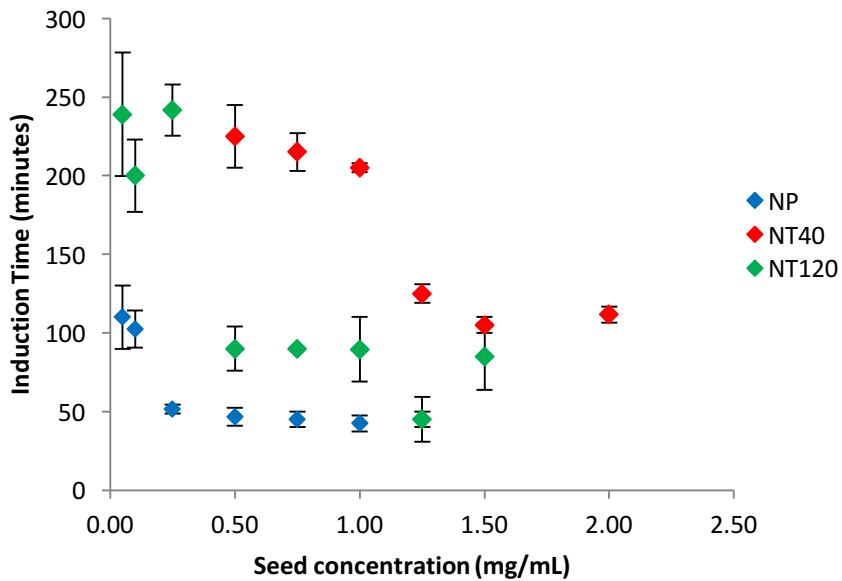
Figure 34 Variation in induction time across different seed concentrations

Table 3 Table summarising the concentration of surface area and pore volume at a given seed concentration

Seed Type	Critical seed concentration (mg/mL)	Corresponding surface area per volume (m^2/mL)	Corresponding pore volume per volume (cm^3/mL)
NP	0.175	64	-
NT40	1.125	866	0.918

NT120	0.375	170	0.405

It can also be shown that the crystallisation curve for NP seeds also adopt a much steeper desupersaturation curve than unseeded conditions, while the desupersaturation region for NT120 and NP are parallel to the crystallisation curve for unseeded experiments. The increase or decrease of seed concentration does not change the crystallisation curve.

Based on the knowledge from homogeneous seeding, it has been stated by Melia and Moffitt (1964) that nucleation rate (and hence induction time) is independent of the number of seeds used.[119] The correlation between induction time and seed used by Cayey and Estrin (1967) varies at different seed loading as the change in induction time was not linear and quite random.[238] Limited studies have been conducted on the use of heterogeneous surfaces on the influence of induction time at isothermal conditions. Weichsel et al (2015) investigated the effects of silica particle size used on the induction time of lysozyme but the amount of surface used was controlled.[197] In one study reported on the crystallisation of calcium sulfate in the presence of mesoporous silica particles with pore diameters of 6 nm, the induction time decreased linearly with the increase of these heterogeneous seeds and was attributed to the increased pore volumes present.[239] However, these particles can play a different role in protein crystallisation - While in small molecular crystallisation, the optimum pore should accommodate a certain number of lattice sites or a solute-rich cluster of defined size, as reported in various literature that was discussed in Chapter 2.4, the pores were selected to accommodate one protein molecule. As the cluster formation step for lysozyme was demonstrated in literature discussed in Chapter 2, these pores on nanotemplates might not be as efficient in promoting crystallisation in flow conditions. Allan and co-workers (2008) discussed the role of shear on the effects of cluster size by theoretical studies, which in turn affected the nucleation rate. It was demonstrated that shear can both encourage the growth of these clusters and break up a cluster into smaller ones. The relative contributions of these two effects varies depending on the shear.[240] This disruption reported might affect in crystallisation behaviour such that the role of the pores might differ when stirred, that there might be an increasing importance in the ability of the pore to accommodate a cluster of molecule.

5.3.4 Effects of Stirring Rate

The effects of stirring rate at different supersaturation was investigated. No apparent correlation was seen for unseeded experiments (Figure 35). At 17.5, the induction time increases with increased stirring speed, when the concentration was reduced to 15.5 mg/mL.

The contribution of different stirring rate to the coalescence or breakage of cluster is not clearly determined here and detailed studies was not discussed.

The effects of stirring in the presence of seeds were also investigated (Figure A2 in Appendix). Increased stirring generally decreased the induction time for crystallisation in the presence of NP seeds. This agrees with the hypothesis that the increased stirring increases the coalescence between small clusters into larger clusters sufficient for nucleation.[240] However, for the case of NT40, the increased stirring rate either had the opposite (at 17.5 mg/mL) or no effects (13.5 mg/mL) at all.
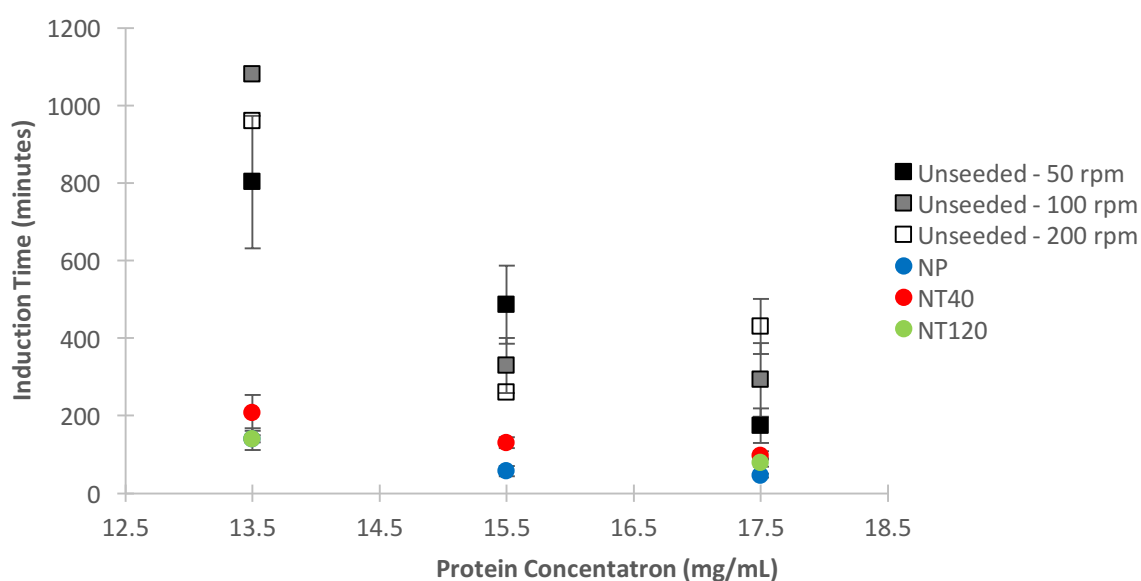
Figure 35 Comparing the effects of varying stirring speed and the presence of seed at protein concentrations investigated.

5.4   Summary

In our study, the induction time was monitored with UV-vis to obtain crystallisation profiles of the conditions studied, and from the curves the induction times was determined. The gradient of the desupersaturation was also compared qualitatively to add to the understanding of the process.

At the conditions studied for the crystallisation of HEWL, the nanotemplates NT40 was not the most effective nucleant in reducing induction time, despite having the highest surface area and the closest size match as the lysozyme molecule, as the use of NT40 consistently gave a higher induction time compared to the use of other seed types. This could be attributed to the different methods employed. Vapour diffusion was used in the experiments by Shah and co-workers, where it is possible to achieve crystallisation with the supersaturation gradually

increasing from an initially metastable condition, and the use of nanotemplates may increase the possibility of achieving nucleation at a supersaturation where nucleation rate is low. The studies on mesoporous silica occurred at conditions in which crystallisation can occur in the absence of heterogeneous nucleants. Also, the proposed mechanism of these nanotemplates was to increase the local concentration of correctly folded proteins near the surface.[28] As discussed previously (Chapter 2.4), formation of pre-nucleation cluster is frequently reported for the crystallisation of lysozyme, and that the effects of shear was reported to affect cluster formation, coalescence and breakage.

Currently, it remains unclear as to why non-porous silica can reduce the induction time by a larger extent compared to porous materials. Such surface properties have been suggested to drive supersaturation via capillary condensation; as well as the stabilisation of native protein conformation. A comparison was made for the induction time as a fraction of the unseeded induction time, t_{ind}(seeded)/t_{ind}(unseeded) in the presence of all seed types at protein concentrations 13.4 and 17.5 mg/mL at 1.25 mg/mL to quantify the efficiency of these seeds at different supersaturation. This suggests that there is a possibility that the effectiveness of the seed type is dependent on the protein supersaturation.

The effect of different heterogeneous seed concentration of mesoporous silica on the induction time was also determined. Across the seed concentration tested at 17.5 mg/mL of HEWL, two induction time values are obtained for all seed types. Above and below a critical seed concentration, the induction time remains unchanged. It appears at high seed concentration, the effects of surface properties (porous vs non-porous) has a bigger effect than changing the seed concentration. Below the critical seed concentrations, porous seeds resulted in induction time in the ranges at which unseeded HEWL nucleates, suggesting a reduced effect at the seed concentrations. A reduction in effects of NP seeds are also observed below the critical seed concentrations, but the induction time was still before the unseeded experiments.

6. Crystallisation of Catalase

This set of experiments involves the crystallisation from bovine liver catalase at two different scales. Conditions were chosen at first at 0.5 mL scale based on the crystals obtained and induction time. Effects on nanotemplates were investigated at 20 mL. Like the experiments reported in Chapter 5, the effects of nanotemplates was compared for a range of seed concentration. The seeds studied were NT120 and NP.

6.1   Materials
Catalase from Bovine Liver (BLC)

This protein catalyses the conversion of H_2O_2 to H_2 and O_2, protecting cells from the toxic effects of hydrogen peroxide.[241] Catalase consist of a tetrameric structure of four identical subunits, ~ 250 kDa. Crystallographic studies reported an approximate molecular diameter of 80 Å.[242] Catalase expressed in a widely different range of organisms are described to have great structural similarities, and the common folding structure is described as a 'catalase fold'.[243] Catalase is an acidic protein as it contains more acidic than basic residues.[244]

Catalase was first crystallised by Sumner (1937), who extracted the protein using a series of steps involving dioxane, followed by the precipitation with ammonium sulfate. Needle and plate like structures have been reported, while most crystals observed were plate-like.[135] Catalase was reported to have large solvent channels that contributes to a low crystal density.[242]

BLC has a pI of 5.4, and was reported to be a table at a range of pH values between pH 3-10.[245] Microcalorimetric studies reported a denaturation temperature of ~325 K, however, endothermic transition was observed from 310 K,[246] this defines the possible working range for the crystallisation experiment.

6.2   Method
6.2.1   Solution Preparation
Stock solutions of buffer was prepared by dissolving crystalline potassium phosphate dibasic (≥98.0 % purity, Sigma Aldrich) into analytical grade deionised water and the pH was adjusted to pH 7. Precipitant solution was prepared by dissolving 50 mM potassium phosphate and the precipitant components, PEG 4000 (Fischer) (6-10 % m/v) and 2-methyl-2,4-pentanediol (MPD) (≥99% purity, 820819, Merck) (10% v/v) into analytical grade deionised water and the pH was adjusted to pH 7 (Protein is negatively charged at the crystallisation conditions studied).

Lyophilised catalase (2000-5000 units, ≥60% purity, C9322, Sigma) was dissolved in buffer solution detailed above. In this process roughly 40% of the mass formed insoluble aggregates. The solution is therefore then centrifuged and filtered off using Millipore 0.22 μm vacuum filters. The filtered solution is buffer exchanged by ultrafiltration to minimise the additives present that was dissolved with the catalase (lyoprotectant, cryoprotectants, stabilisers) into the solution: The protein solution was transferred to centrifugal concentrators (Molecular weight cut off, MWCO – 5000 Da) and were centrifuged using a swing bucket centrifuge at 5°C and at a relative centrifugal force at 3000 g such that the volume of protein solution reduced from 20 mL to ~5 ±2 mL. To avoid the aggregation of protein due to concentration being too high, the dilute solutions were not concentrated by more than 4 times in each step. This concentrated solution was then diluted with new buffer solutions back to 20 mL. This process was repeated 3 times such that if any additives are present in solution, the quantity is reduced by ~64 times. These solutions were stored overnight in a fridge and are not kept for over 48 hours before they are used for crystallisation experiments. Prior to start of experiment, the solution was concentrated again to achieve an excess concentration than 2 times the starting concentration of each experiment. The centrifugation process cannot accurately control the final concentration; the desired concentration was reached by the addition of potassium phosphate to achieve 2 times the concentration of the starting concentration of each experiment as monitored using UV-vis with Nanodrop ND2000c.

Protein and precipitant solutions prepared from the above process were filtered with Millipore 0.22 μm syringe filters again immediately before start of crystallisation experiments described below.

6.2.2   Solubility Determination

Compared to lysozyme, there is an absence in data published on the solubility of most proteins, and catalase is an example. Catalase was chosen as the protein to be studied as it is feasible to use larger quantities of the protein to not only conduct scaled up experiments, but also to conduct solubility analysis using the protein. A 'working phase diagram' was published on similar crystallisation recipe to determine the operating protein concentration.[26] However, this corresponds to experiments conducted in vapour diffusion method, where both the protein and precipitant concentration increases over time, and the assessment of conditions were based on whether crystals or precipitations were present or absent.

Various methods of solubility determination were discussed and evaluated in Chapter 2. Various studies in the batch crystallisation of protein has determined the solubility with crystallisation in microscale, in which the residual concentration of the crystallised solution was determined by concentration measurements (Castro et al, 2016; Huettmann et al, 2015), this method was demonstrated to be feasible. This method was used as the technique for solubility determination. The possibility of experiments affected by the cessation of growth is acknowledged.

Prior to solubility determination, a series of protein solution of concentrations 5-35 mg/mL (0.5 mL) were set up for conditions of 4-10% PEG 4000 and were kept in HPLC vials at 20 °C. The ranges of conditions in which crystals were observed after three days were chosen as starting condition for solubility determination. The use of a higher precipitant concentration was preferred for the design of this experiment as this allows for crystallisation to occur at a lower protein concentration. The solubilities of catalase at 8% and 10% PEG 4000 were studied in detail. Catalase solutions of concentration 15 mg/mL and 20 mg/mL were set up in the two precipitant concentration and were stirred at 100 rpm at 20 °C. Solution concentration was monitored for one week to ensure an equilibrium is being reached.

6.2.3   Batch Crystallisation
0.5 mL experiments

Protein solutions prepared were pre-mixed with precipitant solution of 16% PEG 4000 and 10 % MPD at a ratio of 50:50, making up to 0.5 mL of volume within HPLC vials. Solutions were then immediately incubated at 20 °C and stirred at the specified stirring rate with Crystal 16 (Mettler Toledo). Magnetic stirrers of dimensions 7 x 2 mm were used in stirred experiments and are not placed within vials containing unstirred experiments. The crystallisation profile was monitored by the laser transmission through solution based on the built-in optics within equipment. Offline, real-time concentration monitoring using UV-vis was not possible with the equipment configuration. Initial concentration was therefore assumed to be half of the concentration of protein stock solution. Due to the configuration of the set-up, solutions are prepared in sets of four, an error of ±5 minutes is therefore associated with the precipitant addition and pre-mixing for each sample, and varies from sample to sample, which is not accounted for in the data reported in 6.3.3.

Seeded Experiments: Mass of seed corresponding to 1 mg/mL and 2 mg/mL were weighed directly into HPLC vials, to which the crystallisation solutions described above are added. The experimentation was carried out in a similar manner as the unstirred experiments described.

20 mL stirred (EasyMax)

Protein solution prepared was diluted by precipitant solution of 16% PEG 4000 and 10 % MPD, making up to 20 mL of volume within an EasyMax vessel. Solutions were incubated at 20 °C and stirred at the specified stirring rate within EasyMax (Mettler Toledo). The crystallisation process was monitored with both turbidity and UV-vis.

Seeded Experiments: Mass of seed corresponding to the seed concentrated investigated were weighed directly into EasyMax glass vessel, to which the crystallisation solutions are added in the manner described above.

Turbidity Calibration: The signal response of each turbidity probe in a catalase solution was adjusted to <10 %. The turbidity probes are then placed in a vial isolated from all incoming light source and was adjusted to >90%.

Due to the presence of different amount of seed used, results presented in Chapter 6.3 were baseline corrected to align with each other. This was achieved by subtracting each data point by the signal response at the first data.

UV-vis spectroscopy: The molar extinction coefficient coefficient was calculated using the ProtParam tool by ExPASy Bioinformatics Resource Portal based on the sequence of a monomer of BLC, which was obtained from Uniprot (Code: P00432). The value was found to be 64540 M^{-1}cm for the monomeric unit of 59.9 kDa. A$_{260}$/A$_{280}$ was 0.62 for BLC.

6.3   Results and Discussion

6.3.1   Unstirred (Preliminary) experiment

Crystals were reproducibly observed at 20 and 15 mg/mL at 8 % PEG 4000 (Figure 36A and B), and also, 25 mg/mL at 6% PEG 4000. Plate-like crystals were obtained for both protein concentrations at 8 %. The angles and edges of the crystals were quite well defined for 20 mg/mL, crystals observed in this condition are of a width of ~200 μm and its length at around 300 μm. Whereas at 15 mg/mL crystals obtained were elliptical and with a larger aspect ratio, with the breath similar to the width of the 20 mg/mL crystals and length reaching almost 1 mm. The effects of seeds in unstirred conditions was not studied due to the inability to keep the silica suspended.
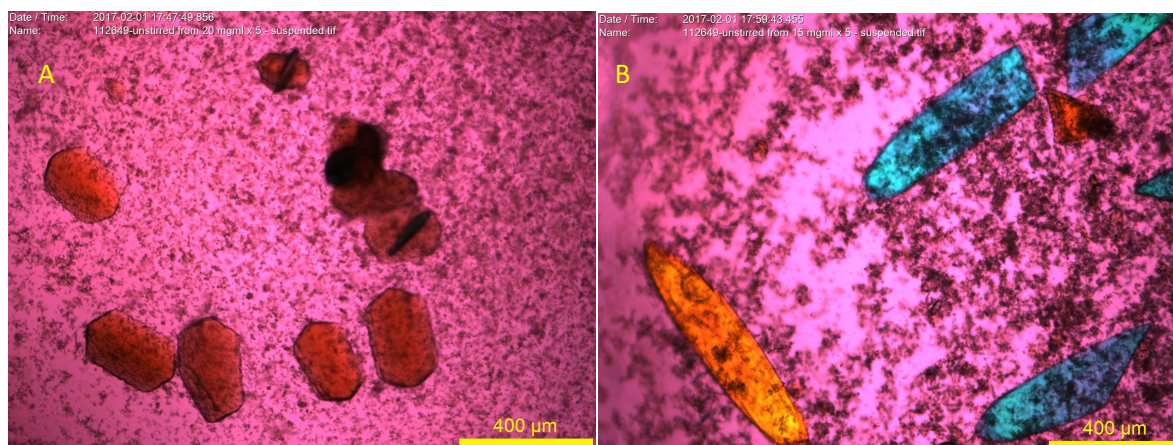
Figure 36 Crystals of catalase at 8 % PEG 4000 content at a) 20 mg/mL and b) 15 mg/mL

6.3.2 Solubility Determination

Crystallisation experiments conducted at 8% and 10% PEG 4000 concentration both equilibrated at a mean value around 10.523 mg/mL and 8.53 mg/mL respectively, and these value was taken as the solubility of the protein at these precipitant concentrations (Figure 37 and 38). The solubility for each condition was assumed at this value as further decrease in protein concentration from 10 mg/mL was minimal after < 1 week for all conditions tested. The proximity of these two values suggested that the solubility curve is fairly flat at the precipitant concentrations investigated. Based on the negative exponential relationship between the solubility of protein and the precipitant concentration as represented in Equation 1, precipitant concentration is quite high with respect to the phase diagram. It should however be acknowledged that the solubility determined with this method, an error due to cessation of growth is expected. This phenomenon is common in protein crystallisation in which its details is discussed in Chapter 2, and can occur at both nucleation and metastable zone. If results are expected by this phenomenon, the true solubility values of catalase in these conditions are expected to be lower.

The only other known report on the phase behaviour of BLC in similar systems was by Chayen and co-workers (2001), in which the supersolubility curve was obtained by microbatch methods. In their reported studies a similar range of PEG concentrations (6-14 % m/v) was also reported to yield crystals, but PEG 6000 was used instead of PEG 4000 and the reported pH was higher by 0.5. MPD (5% v/v) was also used as the precipitant but in different buffer conditions. The supersolubility limit was reported in literature was ~10.5 mg/mL for 8 % m/v PEG, and can expect to be lower in the conditions described in this chapter as the pH reported here is closer to the isoelectric point. However, this would suggest a substantial error in the

solubility here. Unfortunately, there is no information in literature regarding the solubility of BLC crystallised in similar systems.

The variation in concentration in $t < 24$ hours for the discussed experiments was summarised in Figure 39. At a PEG 4000 concentration of 8 %, an induction time of 2.3 hours was obtained for 18 mg/mL and 3.3 hours for 14 mg/mL. A time delay from t=0 in concentration decrease was not observed for 10 % PEG 4000 mg/mL. While the solubility was similar for the two precipitant concentrations, the induction time varied between these two conditions. While this does not correlate with CNT, most evidence against CNT reported in literature that was discussed in Chapter 2 have discussed the deviation of nucleation rate from CNT predictions. The induction time obtained for these conditions also helped identify the conditions suitable to conduct seeding experiments in. Conditions at 10 % PEG 4000 was not considered for further experiments due to the instantaneous nucleation observed. Experiments described in the rest of this chapter will therefore be at 8% PEG 4000.
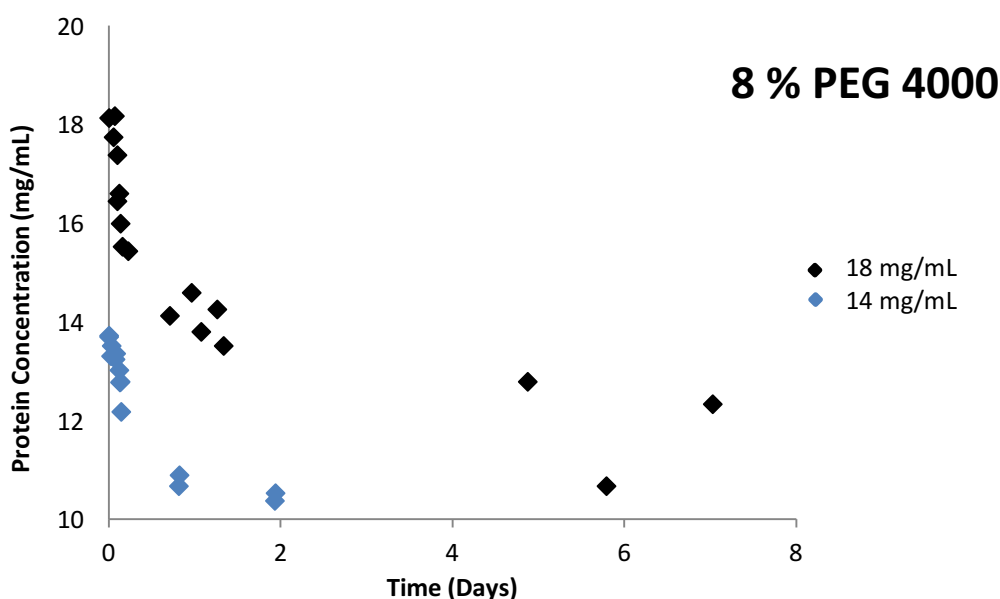
Figure 37 Changes in concentration throughout one week for crystallisation of different starting concentrations as indicated as t=0 with 8% PEG 4000 precipitant content
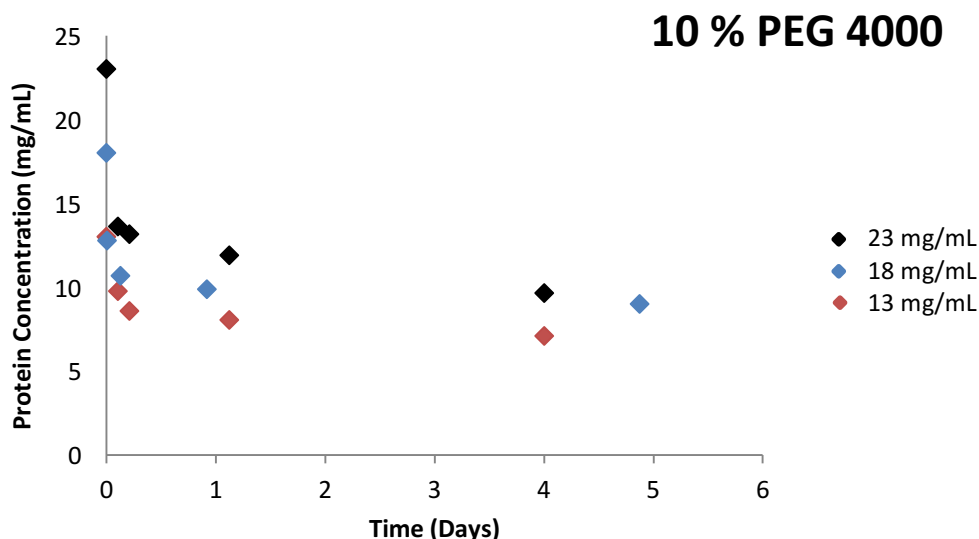
10 % PEG 4000

Figure 38 Changes in concentration throughout one week for crystallisation of different starting concentrations as indicated as t=0 with 10 % PEG 4000 precipitant content
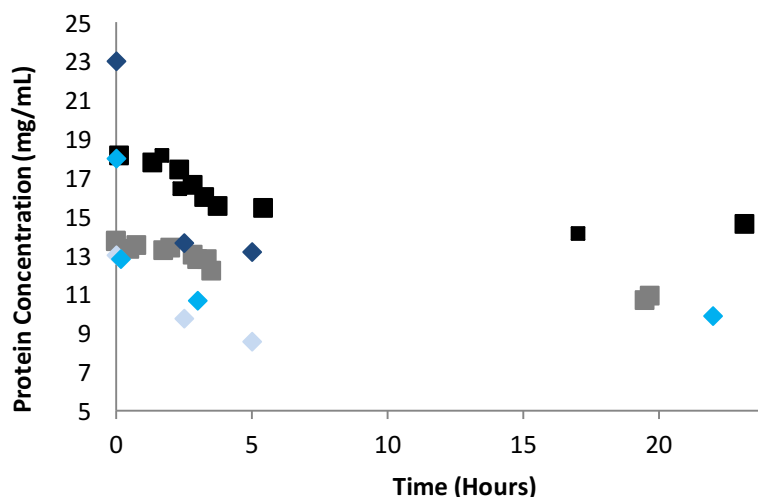
Figure 39 Concentration changes of catalase at 8% (■) and 10% (◆) PEG 4000. The initial concentration can be deduced at time = 0 from graph, also represented with varying colour of data point for ease of interpretation.

6.3.3 Stirred Crystallisation Experiment (0.5 mL) in Crystal 16

The induction time presented in this section is characterised by a rapid decrease in transmission signal and are represented in Figure 40 and 41.

At 20 mg/mL, the induction time was < 5 minutes when stirred in a working volume of 0.5 mL (Figure 40). At 200 rpm, the transmission signal drops from >96% as expected, suggesting initial solution was sufficiently clear. However, at 100 rpm, the initial transmission signal was below 50% and remained so until the offset in transmission signal. This suggests solids were already present at the start of experiment. The reduced transmission response was

unlikely the result of the presence of dust particles: Figure 42 shows the transmission signals at different seed concentrations in a blank solution (catalase concentration = 0 mg/mL) across 4 hours. Even at a seed concentration of 3 mg/mL, which was beyond the highest seed concentration investigated in experiments presented in this chapter, the transmission signal remains higher than the 20 mg/mL solution across 250 minutes at the same stirring rate. The presence of a constant transmission value prior to the sharp decrease in the reading suggested that nucleation was not instantaneous. Detailed investigation on the effects of stirring rate at this condition was difficult as the induction time was not sufficiently large for comparison.

When the protein concentration was reduced to 15 mg/mL (Figure 41), the induction time was clearly delayed at 100 rpm compared to the 20 mg/mL experiments, nucleation occurred only after 50-90 minutes. The variation in induction time was much greater than the induction time of the 20 mg/mL experiments itself. When the stirring rate was increased to 200 rpm, induction time was reduced to ~10 minutes. At 400 rpm, the transmission signal began below 100% and decreased gradually until the curve follows similar curvature as the 200 rpm experiments. As discussed, an error of ± 5 minutes is associated with the preparation of this method, which contributes significantly to the induction times under the conditions that results in an induction time of <20 minutes, these conditions are therefore not investigated further at this scale.

The transmission signals through solution and turbidity depend on both the cross sectional area of a particle and the number of particles in solution the beam source travels, as discussed. Nucleation is assumed to dominate at t_{ind}, the gradient of decrease in transmission at t_{ind} was therefore compared. At 20 mg/mL, this gradient was similar between 100 and 200 rpm, while a clear difference was observed between the same stirring rate for 15 mg/mL, also suggesting the influence of stirring on nucleation rate.
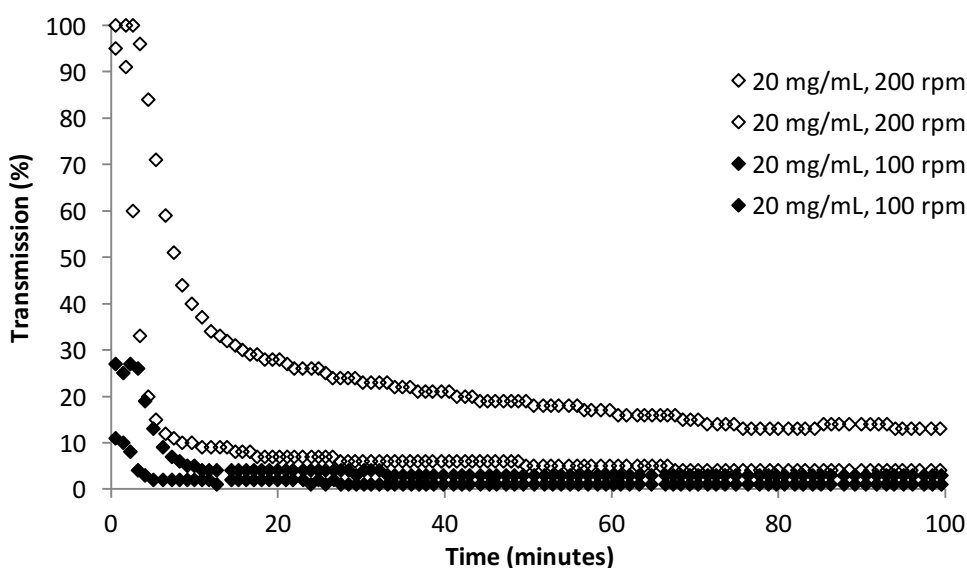
Figure 40 Crystallisation profiles of varying stirring rates for unseeded crystallisation experiments at 20 mg/mL obtained from light transmission measurement through reaction volume.
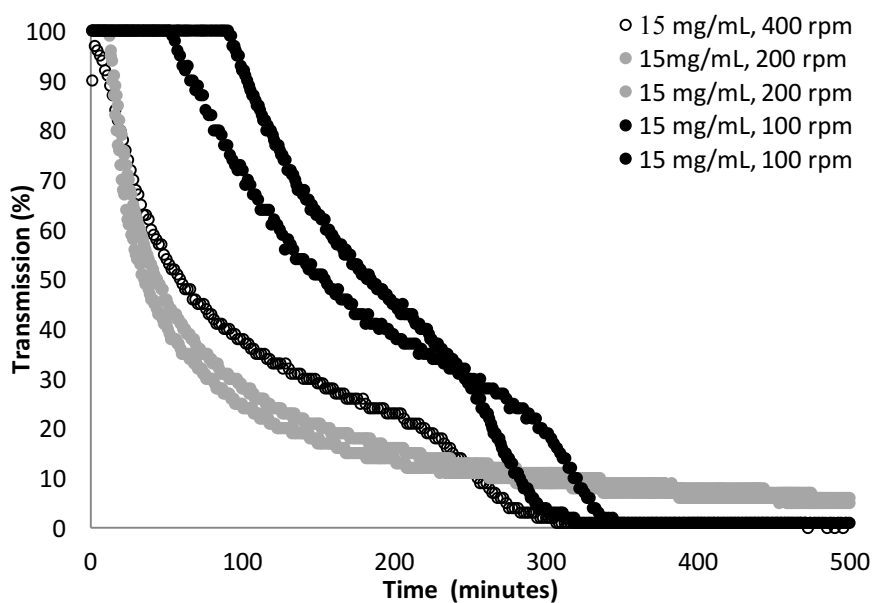
Figure 41 Crystallisation profiles of varying stirring rates for unseeded crystallisation experiments at 15 mg/mL obtained form light transmission measurement through reaction volume

Based on results from Figure 41, catalase concentration of 15 mg/mL at a stirring rate of 100 rpm was used for all stirred crystallisation at 0.5 mL scale. Seeded experiments were attempted (Figure 43). However, in the presence of seed, the induction time was not evident from the changes in transmission signals. The initial transmission signal at seed concentration 2 mg/mL was also significantly lower than that of a blank solution containing the same amount of silica while the seed concentration at 1 mg/mL was only 10 % lower than its baseline value

(Figure 42). The gradients for the immediate decrease in transmission for these seeded experiments were both similar with the unseeded experiments.
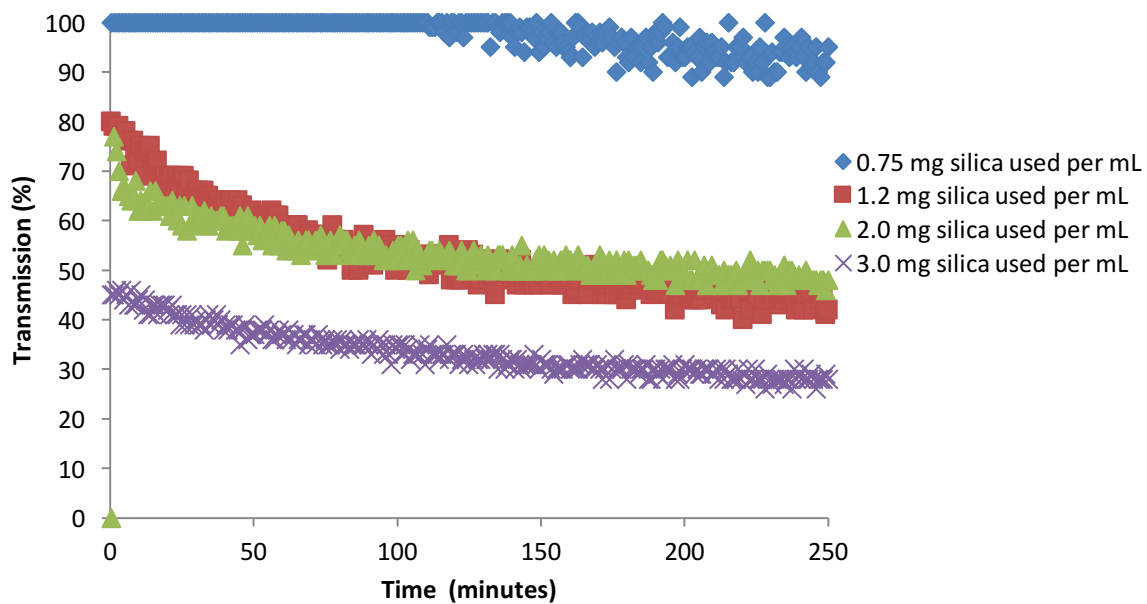
Figure 42 Baseline measurements of transmission signals resulted from different seed concentrations at 100 rpm.
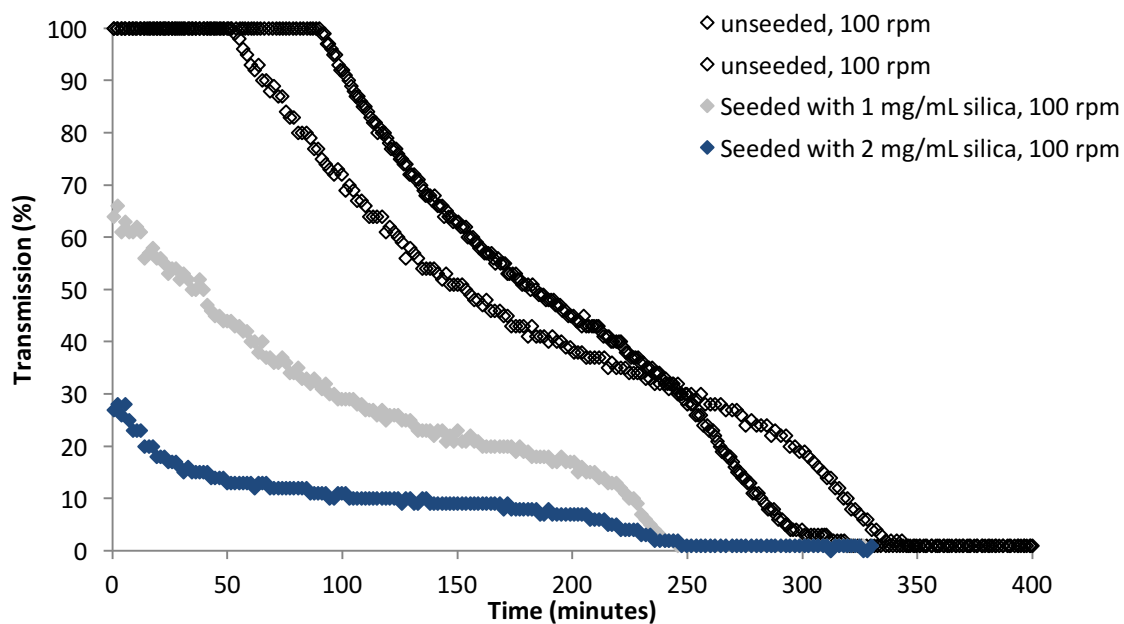
Figure 43 Catalase crystallisation at 15 mg/mL obtained at different seed concentrations.

6.3.4 EasyMax Experiments (20 mL)

Experiments described in this section were conducted at 17.5 mg/mL and 13.4 mg/mL and 100 rpm. A range of seed concentration (0.1 mg/mL to 2.5 mg/mL) was investigated at these concentrations.

Crystals were formed at both protein concentrations when operated at 100 rpm in the scale described. Aliquots of crystallisation solutions were withdrawn for imaging, no qualitative differences in appearance of crystals is observed at different supersaturation. Crystals captured from imaging are all 10-20 μm in the longest dimension and varying dimensions in width. Plate-like crystals are obtained, with shapes less well defined than those obtained at unstirred conditions. The shape of crystals obtained also resembled the crystal structures that was reported in Tsekova et al (2012)[204] although the precipitant conditions are slightly different: Within the literature discussed, PEG 6000 was used instead of PEG 4000, crystals were grown at pH 8.4 was used instead of pH 7. (Figure 44 A and B respectively)
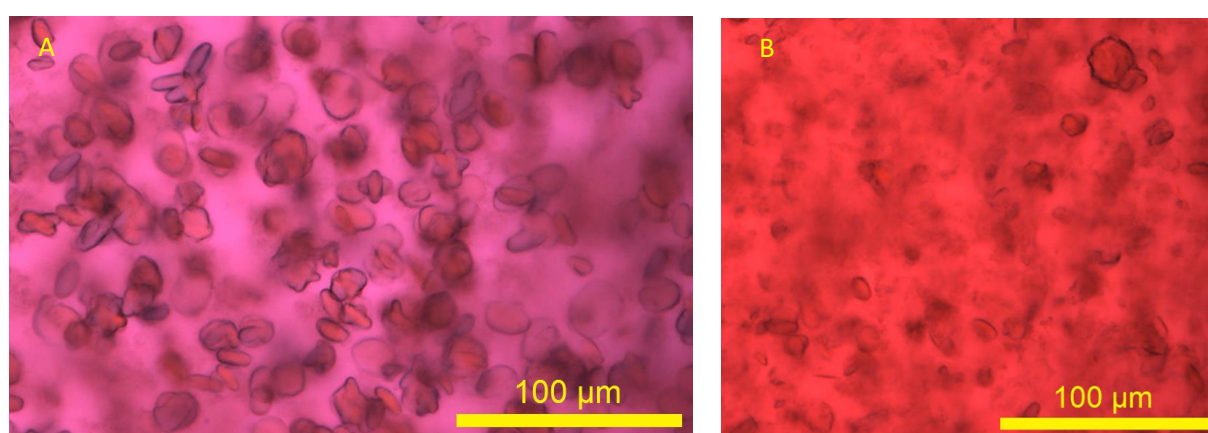
Figure 44 Catalase crystals crystallised in a stirred (100 rpm) vessel at a working volume of 20 mL and a concentration of A) 17.5 mg/mL and B) 13.4 mg/mL.

Turbidity/ Concentration Correlation:

The simultaneous measurement of protein solution concentration and the baseline corrected turbidity values were represented in Figure A3-5 (Appendix). The three separate experiments showed some correlation between the two techniques – both UV-vis and turbidity was suitable in differentiating the induction time between different supersaturations. However, the sensitivity of the turbidity probe with respect to UV-vis concentration measurements vary with varying conditions. The turbidity probe was able to pick up on the induction time prior to UV-vis spectroscopy for 17.5 mg/mL protein concentration seeded with 1.0 mg/mL of NT120, while it was less sensitive at high seed concentration (2.5 mg/mL) as well as at 13.4 mg/mL unseeded catalase solution. It can be seen that the size of the silica particles (Figure 21) was significantly larger than the BLC crystals obtained at this scale (Figure 44), this might contribute significantly towards the measurement towards the induction time when the nuclei is formed according to Equation 27. The following experiments therefore focuses on using UV-vis spectroscopy to obtain crystallisation profiles.

Crystallisation of BLC under effects of seeds

Figure 45 and Figure 46 shows the induction time at various seed concentrations of NP seeds and NT120 respectively at supersaturation of 17.5 ± 0.5 mg/mL. Unseeded experiments are taken from the solubility studies reported above.

The effect of NT120 was not evident from Figure 46, induction time values obtained for experiment across all seed concentration (0-1.5 mg/mL) were within error of each other. In the presence of NP seeds (Figure 45), the gradient of concentration decrease was sharp compared to other conditions since t=0 at high seed concentration (0.6 – 2.5 mg/mL) and remained so thruoghout the first 5 hours of experiment, suggesting that nucleation was instantaneous. At a reduced seed concentration (0.1 – 0.25 mg/mL), the point at which a further decrease in protein concentration gradient was at around 1 hour, which can be taken as the induction time. The rate of desupersaturation was rather low, a drop in protein concentration was ~ 3 mg/mL in 3 hours for most conditions, and the turbidity increase in these experiments were ~6 %.

Similar to the data obtained for HEWL, the variation in amount of seeds in the range of seed concentration used does not have great effect on the induction time. In the case of BLC at 17.5 mg/mL, these ranges like below 1.0 mg/mL of silica seeds. At low seed concentrations, experiments in the presence of NT120 demonstrated no clear effects in shortening the induction time of the 17.5 mg/mL experiments in the seed concentration investiated. The crystallisation curve followed the same shape as that of the unseeded experiment. For this reason even lower seed concentrations levels were not investigated. This is similar to the results obtained from seed concentration experiments for crystallisation at HEWL 17.5 mg/mL seeded with NT40. While the NP seeds are weakly effective at high seed concentrations. In both cases, the shape of the desupersaturaiton of the crystallisation profile was similar to that of unseeded experiments. As data represented the change in concentration, it is difficult to extract information on the relative contribution of nucleation and growth that resulted in the shape of the graph.

Like in the HEWL experiments, the use of different seed type however has influenced the induction time. To highlight the instantaneous nucleation of crystallisation in the presence of NP seeds, the crystallisation profile in Figure 45 and 46 were modified, and expressed in the form of Figure 47, where the concentration at time t, c_t, was expressed as a difference to the starting conditions. Unseeded experiments are represented by the unfilled squares, where the

blue circles and green rhombuses represent crystallisation experiments seeded with NP and NT120 seeds respectively. While the unseeded experiments and NT120 experiments coincide in the same region, the concentration decrease of the NP seeded experiments was consistently lower than the other two conditions.
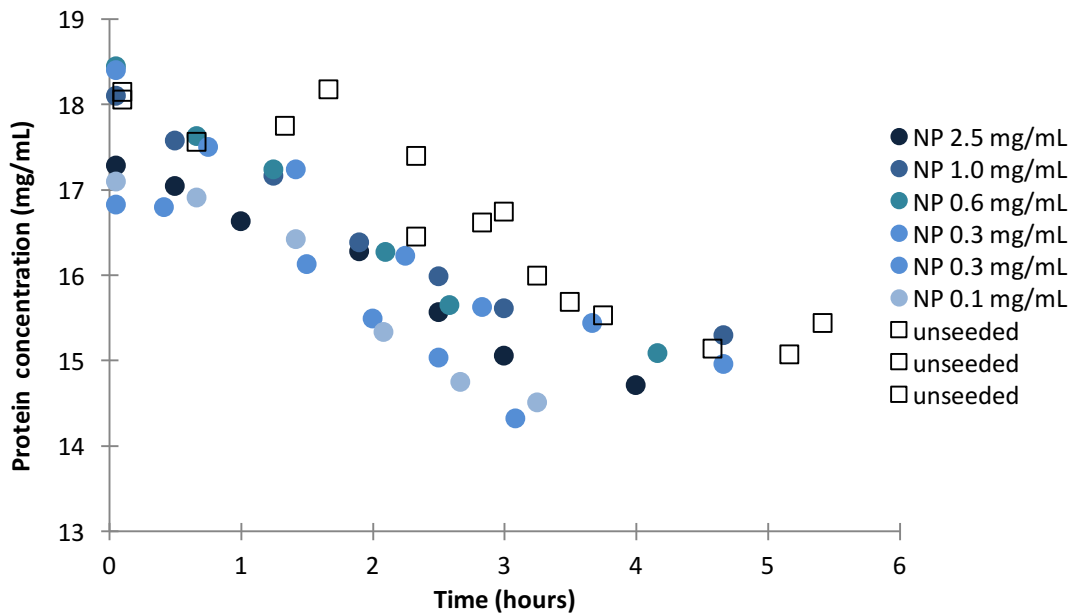
Figure 45 Crystallisation experiments at catalase concentration of 17.5±0.5 mL in the presence of various amounts of NP seeds.
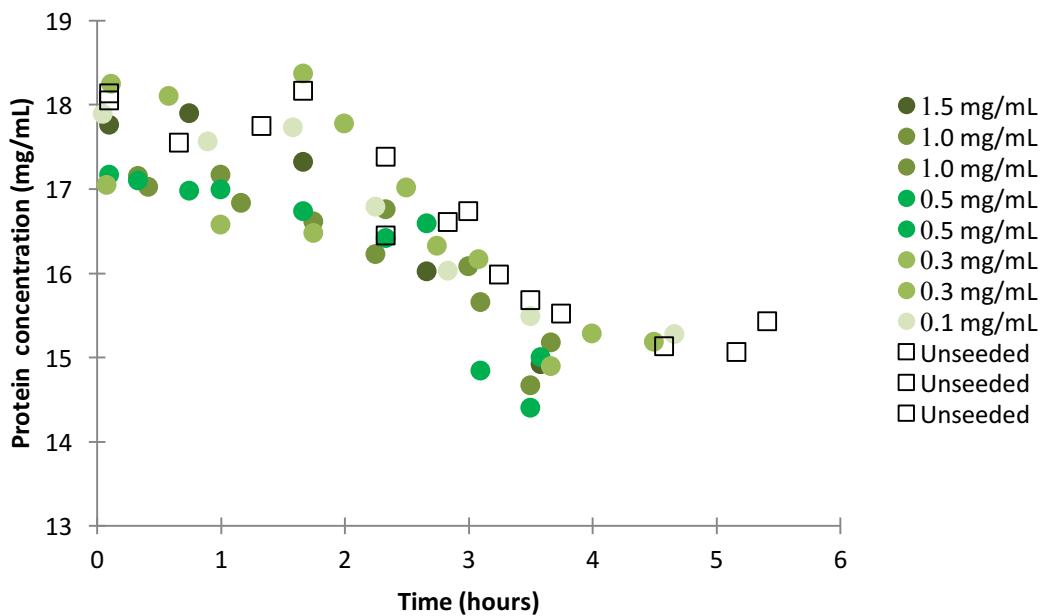
Figure 46 Crystallisation experiments at catalase concentration of 17.5±0.5 mg/mL in the presence of various amounts of NT120 seeds
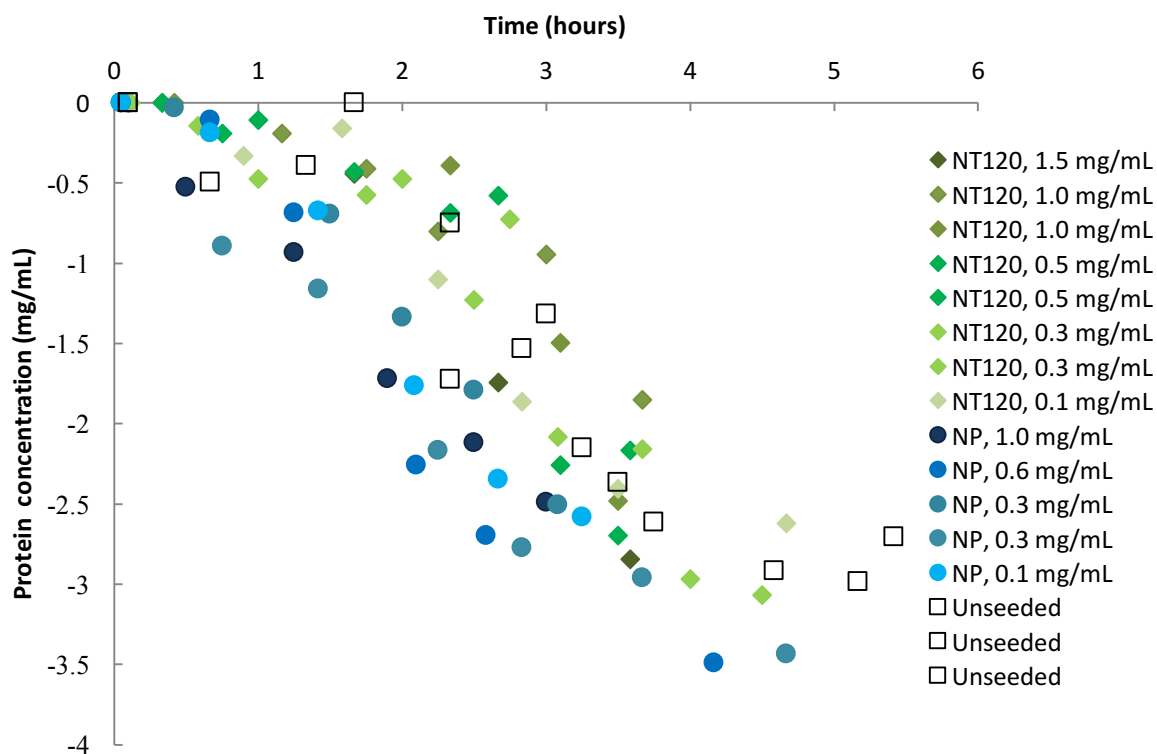
Figure 47 Comparison of crystallisation profile at BLC concentration of 17.5 mg/mL in the presence of NT120 seeds (green) and NP seeds (blue). Concentrations were normalised from Figure 45 and 46 to the difference from starting concentration (c_t-c_0) for clarity.

As the effects of these seeds on the induction time was not very clearly demonstrated in experiments discussed above, these experiments were therefore repeated at a lower protein concentration level (13.4 ± 0.5) mg/mL. Based on data obtained from crystallisation experiment with HEWL, and from [105], it was expected that at a reduced supersaturation (within the supersaturation range >1), the effects of heterogeneous surfaces on nucleation data would be clearer.

The crystallisation profiles at a reduced concentration (13.4 mg/mL) in the presence of NT120 (seed concentration at 0.25 to 1.0 mg/mL) was summarised in Figure 48. To consider each experiment separately, the induction time at a seed concentration of 0 – 1.0 mg/mL can be considered to be ~ 3 hours, this range of seed concentration had not reduced the induction time compared to the unseeded experiments, the crystallisation profiles also follow the same shape as those of unseeded experiment before and after the induction time. As expected, the decrease in protein concentration occurred more slowly than at 17.5 mg/mL. In these experiments, the protein concentration decreased by 2 mg/mL 3 hours after the induction time. However, at seed concentration level 2.0 mg/mL, the induction time has been reduced to ~ 2 hours. More interestingly, the use of such high seed concentration seemed to result in a different

shape of crystallisation profile, characterised by a steeper decrease in concentration after induction time (~ 1 mg/mL per hour). The induction time and crystallisation curve in these seed concentration was similar to the experimental conditions at 17.5 mg/mL.

The effects of NP seeds on the induction time was demonstrated less clearly at reduced concentration of 13.4 mg/mL than at 17.5 mg/mL experiments (Figure 49). Like for the unseeded experiments, the induction time remained at ~3 hours. For seed concentrations below 1.0 mg/mL, the shape of crystallisation curve did not change either.
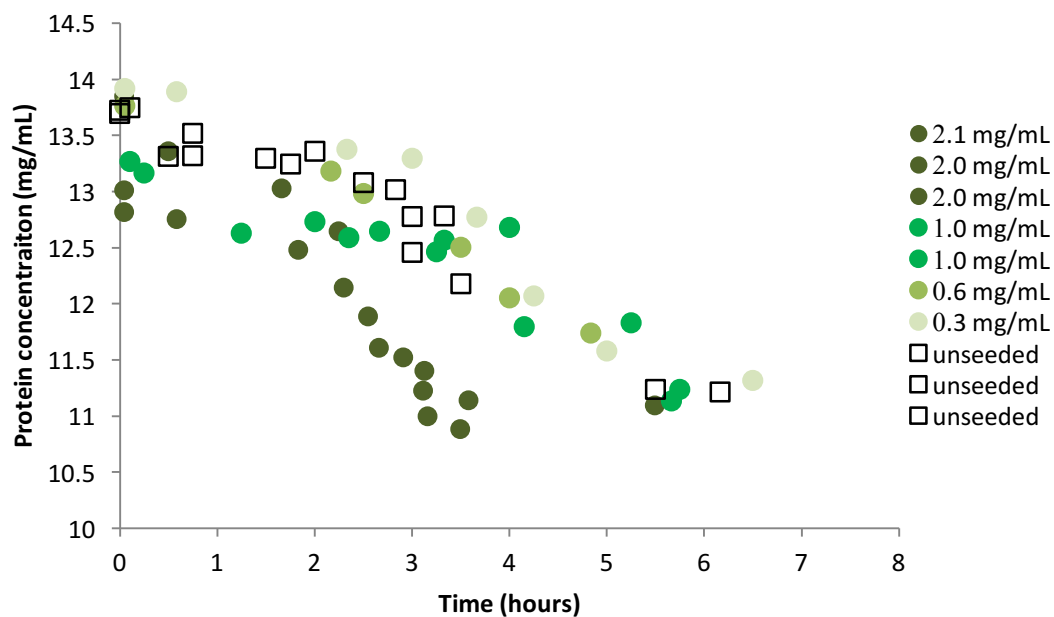
Figure 48 Crystallisation of catalase (13.4±0.5 mg/mL) in the presence of various amounts of NT120 seeds.
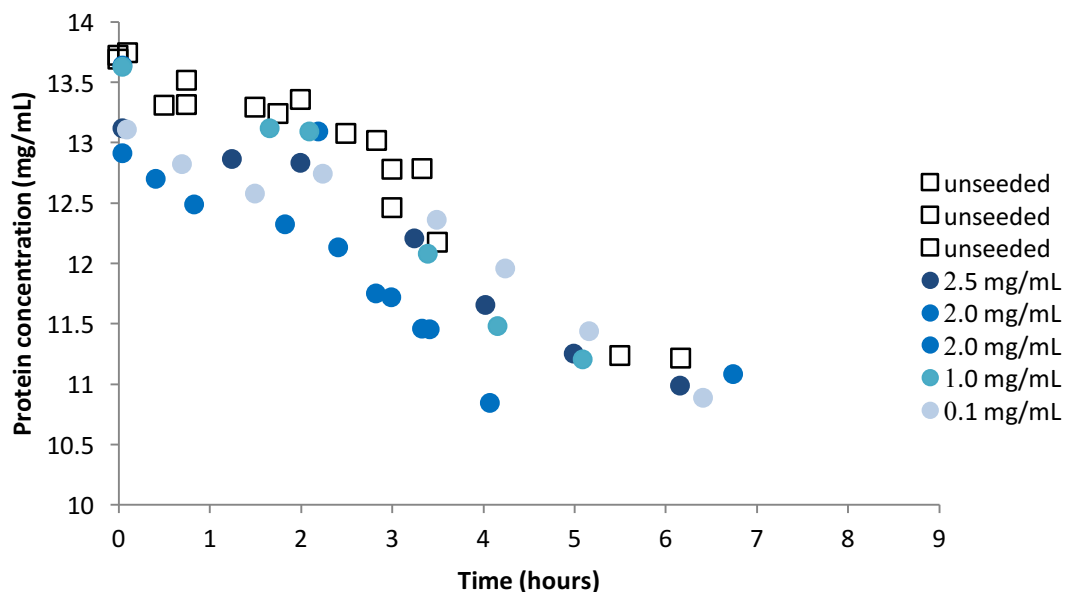
Figure 49 Crystallisation of catalase (13.4±0.5 mg/mL) in the presence of various amounts of NP seeds.

As a NT120 concentration of 2.0 mg/mL demonstrated some effects in reducing the induction time, the crystallisation was therefore plotted as c_t-c_0 over time t (Figure 50). This was compared against the unseeded experiments as well as the NP seeded experiments of similar seed concentration or higher. The concentration decrease from c_0 in the presence of 2.0 mg/mL NT120 seeds were consistently larger than the other seed conditions.
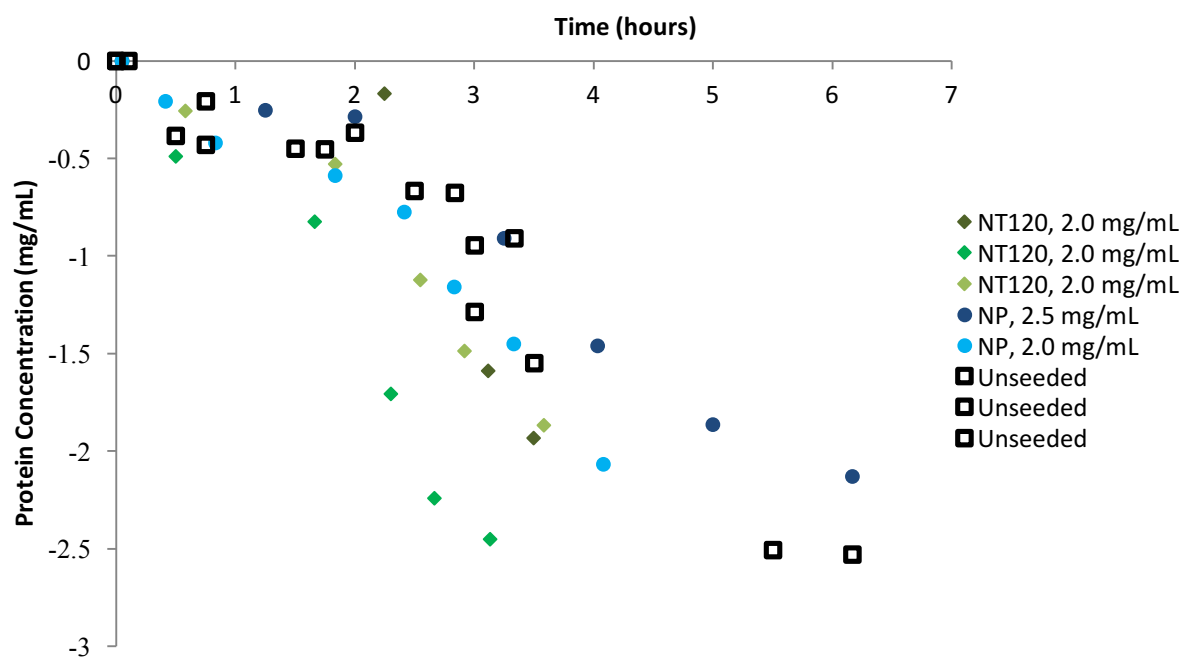
Figure 50 Comparison of crystallisation profile at BLC concentration of 13.4 mg/mL in the presence of NT120 seeds (green) and NP seeds (blue). Concentrations were normalised from Figure 48 and 49 to the difference from starting concentration (c_t-c_0) for clarity.

6.4   Comparison with HEWL Experiments

Overall, the influence of the seed used and the seed concentration was less apparent in the conditions explored for BLC, compared to HEWL, even at a reduced supersaturation. Most of the seeded crystallisation at seed concentrations of 1.0 mg/mL or below display the same crystallisation behaviour as unseeded experiments. It is known that the contributions of heterogeneous nucleation are larger at lower supersaturation[105]. However, at the BLC concentrations studied, the supersaturation ratio was <2. The relationship between seed concentration and induction time observed for HEWL was also observed in the case of BLC crystallisation. The slight reduction in induction time was observed for 2.0 mg/mL NT120 at 13.4 mg/mL and also giving a shape at the desupersaturation region of the crystallisation curve. The induction time remained unchanged for most seeded experiments at the protein concentrations investigated from 0 mg/mL to 1 mg/mL seed concentration. For HEWL, a range

of seed concentration were observed to be effective in reduction of induction time, and the induction time values were constant at those ranges, while at a reduced seed concentration, the crystallisation curve coincided with those from unseeded experiments. The region in which seed was demonstrated to be effective was not extensively studied for BLC.

In comparison, in experiments reported in Chapter 5, the rate of desupersaturation was significantly lower at all conditions. A desaturation of 1.2 mg/mL per hour after induction time was observed for crystallisation experiments at 17.5 mg/mL, and even less so for 13.4 mg/mL (0.8 mg/mL per hour). While as expected, the size of the BLC crystals grown in these conditions are significantly smaller than those in quiescent mode, the variation in crystal size as seen from microscopy (Figure 44) was not very large. It is possible that after the induction time, limited nucleation events took place as the crystals nucleated at t_{ind} grew, as the supersaturation ratio determined from solubility measurement was quite low (Solubility was 10.5 mg/mL), i.e. the protein concentration in solution was close to equilibrium. The increase in turbidity was also quite slow after the induction time determined using the technique (Figure A3-5, Appendix). However, turbidity measurements is affected by both the size and concentration of solid particles formed, which cannot be used effectively in the set-up used to decouple whether there is an increase in particles observed or the increase in particle size. To confirm the desupersaturation is mainly attributed to the growth rate, real-time particle size monitoring technique would be required to decouple nucleation and growth processes occuring within.

Considering only the conditions in which seeds have an effect, it can be suggested the relative performance of different seed types vary depend on protein concentration. At high concentration, NP seeds had greater effect, while at lower protein concentration, the pores present in NT120 had a greater effect. This was not observed in the conditions that was used in the crystallisation of HEWL in Chapter 5. While an appreciable difference between the induction time for NP and NT40 seeds were observed at 17.5 mg/mL HEWL concentrations, this difference was less clear when the supersaturation was reduced, in which we suggested a protein concentration dependency, which is echoed in results presented in this chapter.

The crystallisation profile at seed concentration (2.0 mg/mL) of NT120 that showed some effects of nucleation in Figure 48 was compared to the same concentration of NP seeds.

6.5   Summary

Crystallisation of BLC was scaled up from 0.5 mL to 20 mL. The solubility of BLC in 8 % precipitant (PEG 4000) concentration was first determined to be 10.5 mg/mL, and the crystallisation condition was chosen based on the induction time.

Stirred experiments were performed. At 0.5 mL scale, all crystallisation experiment resulted in a short induction time, at 15 mg/mL, nucleation occurred within two hours, while at 20 mg/mL, the induction time was down to < 5 minutes. Immediate crystal formation was observed in the presence of NT120 seeds.

Minimal effects of both silica seeds were observed at 17.5 mg/mL, except at 2.0 mg/mL NT120 seed concentration, in which a steeper desupersaturation gradient was also seen, which was not observed at any other conditions tested in the crystallisation of BLC in this chapter. Like with HEWL crystallisation, the effects of seed concentration were not observed for a given range (0-1.0 mg/mL for BLC), except for the seed concentrations described for NT120, supporting the seed concentration relationship with the resultant induction time.

A protein supersaturation dependency was proposed for the effects of different seed types on the reduced induction time for HEWL experiments. Results shown in this chapter shows some support with this proposal.

7. Conclusion and Future Outlook

7.1 Conclusion

The aim of this project was to add to the understanding in the role of mesopores on heterogeneous surfaces on the crystallisation of proteins. Prior to the studies reported within, it was expected that pores of specifically designed diameters would be most efficient in promoting protein crystallisation.

The effects of these seeds were studied on two proteins of different hydrodynamic radii (HEWL: 4.5 nm, BLC: 10 nm). In order to make a comparison between different heterogeneous seed types, it is necessary to firstly understand the effects of different surface properties in driving the nucleation process. Non-functionalised silica based seeds were used for the studies, and the pore size of the seeds used were selected based on the relationship that Shah et al (2011) developed, and on the protein investigated, and were compared against non-porous silica. These seeds were purchased and their surface properties (specific pore volume, BJH pore diameter and specific surface area) were characterised prior to crystallisation experiments. The hydrodynamic diameter of the protein was also characterised for comparison to be made.

While the thermodynamic properties of HEWL was frequently reported in literature, stirred batch crystallisation of HEWL was applied directly for the studies with nanotemplates and other seeds. The amount of seed used was controlled. The solubility was determined to be ~ 10 mg/mL for BLC prior to crystallisation experiment, and the crystallisation processes were investigated at 0.5 mL and at 20 mL. Certain observations for the crystallisation behaviours of the two proteins are summarised below.

Relationship between seed type and protein   With HEWL, it was found that in all conditions employed in the studies described, the induction time was reduced by the largest extent in the presence of NP seeds. The effects of the NP seeds are much clearer in the case of HEWL. In the HEWL concentration studied, the increase of induction time upon reducing the protein concentration in the presence of seed varies depending on both seed types used and the protein supersaturation. The effects of heterogeneous seeds are not very apparent in the crystallisation of BLC experiments. While NP seeds resulted in a slight reduction of induction time at BLC concentration of 17.5 mg/mL, the effect of NT120 is more apparent at a reduced BLC concentration at 13.4 mg/mL. Results from HEWL demonstrated that different seed types have different effects on reducing the induction time for a given protein concentration. It has

also demonstrated that the relative effects of pores also depend on the supersaturation of the protein as well.

The Effects of Seed Concentrations Studies on the induction time at various seed concentrations of different seeds allows for the quantification of the effects of varying surface properties (surfaces area vs pore volume), and also to identify suitable seed concentration levels for the use of heterogeneous seeds in protein crystallisation, as this has not been extensively reported. While it has been reported that the induction time decreases linearly with increased loading of mesoporous seeds in batch crystallisation of inorganic molecules, this is not the case reported here for the batch crystallisation of proteins, when the diameter of the pores used here was sufficient to accommodate one protein, whereas literature that was discussed in Chapter 2.4 described the optimal pore sizes for high nucleation rates in terms of number of lattice sites or size of cluster of small molecules. With the exceptions for certain conditions of BLC crystallisation (which possibly requires screening of a larger seed concentration range), experiments reported here demonstrate a critical seed concentration. Below this value, a larger induction time is achieved. With the exception of NP seeds in the crystallisation of HEWL, in which a reduction in induction time was observed at all seed concentration, all seeds used in studies with both HEWL and BLC result in an induction time that is similar to the induction time of unseeded condition and was constant below the critical concentration. A shorter induction time was obtained for induction time higher than the critical seed concentration.

7.2 Future Work

In order for the nanotemplates to be used as a type of heterogeneous seed for the industrial crystallisation of proteins, crystal properties such as crystal size distribution and purity would ultimately have to be demonstrated. However, more understanding is required regarding both the batch crystallisation process and the mechanisms of protein nucleation within pores. Summarised below are the different aspects of related work that is of interest to the author of the thesis:

Detailed investigation of phase behaviour of proteins

While the knowledge of protein phase behaviour is not of great novel potentials, in order to investigate in the supersaturation dependency (discussed next), detailed understanding of the phase behaviour of each protein studied is required. Experiments presented in Chapter 5 and 6 were conducted with the knowledge of solubility. Extensive studies to obtain phase diagrams of different proteins (by finding the supersolubility limits) would assist with

identifying the suitable conditions to demonstrated the supersaturation dependency. In the experiments conducted, the desupersaturation was of interest to the studies in understanding the role of surface mesopores on the batch crystallisation of proteins. However, decoupling this behaviour into nucleation and growth was not possible. The knowledge in supersolubility limit and metastable zone widths would also contribute to this.

Investigation of supersaturation dependency

It has been demonstrated that in stirred batch crystallisation of proteins, the optimal pore size used to reduce induction time of crystallisation does not only depend on the size of proteins, but also the protein concentration. This does not necessarily contradict previous experimental results on the use of nanotemplates on protein crystallisation, as most experimental data reported on the role of pores are demonstrated through vapour diffusion, in which the supersaturation as a function of time varies throughout the experiment. The effects of optimal pore size can be quantified in terms of supersaturation, and establish a relative contribution between size of pores and supersaturation.

While different theoretical models have demonstrated an optimal pore size for crystallisation of small molecules, which was associated with the critical nuclei radius or a number of lattice sites, experimental results reported in literature of protein crystallisation within mesopores reports the involvement of pores of size at the same order as size of protein. The investigation of optimal pore size experimentally at different supersaturation would complement with existing knowledge published.

Control of seed properties

As discussed, there are many other surface properties that are known to affect the role of nanotemplates on the crystallisation behaviour of proteins. Surface area and the pore volumes were the first variables investigated. It is known that other seed properties, such as particle sizes, surface chemistry and surface roughness influences the protein crystallisation behaviour, a literature review of these properties were presented and discussed in Chapter 3.3. Investigation of the effects of inter-pore spacing on nucleation can also contribute to the understanding in the mechanism of protein crystallisation in confined pores. While the focus of the project was not on the controlled synthesis of nanotemplates, the control of these properties would be important for the improved understanding of the role of heterogeneous seeds. Unfunctionalised silica was used such that the surface chemistry was controlled,

particles of narrow size range can be used. The control of particle size and roughness would involve a detailed investigation of the mesoporous silica design strategies.

Investigation of critical seed concentration

It is evident from the result that a 'critical seed concentration' exists for all seed types, and was observed for HEWL and certain conditions in BLC crystallisation. This value varies depending on the seed types used (see HEWL data) and also the choice of proteins. The effect of protein supersaturation on this value has not been demonstrated within this thesis, knowledge on this property can help identify the seed load to be used for the next step (supersaturation dependency) described below, and also to serve as part of the understanding to predict seed concentration required.

Appendix
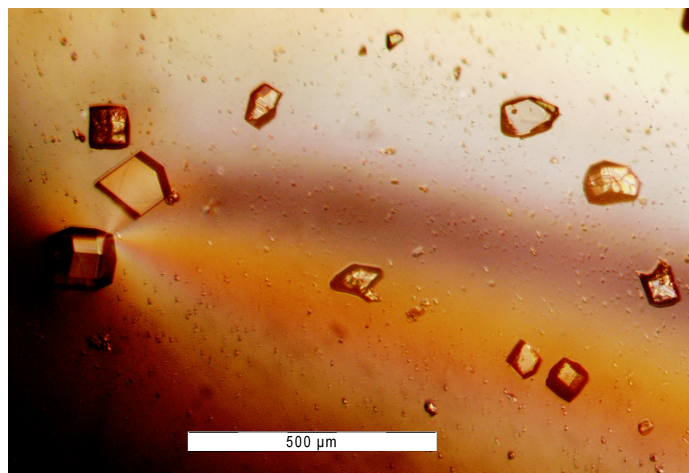
Figure A1 Crystals of lysozyme grown in quiescent mode at 13.5 mg/mL (5 °C)
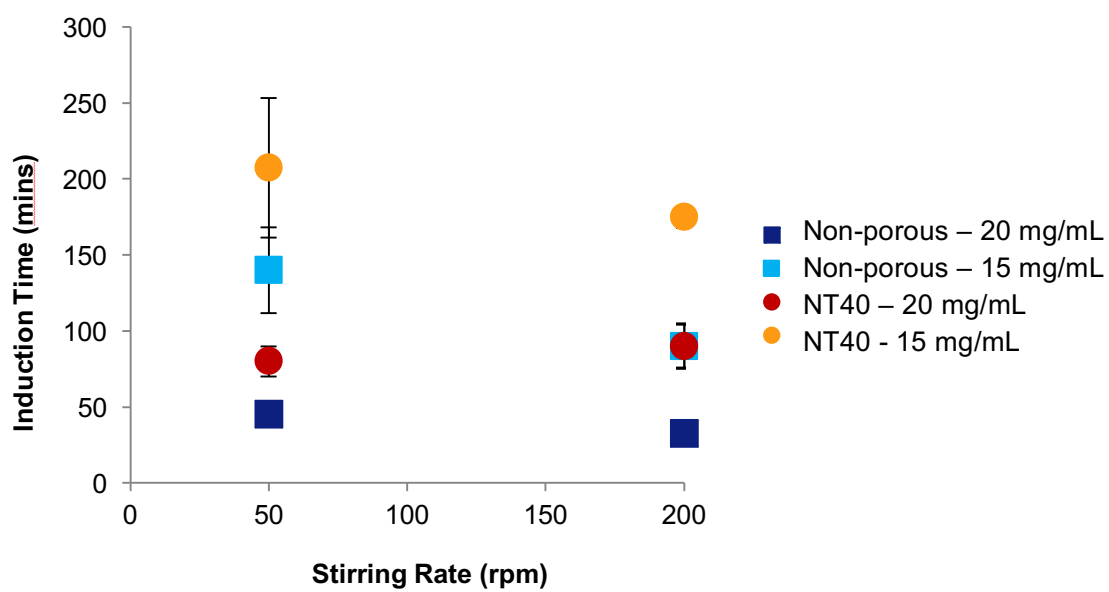
Figure A2 Effects of stirring rate on the effects of different types at different HEWL concentration
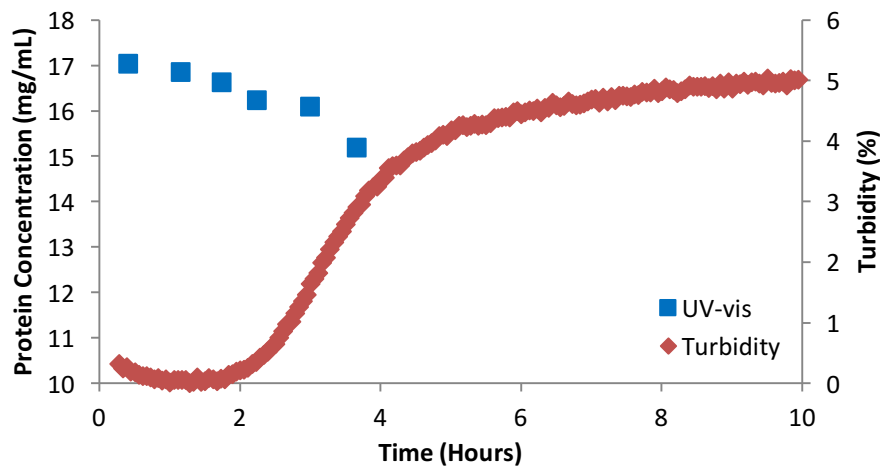
Figure 51 UV-vis/ induction time correlation for crystallisation at 17.5 mg/mL BLC concentration, in the presence of 1.0 mg/mL of NT120 seeds
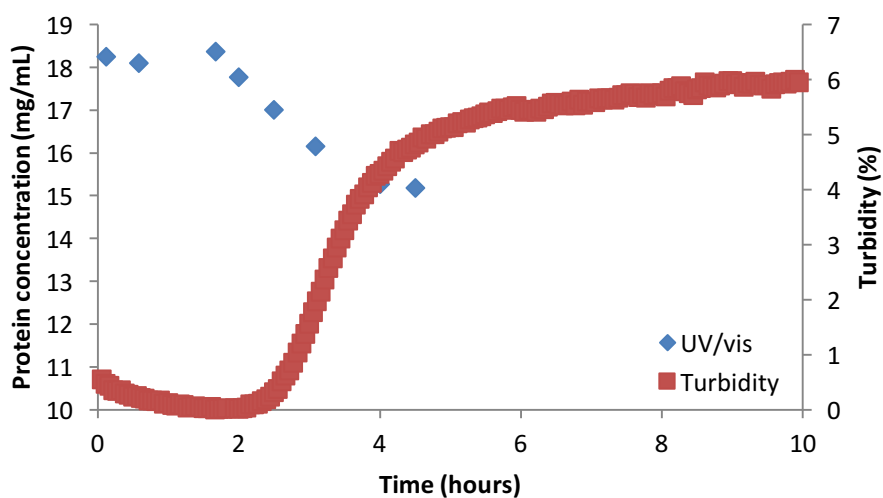
Figure 52 UV-vis/ induction time correlation for crystallisation at 17.5 mg/mL BLC concentration, in the presence of 2.5 mg/mL seeds
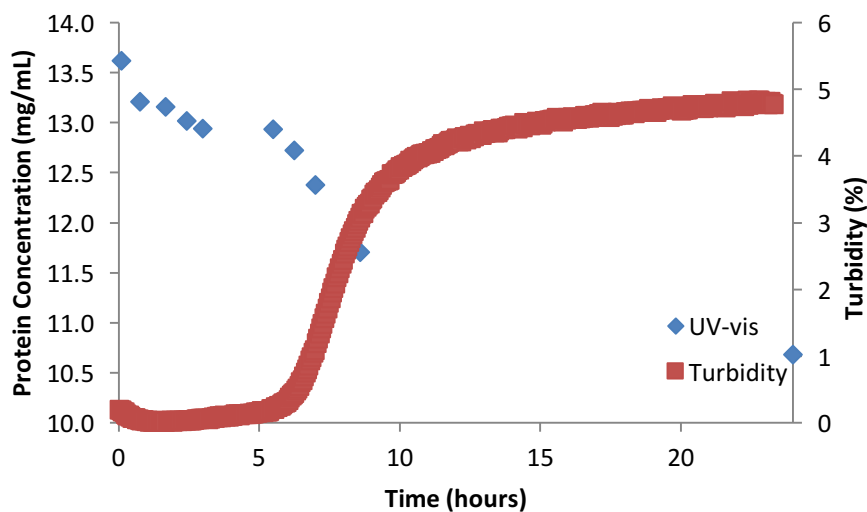
Figure 53 UV-vis/ induction time correlation for crystallisation at 13.5 mg/mL BLC concentration (unseeded)

References

1. Guiochon, G. & Beaver, L. A. Separation science is the key to successful biopharmaceuticals. *J. Chromatogr. A* **1218,** 8836–8858 (2011).

2. Basu, S. K., Govardhan, C. P., Jung, C. W. & Margolin, A. L. Protein crystals for the delivery of biopharmaceuticals. *Expert Opin. Biol. Ther.* **4,** 301–317 (2004).

3. Walsh, G. Biopharmaceutical_benchmarks_2010. **28,** 1–10 (2010).

4. Walsh, G. Biopharmaceutical benchmarks 2014. *Nat. Biotechnol.* **32,** 992–1000 (2014).

5. Langer, E. S. Alleviating Downstream Process Bottlenecks. *Genetic Engineering & Biotechnology News* (2011).

6. dos Santos, R., Carvalho, A. L. & Roque, A. C. A. Renaissance of protein crystallization and precipitation in biopharmaceuticals purification. *Biotechnol. Adv.* **35,** (2016).

7. Langer, E. Trends in Downstream Bioprocessing, in The Science and Business of Biopharmaceuticals. *Biopharm Imternational* 32–36 (2013).

8. Przybycien, T. M., Pujar, N. S. & Steele, L. M. Alternative bioseparation operations: Life beyond packed-bed chromatography. *Curr. Opin. Biotechnol.* **15,** 469–478 (2004).

9. Chung, S. H., Ma, D. L. & Braatz, R. D. Optimal seeding in batch crystallization. *Can. J. Chem. Eng.* **77,** 590–596 (1999).

10. Aamir, E., Nagy, Z. K. & Rielly, C. D. Optimal seed recipe design for crystal size distribution control for batch cooling crystallisation processes. *Chem. Eng. Sci.* **65,** 3602–3614 (2010).

11. Low, D., O'Leary, R. & Pujar, N. S. Future of antibody purification. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **848,** 48–63 (2007).

12. Cohn, E. J. THE PHYSICAL CHEMISTRY OF THE Proteins. *Physiol. Rev.* **5,** 349–437 (1925).

13. Giegé, R. A historical perspective on protein crystallization from 1840 to the present day. *FEBS J.* **280,** 6456–6497 (2013).

14. Shenoy, B., Wang, Y., Shan, W. & Margolin, A. L. Stability of crystalline proteins. *Biotechnol. Bioeng.* **73,** 358–369 (2001).

15. Drenth, J. & Haas, C. Protein Crystals and their Stability. **122,** 107–109 (1992).

16. Pechenov, S., Shenoy, B., Yang, M. X., Basu, S. K. & Margolin, A. L. Injectable controlled release formulations incorporating protein crystals. *J. Control. Release* **96,** 149–158 (2004).

17. Ducruix, A. & Giegé, R. *Crystallization of Nucleic Acids and Proteins*. (Oxford University Press Inc, 1999).

18. McPherson, A. *Crystallisation of Biological Macromolecules*. (Cold Spring Harbour Laboratory Press, 1999).

19. Chayen, N. E. & Saridakis, E. Protein crystallization: from purified protein to diffraction-quality crystal. *Nat. Methods* **5,** 147–153 (2008).

20. Zang, Y., Kammerer, B., Eisenkolb, M., Lohr, K. & Kiefer, H. Towards protein crystallization as a process step in downstream processing of therapeutic antibodies: Screening and optimization at microbatch scale. *PLoS One* **6,** 3–10 (2011).

21. Huettmann, H., Berkemeyer, M., Buchinger, W. & Jungbauer, A. Preparative crystallization of a single chain antibody using an aqueous two-phase system. *Biotechnol. Bioeng.* **111,** 2192–2199 (2014).

22. Khurshid, S., Saridakis, E., Govada, L. & Chayen, N. E. Porous nucleating agents for protein crystallization. *Nat. Protoc.* **9,** 1621–33 (2014).

23.   Bergfors, T. Seeds to crystals. *J. Struct. Biol.* **142,** 66–76 (2003).

24.   Gregg, S. J. & Sing, K. S. W. *Adsorption, Surface Area and Porosity*. (Academic Press, 1982).

25.   Lu, J. & Liong, M. Biocompatibility, Biodistribution, and Drug Delivery Efficiency of Mesoporous Silica Nanoparticles for Cancer Therapy in Animals. *Small.* **6,** 1794–1805 (2011).

26.   Chayen, N. E., Saridakis, E., El-Bahar, R. & Nemirovsky, Y. Porous silicon: an effective nucleation-inducing material for protein crystallization. *J. Mol. Biol.* **312,** 591–595 (2001).

27.   Chayen, N. E., Saridakis, E. & Sear, R. P. Experiment and theory for heterogeneous nucleation of protein crystals in a porous medium. (2005).

28.   Shah, U. V., Williams, D. R. & Heng, J. Y. Y. Selective crystallization of proteins using engineered nanonucleants. *Cryst. Growth Des.* **12,** 1362–1369 (2012).

29.   Kresge. Ordered Mesoporous Molecular Sieves Synthesised by a Liquid Crystal Template Mechanism. *Lett. to Nat.* 710–712 (1992).

30.   Shah, U. V., Allenby, M. C., Williams, D. R. & Heng, J. Y. Y. Crystallization of proteins at ultralow supersaturations using novel three-dimensional nanotemplates. *Cryst. Growth Des.* **12,** 1772–1777 (2012).

31.   Mullin, J. W. *Crystallization*. (Butterworth-Heinemann, 2001). doi:10.1002/0471238961.0318251918152119.a01.pub3

32.   Myerson, A. S. *Handbook of Industrial Cryrstallization*. (Butterworth-Heinemann, 2002).

33.   Gunton, J. D., Shiryayev, A. & Pagan, D. L. *Protein Condensation: Kinetic Pathways to Crystallization and Disease*. (Cambridge University Press, 2007).

34.   Creighton, T. E. Protein folding. *Biochem. J.* **270,** 1–16 (1990).

35.   Jeong, S. H. Analytical methods and formulation factors to enhance protein stability in solution. *Arch. Pharm. Res.* **35,** 1871–1886 (2012).

36.   Manning, M. C., Patel, K. & Borchardt, R. T. Stability of Protein Pharmaceuticals. *Pharmaceutical Research: An Official Journal of the American Association of Pharmaceutical Scientists* **6,** 903–918 (1989).

37.   Amin, S., Barnett, G. V., Pathak, J. A., Roberts, C. J. & Sarangapani, P. S. Protein aggregation, particle formation, characterization & rheology. *Curr. Opin. Colloid Interface Sci.* **19,** 438–449 (2014).

38.   Mahler, H.-C., Friess, W., Grauschopf, U. & Kiese, S. Protein Aggregation: Pathways, Induction Factors and Analysis. *J Pharm Sci* **98,** 2909–2934 (2009).

39.   Sample Preparation for Crystallization : Crystal Growth 101. *Hampton research*

40.   McPherson, A. Introduction to protein crystallization. *Methods* **34,** 254–265 (2004).

41.   McPherson, A., Nguyen, C., Larson, S. B., Day, J. S. & Cudney, B. Development of an alternative approach to protein crystallization. *J. Struct. Funct. Genomics* **8,** 193–198 (2007).

42.   McPherson, A. & Cudney, B. Searching for silver bullets: An alternative strategy for crystallizing macromolecules. *J. Struct. Biol.* **156,** 387–406 (2006).

43.   Larson, S., Day, J., Nguyen, C., Cudney, R. & McPherson, A. Progress in the Development of an Alternative Approach to Macromolecular Crystallization. *Cryst. Growth Des.* **8 no8,** 3038–3052 (2008).

44.   Chernov, A. A. Protein crystals and their growth. *J. Struct. Biol.* **142,** 3–21 (2003).

45.   Durbin, S. D. & Feher, G. Protein Crystallization. *Bull. Menninger Clin.* **77,** 474 (2009).

46. Vekilov, P. G. & Alexander, J. I. D. Dynamics of layer growth in protein crystallization. *Chem. Rev.* **100,** 2061–2089 (2000).

47. Crick, F. H. C. X-Ray Diffraction of Protein Crystals. *Methods Enzymol.* **4,** 127–146 (1957).

48. Deller, M. C. & Rupp, B. Crystallisation of Proteins and Macromolecular Complexes: Past, Present and Future. *eLS* **John Wiley,** (2014).

49. McPherson, A. & Gavira, J. A. Introduction to protein crystallization. *Acta Crystallogr. Sect. FStructural Biol. Commun.* **70,** 2–20 (2014).

50. Luft, J. R., Wolfley, J. R. & Snell, E. H. What's in a drop? Correlating observations and outcomes to guide macromolecular crystallization experiments. *Cryst. Growth Des.* **11,** 651–663 (2011).

51. Vekilov, P. G. [6] Solvent Entropy Effects in the Formation of Protein Solid Phases. *Methods Enzymol.* **368,** 84–105 (2003).

52. Hampton Research. Salt or Protein Crystals ? *Crystal Growth 101* 1–2 (2016).

53. Jones, A. . Crystallization Process Systems. (2002). doi:10.1016/B978-075065520-0/50007-9

54. Kashchiev, D. & van Rosmalen, G. M. Review: Nucleation in solutions revisited. *Cryst. Res. Technol.* **38,** 555–574 (2003).

55. Oxtoby, D. W. Homogeneous nucleation : theory and experiment. *J Phys Condens Matter* **4,** 7627–7650

56. Asherie, N. Protein crystallization and phase diagrams. **34,** 266–272 (2004).

57. Rupp, B. Origin and use of crystallization phase diagrams. *Acta Crystallogr. Sect. FStructural Biol. Commun.* **71,** 247–260 (2015).

58. de Coppet, L. C. No Title. *Ann Chim Phys* **10,** 457–527 (1907).

59. Young, S. W. Crystallisation in Supercooled Liquids. *J Am Chem Soc* **33,** 148–162 (1911).

60. Kashchiev, D. The kinetic approach to nucleation. *Cryst. Res. Technol.* **19,** 1413–1423 (1984).

61. Majeed, S. *et al.* Enhancing protein crystallization through precipitant synergy. *Structure* **11,** 1061–1070 (2003).

62. Rupp, B. & Wang, J. Predictive models for protein crystallization. *Methods* **34,** 390–407 (2004).

63. Green, A. A. Studies in the Physical Chemistry of the Proteins. VIII. The Solubility of Hemoglobin in Concentrated salt Solutions. A study of the Salting Out of Proteins. *J. Biol. Chem.* **93,** 495–516 (1931).

64. Feher, G. & Kam, Z. Nucleation and Growth of Protein Crystals: General Principles and Assays. *Methods Enzymol.* **114,** 77–112 (1985).

65. Shih, Y., Prausnitz, J. M. & Blanch, H. W. Some Characteristics of Protein Precipitation by Salts. *Biotechnol. Bioeng.* **40,** 1155–1164 (1992).

66. Baldwin, R. L. How Hofmeister ion interactions affect protein stability. *Biophys. J.* **71,** 2056–2063 (1996).

67. Hofmeister, F. Zur Lehre von der Wirkung der Salze. *Arch. Exp. Pathol. Pharmakol* **24,** 247 (1888).

68. Zhang, Y. & Cremer, P. S. Interactions between macromolecules and ions: the Hofmeister series. *Curr. Opin. Chem. Biol.* **10,** 658–663 (2006).

69. Dumetz, A. C., Snellinger-O'Brien, A. M., Kaler, E. W. & Lenhoff, A. M. Patterns of protein – protein interactions in salt solutions and implications for protein crystallization. *Protein Sci.* **16,** 1867–1877 (2007).

70.   Arakawa, T. & Timasheff, S. N. Preferential interactions of proteins with salts in concentrated solutions. *Biochemistry* **21,** 6545–6552 (1982).

71.   Polson, A., Potgieter, G. M., Largiers, J. F., Mears, G. E. F. & Joubert, F. J. The fractionation of protein mixtures by linear polymers of high molecular weight. *Biochim. Biophys. Acta - Gen. Subj.* **82,** 463–475 (1964).

72.   McPherson, A. Crytsllization of Proteins from Polyethylene Glycol. *J. Biol. Chem.* **251,** 6300–6303 (1976).

73.   Atha, D. H. & Ingham, K. C. Mechanism of precipitation of proteins by polyethylene glycols. Analysis in terms of excluded volume. *J. Biol. Chem.* **256,** 12108–12117 (1981).

74.   Kramer, R. M., Shende, V. R., Motl, N., Pace, C. N. & Scholtz, J. M. Toward a Molecular Understanding of Protein Solubility : Increased Negative Surface Charge Correlates with Increased Solubility. *Biophys. J.* **102,** 1907–1915 (2012).

75.   Kantardjieff, K. A. & Rupp, B. Protein isoelectric point as a predictor for increased crystallization screening efficiency. *Bioinformatics* **20,** 2162–2168 (2004).

76.   Kashchiev, D. *Nucleation: Basic Theory and Application*. (Butterworth-Heinemann, 2000).

77.   Garside, J. Industrial crystallization from solution. *Chem. Eng. Sci.* **40,** 3–26 (1985).

78.   Vekilov, P. G. & Vorontsova, M. A. Nucleation precursors in protein crystallization. *Acta Crystallogr. Sect. FStructural Biol. Commun.* **70,** 271–282 (2014).

79.   Gibbs, J. W. On the equilibrium of heterogeneous substances. *Trans. Connect. Acad. Sci* **3,** 108–248 (1876).

80.   Gibbs, J. W. On the equilibrium of heterogeneous substances. *Trans. Connect. Acad. Sci* **16,** 343–524 (1878).

81.   Garcia-Ruiz, J. M. Nucleation of protein crystals. *J Struct. Biol* **142,** 22–31 (2003).

82.   DeYoreo, J. J. & Vekilov, P. G. Principles of Crystal Nucleation and Growth. *Rev. Mineral. Geochemistry* **54,** 57–93 (2003).

83.   Nanev, C. N. Protein crystal nucleation : Recent notions. **12,** 4–12 (2007).

84.   Merikanto, J., Zapadinsky, E., Lauri, A. & Vehkam??ki, H. Origin of the failure of classical nucleation theory: Incorrect description of the smallest clusters. *Phys. Rev. Lett.* **98,** 1–4 (2007).

85.   Sear, R. P. Nucleation via an unstable intermediate phase. *J Chem Phys* **131,** 74702 (2009).

86.   Vekilov, P. G. Two-step mechanism for the nucleation of crystals from solution. *J. Cryst. Growth* **275,** 65–76 (2005).

87.   Erdemir, D., Lee, A. Y. & Myerson, A. S. Nucleation of crystals from solution: Classical and two-step models. *Acc. Chem. Res.* **42,** 621–629 (2009).

88.   Sear, R. P. The non-classical nucleation of crystals: microscopic mechanisms and applications to molecular crystals, ice and calcium carbonate. *Int. Mater. Rev.* **57,** 328–356 (2012).

89.   Galkin, O. & Vekilov, P. G. Are nucleation kinetics of protein crystals similar to those of liquid droplets? *J. Am. Chem. Soc.* **122,** 156–163 (2000).

90.   Vekilov, P. G. Dense Liquid Precursor for the Nucleation of Ordered Solid Phases from Solution. *Cryst. Growth Des.* **4,** 671–685 (2004).

91.   Sear, R. P. The non-classical nucleation of crystals: microscopic mechanisms and applications to molecular crystals, ice and calcium carbonate. *Int. Mater. Rev.* **57,** 328–356 (2012).

92.   ten Wolde, P. R. & Frenkel, D. Enhancement of Protein Crystal Nucleation by Critical Density Fluctuations. *Science (80-.).* **277,** 1975–1978 (1997).

93.   Muschol, M. & Rosenberger, F. Interactions in Undersaturated and Supersaturated Lysozyme Solutions: Static and Dynamic Light Scattering Results. *J. Phys. Chem.* **103,** 10424–10432 (1995).

94.   Lin, H., Rosenberger, F. & Nadarajah, A. Convective-diffusive transport in protein crystal growth. **248,** (1995).

95.   Vorontsova, M. A., Maes, D. & Vekilov, P. G. Recent advances in the understanding of two-step nucleation of protein crystals. *Faraday Discuss.* **179,** 27–40 (2015).

96.   Zaccaro, J., Matic, J., Myerson, A. S. & Garetz, B. A. Nonphotochemical, Laser-Induced Nucleation of Supersaturated Aqueous Glycine Produces Unexpected γ-Polymorph. *Cryst. Growth Des.* **1,** 5–8 (2001).

97.   Bonnett, P. E., Carpenter, K. J., Dawson, S. & Davey, R. J. Solution crystallisation via a submerged liquid-liquid phase boundary: oiling out. *Chem. Commun.* 698–699 (2003). doi:10.1039/B212062C

98.   Myerson, A. S. & Trout, B. L. Nucleation from Solution. *Science (80-.).* **341,** 855–856 (2013).

99.   Wolf, S. E., Leiterer, J., Kappl, M., Emmerling, F. & Tremel, W. Early Homogenous Amorphous Precursor Stages of Calcium Carbonate and Subsequent Crystal Growth in Levitated Droplets. *J. Am. Chem. Soc.* **130,** 12342–12347 (2008).

100.   Kawasaki, T. & Tanaka, H. Formation of a crystal nucleus from liquid. *Proc. Natl. Acad. Sci.* **107,** 14036–14041 (2010).

101.   Vekilov, P. G. The two-step nucleation of crystals in solution. *Nanoscale* **1270,** 2346–2357 (2010).

102.   Gliko, O. *et al.* Metastable liquid clusters in super- and undersaturated protein solutions. *J. Phys. Chem. B* **111,** 3106–3114 (2007).

103.   Chayen, N. E. Rigorous filtration for protein crystallization. *J. Appl. Crystallogr.* **42,** 743–744 (2009).

104.   Blow, D. M., Chayen, N. E., Lloyd, L. F. & Saridakis, E. Control of nucleation of protein crystals. *Protein Sci.* **3,** 1638–1643 (1994).

105.   Sohnel, O. & Mullin, J. W. Interpretation of crystallization induction periods. *J. Colloid Interface Sci.* **123,** 43–50 (1988).

106.   Liu, Y. X., Wang, X. J., Lu, J. & Ching, C. B. Influence of the roughness, topography, and physicochemical properties of chemically modified surfaces on the heterogeneous nucleation of protein crystals. *J. Phys. Chem. B* **111,** 13971–13978 (2007).

107.   Young, T. An Essay on the Cohesion of Fluids. *Philos. Trans. R. Soc. London* **95,** 65–87 (1805).

108.   Sear, R. P. Nucleation at contact lines where fluid–fluid interfaces meet solid surfaces. *J. Phys. Condens. Matter* **19,** 466106 (2007).

109.   Meel, J. A. Van, Sear, R. P. & Frenkel, D. Design Principles for Broad-Spectrum Protein-Crystal Nucleants with Nanoscale Pits. **205501,** 1–4 (2010).

110.   van Meel, J. a., Liu, Y. & Frenkel, D. Mechanism of two-step vapour–crystal nucleation in a pore. *Mol. Phys.* **8976,** 1–13 (2015).

111.   Evans, R. Fluids adsorbed in narrow pores: phase equilibria and structure. *J. Phys. Condens. Matter* **2,** 8989 (1990).

112.   Sear, R. P. Nucleation: theory and applications to protein solutions and colloidal suspensions. *J. Phys. Condens. Matter* **19,** 33101 (2007).

113.   Page, A. J. & Sear, R. P. Crystallization controlled by the geometry of a surface. *J. Am. Chem. Soc.* **131,** 17550–17551 (2009).

114. Sholl, C. A. & Fletcher, N. H. Decoration criteria for surface steps. *Acta Metall.* **18,** 1083–1086 (1970).

115. Sliwinska-Bartkowiak, M. *et al.* Freezing behavior in porous glasses and MCM-41. *Colloids Surfaces A Physicochem. Eng. Asp.* **187–188,** 523–529 (2001).

116. Page, A. J. & Sear, R. P. Heterogeneous Nucleation in and out of Pores. *Phys. Rev. Lett.* **97,** 65701 (2006).

117. Diao, Y., Harada, T., Myerson, A. S., Hatton, T. A. & Trout, B. L. The role of nanopore shape in surface-induced crystallization. **10,** 2–4 (2011).

118. Garside, J. & Davey, R. J. Invited Review Secondary Contact Nucleation: Kinetics, Growth and Scale-Up. *Chem. Eng. Commun.* **4,** 393–424 (2007).

119. MELIA, T. P. & MOFFITT, W. P. Kinetic Constants in Nucleation from Aqueous Solution. *Nature* **201,** 1024–1025 (1964).

120. Land, T. A., De Yoreo, J. J. & Lee, J. D. An in-situ AFM investigation of canavalin crystallization kinetics. *Surf. Sci.* **384,** 136–155 (1997).

121. Nanev, C. N., Saridakis, E. & Chayen, N. E. Protein crystal nucleation in pores. *Sci. Rep.* **7,** 35821 (2017).

122. Nanev, C. N. How do crystal lattice contacts reveal protein crystallization mechanism? *Cryst. Res. Technol.* **43,** 914–920 (2008).

123. Nanev, C. N. & Dimitrov, I. L. Layered crystals of apo- and holoferritin grown by alternating crystallization. *Cryst. Res. Technol.* **44,** 908–914 (2009).

124. Durbin, S. D. & Feher, G. Crystal growth studies of lysozyme as a model for protein crystallization. *J. Cryst. Growth* **76,** 583–592 (1986).

125. Grant, M. L. & Saville, D. A. The role of transport phenomena in protein crystal growth. *J. Cryst. Growth* **108,** 8–28 (1991).

126. Hirschler, J. & Fontecilla-camps, J. C. Protein crystal growth rates are face-specifically modified by structurally related contaminants. *J. Cryst. Growth* **171,** 559–565 (1997).

127. Vekilov, P. G. & Rosenberger, F. Protein crystal growth under forced solution flow: experimental setup and general response of lysozyme. *J. Cryst. Growth* **186,** 251–261 (1998).

128. Falcón Rodríguez, C., Aguilera Morales, S. & Falcón Rodríguez, F. Cessation of growth in crystals. *J. Cryst. Growth* **208,** 651–656 (2000).

129. Malkin, A. ., Kuznetsov, Y. . & McPherson, A. In situ atomic force microscopy studies of surface morphology, growth kinetics, defect structure and dissolution in macromolecular crystallization. *J. Cryst. Growth* **196,** 471–488 (1999).

130. Nanev, C. N. On the elementary processes of protein crystallization: Bond selection mechanism. *J. Cryst. Growth* **402,** 195–202 (2014).

131. Nanev, C. N. On the Slow Kinetics of Protein Crystallization. *Cryst. Growth Des.* 1533–1540 (2007).

132. Dasgupta, S., Iyer, G. H., Bryant, S. H., Lawrence, C. E. & Bell, J. A. Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. *Proteins Struct. Funct. Bioinforma.* **28,** 494–514 (1997).

133. Dale, G. E., Oefner, C. & D'Arcy, A. The protein as a variable in protein crystallization. *J. Struct. Biol.* **142,** 88–97 (2003).

134. Derewenda, Z. S. & Vekilov, P. G. Entropy and surface engineering in protein crystallization. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62,** 116–124 (2006).

135. Sumner, J. B. The Isolation and Crystallization of the Enzyme Urease. *J. Biol. Chem.* **69,** 435–441 (1926).

136. Hill, A. F., Barnham, K. J., Bottomley, S. P. & Cappai, R. *Protein folding, Misfolding and Disease: Methods and Protocols*. (Humana Press, 2011).

137. Chayen, N. E. Turning protein crystallisation from an art into a science. *Curr. Opin. Struct. Biol.* **14,** 577–583 (2004).

138. Institute, E. B. Uniprot Current Release Statistics. (2017). Available at: http://www.ebi.ac.uk/uniprot/TrEMBLstats. (Accessed: 14th March 2017)

139. RCSB Protein Data Bank Homepage.

140. Shaw Stewart, P. & Mueller-Dieckmann, J. Automation in biological crystallization. *Acta Crystallogr. Sect. FStructural Biol. Commun.* **70,** 686–696 (2014).

141. Berman, H. M. RCSB Protein Data Bank : Overview. (2010).

142. Bernal, J. D. & Crowfoot, D. X-ray Photographs of Crystalline Pepsin. *Nature* **133,** 794–795 (1934).

143. Carter, C. W. & Carter, C. W. Protein Crystallization Using Incomplete Factorial-Experiments. *J. Biol. Chem.* **254,** 2219–2223 (1979).

144. Jancarik, J. & Kim, S.-H. Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallogr.* **24,** 409–411 (1991).

145. Segelke, B. W. Efficiency analysis of sampling protocols used in protein crystallization screening. *J. Cryst. Growth* **232,** 553–562 (2001).

146. Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. Screening and optimization strategies for macromolecular crystal growth. *Acta Cryst. D* **50,** 414–423 (1994).

147. Gorrec, F. Progress in macromolecular crystallography depends on further miniaturization of crystallization experiments. *Drug Discov. Today* **19,** 1505–1507 (2014).

148. DeLucas, L. J. *et al.* Protein crystallization: Virtual screening and optimization. *Prog. Biophys. Mol. Biol.* **88,** 285–309 (2005).

149. Stevens, R. C. Protein Crystallization. *Curr. Opin. Struct. Biol.* **10,** 558–563 (2000).

150. Pusey, M. L. *et al.* Life in the fast lane for protein crystallization and X-ray crystallography. *Prog. Biophys. Mol. Biol.* **88,** 359–386 (2005).

151. Luft, J. R., Newman, J. & Snell, E. H. Crystallization screening: The influence of history on current practice. *Acta Crystallogr. Sect. FStructural Biol. Commun.* **70,** 835–853 (2014).

152. Hampel, a *et al.* Single crystals of transfer RNA from formylmethionine and phenylalanine transfer RNA's. *Science* **162,** 1384–1387 (1968).

153. Rayment, I. Small-Scale Batch Crystallization Ways & Means of Proteins Revisited : An Underutilized Way to Grow Large Protein Crystals. *Structure* **10,** 147–151 (2002).

154. Adachi, H. *et al.* Novel and Simple Screening Methods for Protein Crystallization by Vapor Diffusion Rate Control. *Jpn. J. Appl. Phys.* **43,** 6264–6267 (2004).

155. Chayen, N. E. A novel technique to control the rate of vapour diffusion, giving larger protein crystals. *J. Appl. Crystallogr.* **30,** 198–202 (1997).

156. Otálora, F., Gavira, J. A., Ng, J. D. & García-Ruiz, J. M. Counterdiffusion methods applied to protein crystallization. *Prog. Biophys. Mol. Biol.* **101,** 26–37 (2009).

157. Ng, J. D., Gavira, J. A. & García-Ruíz, J. M. Protein crystallization by capillary counterdiffusion for

applied crystallographic structure determination. *J. Struct. Biol.* **142,** 218–231 (2003).

158.   Otalora, F. & Garcia Ruiz, J. . Computer model of the diffusion/reaction interplay in the gel acupuncture method. *J. Cryst. Growth* **169,** 361–367 (1996).

159.   Garcia-Ruiz, J. M., Gonzalez-Ramirez, L. A., Gavira, J. A. & Otálora, F. Granada Crystallisation Box: A new device for protein crystallisation by counter-diffusion techniques. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58,** 1638–1642 (2002).

160.   Chayen, N. E., Shaw Stewart, P. D. & Blow, D. M. Microbatch crystallization under oil - a new technique allowing many small-volume crystallization trials. *J. Cryst. Growth* **122,** 176–180 (1992).

161.   Reus, M. A., Guguta, C., Kramer, H. J. M. & ter Horst, J. H. *Solubility : Importance , Measurements and Applications*. (2016).

162.   Ataka, M. Protein crystal growth: an approach based on phase-diagram determination. *Phase Transitions* **45,** 205–219 (1993).

163.   Carbonnaux, C., Ries-kautt, M. & Ducruix, A. Reletive effectiveness of various anions on the solubility of acidic Hypoderma lineatum collagenase at pH 7.2. *Protein Sci.* **4,** 2123–2128 (1995).

164.   Boistelle, R., Astier, J. P. & Haser, R. Solubility , phase transition , kinetic ripening and growth rates of porcine pancreatic a-amylase isoenzymes. **123,** 109–120 (1992).

165.   Chayen, N. E. Methods for separating nucleation and growth in protein crystallisation. *Prog. Biophys. Mol. Biol.* **88,** 329–337 (2005).

166.   Rosenberger, F., Howard, S. B., Sowers, J. W. & Nyce, T. A. Temperature dependence of protein solubility - determination and application to crystallization in X-ray capillaries. *J. Cryst. Growth* **129,** 1–12 (1993).

167.   Christopher, G. K., Phipps, a. G. & Gray, R. J. Temperature-dependent solubility of selected proteins. *J. Cryst. Growth* **191,** 820–826 (1998).

168.   Pusey, M. L. & Gernert, K. A method for rapid liquid-solid phase solubility measurements using the protein lysozyme. *J. Cryst. Growth* **88,** 419–424 (1988).

169.   Cacioppo, E., Munson, S. & Lee Pusey, M. Protein solubilities determined by a rapid technique and modification of that technique to a micro-method. *J. Cryst. Growth* **110,** 66–71 (1991).

170.   Baumgartner, K. *et al.* Determination of protein phase diagrams by microbatch experiments: Exploring the influence of precipitants and pH. *Int. J. Pharm.* **479,** 28–40 (2015).

171.   Talreja, S. *et al.* Determination of the phase diagram for soluble and membrane proteins. *J. Phys. Chem. B* **114,** 4432–41 (2010).

172.   Saridakis, E. & Chayen, N. E. Systematic improvement of protein crystals by determining the supersolubility curves of phase diagrams. *Biophys. J.* **84,** 1218–22 (2003).

173.   Yang, M. X. *et al.* Crystalline monoclonal antibodies for subcutaneous delivery. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 6934–9 (2003).

174.   Nagy, Z. K. & Braatz, R. D. Advances and New Directions in Crystallization Control. *Annu. Rev. Chem. Biomol. Eng* **3,** 55–75 (2012).

175.   Liu, J. J., Ma, C. Y. & Wang, X. Z. Imaging protein crystal growth behaviour in batch cooling crystallisation. *Chinese J. Chem. Eng.* **24,** 101–108 (2016).

176.   DeMattei, R. C. & Feigelson, R. S. Controlling nucleation in protein solutions. *J. Cryst. Growth* **122,** 21–30 (1992).

177.   DeMattei, R. C. & Feigelson, R. S. Thermal methods for crystallizing biological macromolecules. *J. Cryst. Growth* **128,** 1225–1231 (1993).

178. Ryu, B. H., Jones, M. J. & Ulrich, J. Crystallization of hen egg white lysozyme by solvent freeze-out: Effect of cooling rate on protein inclusion in the ice layer. *Chem. Eng. Technol.* **33,** 1695–1698 (2010).

179. Ryu, B. H. & Ulrich, J. Controlled nucleation and growth of protein crystals by solvent freeze-out. *Cryst. Growth Des.* **12,** 6126–6133 (2012).

180. Yu, X., Wang, J. & Ulrich, J. Purification of Lysozyme from Protein Mixtures by Solvent-Freeze-Out Technology. *Chem. Eng. Technol.* **37,** 1353–1357 (2014).

181. Diaz Borbon, V. & Ulrich, J. Solvent freeze out crystallization of lysozyme from a lysozyme-ovalbumin mixture. *Cryst. Res. Technol.* **47,** 541–547 (2012).

182. Groß, M. & Kind, M. Bulk Crystallization of Proteins by Low-Pressure Water Evaporation. *Chem. Eng. Technol.* **39,** 1483–1489 (2016).

183. Huettmann, H., Zich, S., Berkemeyer, M., Buchinger, W. & Jungbauer, A. Design of industrial crystallization of interferon gamma: Phase diagrams and solubility curves. *Chem. Eng. Sci.* **126,** 341–348 (2015).

184. Hekmat, D., Hebel, D., Schmid, H. & Weuster-Botz, D. Crystallization of lysozyme: From vapor diffusion experiments to batch crystallization in agitated ml-scale vessels. *Process Biochem.* **42,** 1649–1654 (2007).

185. Smejkal, B. *et al.* Protein crystallization in stirred systems-scale-up via the maximum local energy dissipation. *Biotechnol. Bioeng.* **110,** 1956–1963 (2013).

186. Judge, R. A., Johns, M. R. & White, E. T. Protein purification by bulk crystallization: The recovery of ovalbumin. *Biotechnol. Bioeng.* **48,** 316–323 (1995).

187. Carbone, M. N., Judge, R. A. & Etzel, M. R. Evaluation of a model for seeded isothermal batch protein crystallization. *Biotechnol. Bioeng.* **91,** 84–90 (2005).

188. Maosoongnern, S., Flood, C., Flood, A. E. & Ulrich, J. Crystallization of lysozyme from lysozyme – ovalbumin mixtures : Separation potential and crystal growth kinetics. *J. Cryst. Growth* 0–1 (2016). doi:10.1016/j.jcrysgro.2016.09.049

189. Jacobson, C., Garside, J. & Hoare, M. Nucleation and growth of Microbial Lipase from Clarified Concentrated Fermentation Broths. *Biotechnol. Bioeng.* **57,** 666–675 (1998).

190. Giffard, M. *et al.* Urate oxidase purification by salting-in crystallization: Towards an alternative to chromatography. *PLoS One* **6,** 1–9 (2011).

191. Genck, W. J. *et al.* Recent Advances for Seeding a Crystallization Process: A Review of Modern Techniques. *Chem. Eng. Res. Des.* **80,** 799–805 (2002).

192. D'Arcy, A., Mac Sweeney, A. & Haber, A. Using natural seeding material to generate nucleation in protein crystallization experiments. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **59,** 1343–1346 (2003).

193. Bergfors, T. *Protein Crystallisation*. (International University Line, 2009).

194. McPherson A. & Shlichta, P. Heterogeneous and Epitaxial Nucleation of Protein Crystals on Mineral Surfaces. *Science (80-.).* **239,** 385 LP-387 (1988).

195. Edwards, A. M., Darst, S. A., Hemming, S. A., Li, Y. & Kornberg, R. D. Epitaxial Growth of Protein Crystals on Lipid Layers. *Nature* **1,** 638–653 (1994).

196. Allahyarov, E., Sandomirski, K., Egelhaaf, S. U. & Löwen, H. Crystallization seeds favour crystallization only during initial growth. *Nat. Commun.* **6,** 7110 (2015).

197. Weichsel, U., Segets, D., Janeke, S. & Peukert, W. Enhanced Nucleation of Lysozyme Using Inorganic Silica Seed Particles of Different Sizes. *Cryst. Growth Des.* **15,** 3582–3593 (2015).

198. Wenzel, R. N. Surface Roughness and Contact Angle. *J. Psychosom. Res.* **53,** 1466–1467 (1949).

199.   Curcio, E., Fontananova, E., Di Profio, G. & Drioli, E. Influence of the structural properties of poly(vinylidene fluoride) membranes on the heterogeneous nucleation rate of protein crystals. *J. Phys. Chem. B* **110,** 12438–45 (2006).

200.   Ghatak, A. S. & Ghatak, A. Disordered nanowrinkle substrates for inducing crystallization over a wide range of concentration of protein and precipitant. *Langmuir* **29,** 4373–4380 (2013).

201.   Fermani, S., Falini, G., Minnucci, M. & Ripamonti, A. Protein crystallization on polymeric film surfaces. *J. Cryst. Growth* **224,** 327–334 (2001).

202.   Tosi, G., Fermani, S., Falini, G., Gavira Gallardo, J. A. & Garc??a Ruiz, J. M. Crystallization of proteins on functionalized surfaces. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64,** 1054–1061 (2008).

203.   Falini, G., Fermani, S., Conforti, G. & Ripamonti, A. Protein crystallisation on chemically modified mica surfaces. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58,** 1649–1652 (2002).

204.   Tsekova, D. S., Williams, D. R. & Heng, J. Y. Y. Effect of surface chemistry of novel templates on crystallization of proteins. *Chem. Eng. Sci.* **77,** 201–206 (2012).

205.   Stolyarova, S., Saridakis, E., Chayen, N. E. & Nemirovsky, Y. A model for enhanced nucleation of protein crystals on a fractal porous substrate. *Biophys. J.* **91,** 3857–63 (2006).

206.   Saridakis, E. *et al.* Protein crystallization facilitated by molecularly imprinted polymers. *Proc. Natl. Acad. Sci.* **108,** 18566–18566 (2011).

207.   Minton, A. P. Confinement as a determinant of macromolecular structure and reactivity. *Biophys. J.* **63,** 1090–1100 (1992).

208.   Zhou, H. X. & Dill, K. A. Stabilization of proteins in confined spaces. *Biochemistry* **40,** 11289–11293 (2001).

209.   Mittal, J. & Best, R. B. Thermodynamics and kinetics of protein folding under confinement. **105,** 20233–20238 (2008).

210.   Shah, U. V. *et al.* Crystallisation via novel 3D nanotemplates as a tool for protein purification and bio-separation. *J. Cryst. Growth* (2016). doi:10.1016/j.jcrysgro.2016.09.029

211.   Delmas, T., Roberts, M. M. & Heng, J. Y. Y. Nucleation and Crystallization of Lysozyme: Role of Substrate Surface Chemistry and Topography. *J. Adhes. Sci. Technol.* **25,** 357–366 (2012).

212.   Yi, Y. J. & Myerson, A. S. Laboratory Scale Batch Crystallization and the Role of Vessel Size. *Chem. Eng. Res. Des.* **84,** 721–728 (2006).

213.   Lee, T., Vaghjiani, J., Lye, G. & Turner, M. A systematic approach to the large-scale production of protein crystals. *Enzyme Microb. Technol.* **26,** 582–592 (2000).

214.   Chianese, A., Contaldi, A. and Mazzarotta, B. Primary nucleation of sodium perborate in aqueous solutions,. *J Cryst. Growth* **78,** 279–290 (1986).

215.   Chadwick, K., Chen, J., Myerson, A. S. & Trout, B. L. Toward the rational design of crystalline surfaces for heteroepitaxy: Role of molecular functionality. *Cryst. Growth Des.* **12,** 1159–1166 (2012).

216.   Asefa, T. & Tao, Z. Biocompatibility of Mesoporous Silica Nanoparticles. *Chem. Res. Toxicol.* (2012).

217.   Rengarajan, G. T., Enke, D. & Beiner, M. Crystallization Behavior of Acetaminophen in Nanopores. *Open Phys. Chem. J.* 18–24 (2007). doi:10.2174/187406770701011807

218.   Dwyer, L. M., Michaelis, V. K., O'Mahony, M., Griffin, R. G. & Myerson, A. S. Confined crystallization of fenofibrate in nanoporous silica. *CrystEngComm.* **33,** 395–401 (2015).

219.   Lapidot, T. & Heng, J. Y. Y. Functionalized Mesoporous Silica for the Control of Crystallization Fouling. *Ind. Eng. Chem. Res.* **55,** 11475–11479 (2016).

220. Kubota, N., Doki, N., Yokota, M. & Sato, A. Seeding policy in batch cooling crystallization. *Powder Technol.* **121,** 31–38 (2001).

221. Kashchiev, D., Verdoes, D. & van Rosmalen, G. M. Induction time and metastability limit in new phase formation. *J. Cryst. Growth* **110,** 373–380 (1991).

222. Melvin Avrami & Avrami, M. Kinetics of phase change. I general theory. *J. Chem. Phys.* **7,** 1103–1112 (1939).

223. Aitken, A. & Learmonth, M. P. in *The Protein Protocols Handbook* (ed. Walker, J. M.) 3–6 (Humana Press, 2002). doi:10.1385/1-59259-169-8:3

224. Goldfarb, R., Saidel, L. J. & Mosovich, E. The Ultraviolet Absorption Spectra of Proteins. *J. Biol. Chem.* **193,** 397–404 (1951).

225. Huffman, S., Keyur, S. & Ferraiolo, J. UV-Vis Based Determination of Protein Concentration: Validating and Implementing Slope Measurements Using Variable Pathlength Technology. *BioProcess International* (2014).

226. Lawler, D. M. Spectrophotometry: Turbidimetry and Nephelometry. *Encycl. Anal. Sci.* 343–351 (2005). doi:10.1016/B0-12-369397-7/00718-4

227. United States Pharmacopeia. <788> Particulate Matter in Injections. *Usp* **34,** 326–328 (2011).

228. Moscosa-Santillan, M., Bals, O., Fauduet, H., Porte, C. & Delacroix, A. Study of batch crystallization and determination of an alternative temperature-time profile by on-line turbidity analysis - application to glycine crystallization. *Chem. Eng. Sci.* **55,** 3759–3770 (2000).

229. Fleming, A. Downloaded from http://rsnr.royalsocietypublishing.org/ on February 15 , 2017. *Proc. R. Soc. London, Ser. B* **93,** 306–317 (1922).

230. Blake, C. C. F. *et al.* Structure of Hen Egg-White Lysozyme: A Three-dimensional Fourier Synthesis at 2 Å Resolution. *Nature* **206,** 757–761 (1965).

231. Cacioppo, E. & Pusey, M. L. The solubility of the tetragonal form of hen egg white lysozyme from pH 4.0 to 5.4. *J. Cryst. Growth* **114,** 286–292 (1991).

232. Howard, S. B., Twigg, P. J., Baird, J. K. & Meehan, E. J. The solubility of hen egg-white lysozyme. *J. Cryst. Growth* **90,** 94–104 (1988).

233. Forsythe, E. L., Judge, R. A. & Pusey, M. L. Tetragonal chicken egg white lysozyme solubility in sodium chloride solutions. *J. Chem. Eng. Data* **44,** 637–640 (1999).

234. Alderton, G., Ward, W. H. & Fevold, H. L. Isolation of Lysozyme from Egg White. *J. Biol. Chem.* **157,** 43–58 (1945).

235. Alderton, G. & Fevold, H. L. Direct Crystallization of Lysozyme. *J. Biol. Chem.* **164,** 1–5 (1946).

236. Maosoongnern, S., Diaz Borbon, V., Flood, A. E. & Ulrich, J. Introducing a fast method to determine the solubility and metastable zone width for proteins: Case study lysozyme. *Ind. Eng. Chem. Res.* **51,** 15251–15257 (2012).

237. Galkin, O. & Vekilov, P. G. Nucleation of protein crystals: Critical nuclei, phase behavior, and control pathways. *J. Cryst. Growth* **232,** 63–76 (2001).

238. Cayey, N. W. & Estrin, J. Secondary Nucleation in Agitated, Magnesium Sulfate Solutions. *Ind. Eng. Chem. Fundam.* **6,** 13–20 (1967).

239. Lapidot, T. Heterogeneous Crystallisation of Calcium Sulphate in the Presence of Silica Particles. (2016).

240. Allen, R. J., Valeriani, C., Tanase-Nicola, S., Ten Wolde, P. R. & Frenkel, D. Homogeneous nucleation under shear in a two-dimensional Ising model: Cluster growth, coalescence, and breakup. *J. Chem.*

Phys. **129,** (2008).

241.   Murthy, M. R. N., Reid, T. J., Sicignano, A., Tanaka, N. & Rossmann, M. G. in *The Biological Chemistry of Iron: A Look at the Metabolism of Iron and Its Subsequent Uses in Living Organisms Proceedings of the NATO Advanced Study Institute held at Edmonton, Alberta, Canada, August 13 -- September 4, 1981* (eds. Dunford, H. B., Dolphin, D., Raymond, K. N. & Sieker, L.) 439–458 (Springer Netherlands, 1982). doi:10.1007/978-94-009-7882-9_29

242.   Longley, W. The Crystal Structure of Bovine Liver Catalase: A Combined Study by X-ray Diffraction and Electron Microscopy. *J. Mol. Biol.* **30,** 323–327 (1967).

243.   Díaz, A., Loewen, P. C., Fita, I. & Carpena, X. Thirty years of heme catalases structural biology. *Arch. Biochem. Biophys.* **525,** 102–110 (2012).

244.   Hashemnia, S. *et al.* Diminishing of aggregation for bovine liver catalase through acidic residues modification. *Int. J. Biol. Macromol.* **40,** 47–53 (2006).

245.   Bartoszek, M. & Sułkowski, W. The study of pH influence on bovine liver catalase by means of UV-VIS spectroscopy and spin labelling method. *... Environ. Stud.* **15,** 41–43 (2006).

246.   Bartoszek, M. & Kściuczyk, M. Study of the temperature influence on catalase using spin labelling method. *J. Mol. Struct.* **744–747,** 733–736 (2005).