Characterizing Protein Crystal Nucleation

A Dissertation

Presented to

The Faculty of the Graduate School of Arts and Sciences

Brandeis University

Martin A. Fisher School of Physics

Seth Fraden, Department of Physics, Advisor

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

by

Sathish V. Akella

February, 2014

This dissertation, directed and approved by Sathish V. Akella's committee, has been accepted and approved by the Graduate Faculty of Brandeis University in partial fulfillment of the requirements for the degree of:

DOCTOR OF PHILOSOPHY

Malcolm Watson, Dean of Arts and Sciences

Dissertation Committee:
Seth Fraden, Department of Physics, Chair
Michael F. Hagan, Department of Physics
Dagmar Ringe, Department of Chemistry

To,

Father & Mother

Acknowledgments

I wish to express my sincere thanks to my thesis adviser, Prof. Seth Fraden, for his guidance and support in my research career at Brandeis.

I would like to thank Dr. Aleksey Lomakin of MIT for his valuable insights and suggestions during the discussions.

I am thankful to Prof. Dagmar Ringe and Prof. Michael Hagan for being on my dissertation committee.

I would like to thank Matt McNeely and Greg Widberg for their technical support and teaching the required instrumentation skills.

Last but not least, I would like to thank the National Science Foundation for the financial support.

Abstract

Characterizing Protein Crystal Nucleation

A dissertation presented to the Faculty of
the Graduate School of Arts and Sciences of
Brandeis University, Waltham, Massachusetts

by Sathish V. Akella

We developed an experimental microfluidic based technique to measure the nucleation rates and successfully applied the technique to measure nucleation rates of lysozyme crystals. The technique involves counting the number of samples which do not have crystals as a function of time. Under the assumption that nucleation is a Poisson process, the fraction of samples with no crystals decays exponentially with the decay constant proportional to nucleation rate and volume of the sample. Since nucleation is a random and rare event, one needs to perform measurements on large number of samples to obtain good statistics. Microfluidics offers the solution of producing large number of samples at minimal material consumption. Hence, we developed a microfluidic method and measured nucleation rates of lysozyme crystals in supersaturated protein drops, each with volume of ≈ 1 nL. Classical Nucleation Theory (CNT) describes the kinetics of nucleation and predicts the functional form of nucleation rate in terms of the thermodynamic quantities involved, such as supersaturation, temperature, etc. We analyzed the measured nucleation rates in the context of CNT and obtained the activation energy and the kinetic pre-factor characterizing the nucleation process. One conclusion is that heterogeneous nucleation dominates crystallization.

We report preliminary studies on selective enhancement of nucleation in one of the crystal polymorprhs of lysozyme (spherulite) using amorphous mesoporous bioactive gel-glass[48, 49], $CaO.P_2O_5.SiO_2$ (known as bio-glass) with $2-10$ nm pore-size diameter distribution. The pores act as *heterogeneous* nucleation centers and claimed[46]

to enhance the nucleation rates by molecular confinement. The measured kinetic profiles of crystal fraction of spherulites indicate that the crystallization of spherulites may be proceeding via secondary nucleation pathways.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Determining the structure of a protein (e.g. an enzyme) is an important step towards understanding the function of the protein molecule and designing drugs for certain diseases. In determining the structure of a protein using X-ray diffraction, one needs to crystallize the protein to obtain a large and X-ray diffraction (XRD) quality crystal. Many crystallographers exhaustively scan the protein-precipitant phase space in search of the crystallization conditions which yield XRD quality crystals. The physical processes underlying crystallization are the nucleation and the growth. *Nucleation* is an activated process by which embyros of the stable crystal phase spontaneously appear from the metastable bulk phase, known as *nucleus*. *Growth* is the growth of the nucleus ensuing nucleation. Nucleation is a frequently observed phenomenon in nature. Formation of rain drops in clouds, formation of cataract due to a liquid-liquid phase transition of eye-lens proteins, polymerization of sickle cell anemia hemoglobin[19] are just few examples of wide range of phenomena where nucleation occurs. Nucleation is 'homogeneous' when it occurs spontaneously in the bulk of the metastable phase or 'heterogeneous' when it occurs in contact with foreign substance such as container walls, cat whiskers, or even precipitate of the protein under

study. Almost every nucleation observed in nature is heterogeneous[37]. Classical Nucleation Theory (CNT) attempts to describe the kinetics of nucleation in terms of the thermodynamic parameters involved such as temperature and supersaturation. To characterize nucleation, our approach is to experimentally measure the nucleation rates and analyze them according to the predictions of CNT to obtain the activation energy and extract the interfacial energy between the stable crystalline phase and metastable bulk phase. Here we give a brief introduction to the list (table 1.1) of reported techniques in the literature to measure nucleation rates, in a chronological order.

1. In 1950s, Turnbull[9, 10, 11] developed a drop-based method to characterize nucleation in supercooled liquid metals. The method involves suspending monodisperse droplets of liquid metals in an inert medium (Ethyl alcohol or Methyl cyclopentane) and measuring the total volume change as a function of time. Assuming the solidification of drops occurs at shorter time scales than nucleation, the volume change is a measure of number of solidified droplets. The change in volume decays exponentially with time and the decay constant is proportional to nucleation rate.

2. Vekilov et al[17], developed a drop-based technique to measure nucleation rates of protein crystals. The technique involves placing a drop of protein solution in an inert oil and incubating the sample at temperature T_1 for a period of Δt to nucleate the crystal and subsequently growing the nuclei at temperature T_2. The temperatures T_1 and T_2 are chosen such that only nucleation occurs at T_1 and no nucleation occurs at T_2 but the system is sufficiently supersaturated to grow the nuclei. Under the assumption that the characteristic time scale for growth is much longer than quench period Δt, the number of crystals formed

in a drop varies linearly with Δt, with slope proportional to nucleation rate.

3. Laval[4] and Salmon[5] used Turnbull's method to obtain nucleation rates of KNO_3 crystals in 100nL drops using microfluidics.

4. Veesler[21, 25], Fraden[22] have extended the Vekilov's technique to measure nucleation rates of lysozyme crystal in 100 nL drops using microfluidics. Fraden group developed a microfluidic PhaseChip[23] to increase the supersaturation by varying the concentration instead of temperature. Another advantage with the PhaseChip is that, it can be used to study the temperature-concentration phase diagrams.

5. Darcy and Wiencek[1] measured the enthalpy of lysozyme crystallization using microcalorimetry. Zukoski[6] et al have estimated the nucleation rates from the enthalpy measurements as the heat released during crystallization is a measure of nucleation.

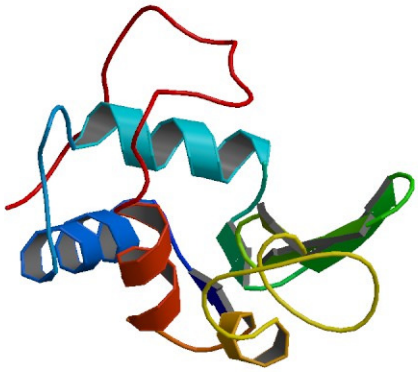
6. ter Horst et al[2, 3] measured the nucleation rates of m-aminobenzoic acid, L-histidine and Isonicotinamide using Turnbull's technique. Instead of crystallized drop fraction, they measured the induction times for the appearance of crystals in drops. Induction time is the elapsed time between the starting of the experiment and the time till a detectable sized crystal appears in a the drop. The measured induction times follow an exponential distribution whose decay constant is proportional to nucleation rate.

With the advent of free-electron lasers (FEL), the constraint of obtaining large and well diffracting crystals is relaxed. FELs produce intense X-ray pulses for periods of time smaller than required for damaging a crystal. The multi-angle diffraction data is obtained by diffracting a large number of randomly oriented small (sub-micron)

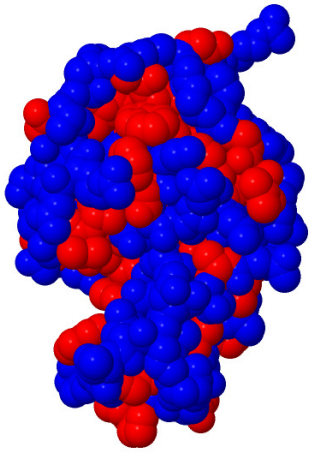
S. No.	Groups	System
1	Turnbull[9, 10, 11], La Mer[14]	Supercooled liquid metals
2	Vekilov[17, 19], **Vessler[21, 25]**, **Fraden[22]**, **Wagner[24]**	Lysozyme, Sickle cell Hemoglobin
3	Darcy[1], Zukoski[6]	Lysozyme
4	ter Horst[2, 3]	m-aminobenzoic acid (m-ABA), L-histidine (L-His), Isonicotinamide
5	**Laval[4]**, **Salmon[5]**	KNO_3

Table 1.1: Nucleation rate measurement techniques reported in literature. The **bolded** references are measurement techniques using microfluidics.

crystals. FEL X-ray diffraction does not need cryo-protection against radiation damage because the crystals are exposed only for a period of time shorter than the time needed for radiation damage to occur and disposed after single use. Usually cryo-protection induces stresses in crystals which lead to distortion in the crystal structure. Another advantage of FEL X-ray diffraction is that the crystal structure can be obtained from small crystals. Therefore, understanding and characterizing nucleation in small volumes can improve diffraction obtained from FEL X-ray sources. Towards that end, we have adopted Turnbull's technique to measure nucleation rates of protein crystals using drop volumes of < 1 nL. This method involves nucleating large number of identical and independent drops at constant temperature and counting number of drops without crystals as a function of time. Under the assumption that nucleation is a Poisson process, the number of drops without crystals decays exponentially with time and the decay constant is proportional to the nucleation rate and the volume of the drop. Lysozyme is an enzyme consisting of 129 amino acid residues, with a molecular weight of 14,500 Daltons. Lysozyme is a globular protein (figure 1.1), roughly ellipsoidal in shape with two minor axes 3.3 nm and a major axis 5.5 nm in diameter[15]. Lysozyme has a net positive charge of $8.5e^-$ at pH



(a) Ribbon structure of lysozyme. The structure highlights the secondary structures (alpha helical and beta sheet structures) of lysozyme.



(b) Ball and stick structure of lysozyme. Blue and red colored regions represent the hydrophillic and hydrophobic regions of lysozyme.

Figure 1.1: Structure of Hen Egg White Lysozyme(HEWL) obtained from protein data bank. (http://www.rcsb.org/pdb/explore/explore.do?structureId=2lyz)

7.0. The isoelectric point of lysozyme is pH 11.2[39]. It is the first enzyme whose structure to be determined in 1960s using X-ray crystallography. Hen Egg White Lysozyme (HEWL) is an inexpensive protein. As a consequence, researchers studied and characterized the nucleation and growth kinetics of lysozyme crystallization under various physical/chemical conditions. The availability of abundant literature on lysozyme crystallization led us to study the nucleation kinetics of lysozyme crystallization using the developed technique. Lysozyme predominantly crystallizes into tetragonal crystal form. However, lysozyme crystals grown above 25℃ transform into orthorhombic crystalline form[16]. Since all the nucleation and growth experiments reported in this work are performed at temperatures $\leq$ 12℃, we only observe tetragonal crystal form. Table 1.2 lists the crystal properties of tetragonal lysozyme[1].

[1]Source: http://www.rcsb.org/pdb/explore/explore.do?structureId=2lyz

Space Group	$P4_32_12$
Unit Cell Dimensions	a = 79.10 Å, b = 79.10 Å, c = 37.90 Å; $\alpha = 90°$, $\beta = 90°$, $\gamma = 90°$.

Table 1.2: Properties of tetragonal lysozyme crystals.

We measure the nucleation rates of lysozyme crystals using the proposed method. Surprisingly, we obtain two nucleation rates at every nucleation rate measurement. Pound and La Mer[14] observed similar phenomenon in nucleation experiments with super-cooled tin. They proposed a nucleation mechanism which accounts for the presence of impurities and predicts the existence of two nucleation rates, 'slow' and 'fast'. The 'fast' process is the result of nucleation on the impurities. We used the Pound and La Mer model[14] to describe the nucleation in our system and obtained 'fast' and 'slow' nucleation rates at every measurement. We analyzed the 'slow' and 'fast' nucleation rates according to the predictions of CNT and obtained the activation energy and the kinetic pre-factor associated with the nucleation. Contrary to conventional wisdom the kinetic pre-factor plays a greater role in determining the nucleation kinetics than does the activation energy. Also, analysis[42] of the kinetic pre-factor indicates that both of the observed nucleation processes are *heterogeneous* in nature. The impurities causing the 'slow' nucleation are lysozyme aggregates ($0.1 - 10$ μm) which are formed due to the depletion interaction induced by the Poly-Ethylene Glycol (PEG) molecules. We speculate that the 'fast' nucleation is primarily induced by the sub-micron (< 220 nm) clusters of lysozyme. We observed that the number of these impurities increases upon lowering the temperature suggesting the formation of lysozyme clusters at high supersaturations. Dynamic Light Scattering (DLS) studies on lysozyme under crystallization conditions by Muschol[26] and Chayen[27] and Small Angle X-ray Scattering (SAXS) studies on supersaturated solutions of lysozyme by Shurtenburger[28] also corroborate the idea of cluster formation.

We also studied the effect of bio-glass on the nucleation kinetics of lysozyme polymorphs. It has been argued[46] that bio-glass enhances the nucleation by molecular confinement. We report preliminary studies on the nucleation kinetics of tetragonal and spherulite polymorphs of lysozyme in the absence and presence of bio-glass. The kinetic profiles of tetragonal and spherulite polymorphs are qualitatively different indicating different mechanisms for the formation of tetragonal and spherulite crystals. We lay out the possible nucleation scenarios and propose further experiments.

Chapter 2

Materials & Methods

2.1 Materials

Lysozyme from chicken egg white is purchased from Sigma Aldrich (Product Number L6876). Without further purification, we dissolved the protein in 0.1 M sodium acetate (Fisher Scientific, Cat. No. S210-500) buffer at pH 4.8 (Thermo Orion pH Meter Model 330). In all our experiments the protein solution is centrifuged (Eppendorf, Centrifuge 5415C) for half an hour at $\simeq$10,000g and filtered through 0.22 μm cellulose acetate filters (VWR International, Cat. No. 28145-477). Lysozyme concentration is measured using Thermo Scientific nano UV-VIS spectrophotometer with lysozyme extinction co-efficient $\epsilon = 2.64$ mL mg^{-1}cm^{-1} at 280 nm. Stock solutions of 20% w/v NaCl (Fisher Scientific, Cat. No. S271-1) and 25% w/v PEG 8kD (OmniPur EMD, Cat. No. 6510) solutions are also prepared in 0.1 M Sodium Acetate buffer at pH 4.8 and filtered through 0.22 μm cellulose acetate filters.

2.2 Methods

2.2.1 Sample Preparation

Nucleation is a rare and random event. In order to get good statistics, one needs to perform large number of identical and independent experiments. Microfluidics offers the solution of producing and storing large number of identical and independent protein drops. In the following sections, we explain the production and storage of drops.

Droplet Generation

We produce emulsion drops using single flow and co-flow microfluidic devices as needed. In a microfluidic device, water-in-oil emulsion drops are produced at the nozzle where an aqueous stream is sheared off by the oil stream (figure 2.1). We use fluorinated oil, HFE-7500 (3M) containing 2% (w/w) EA-surfactant[51] (RainDance Technologies, Inc.) as the oil medium to prevent any adsorption of protein molecules at the droplet and oil interface. The surfactant reduces the shear forces required to produce drops by decreasing the interfacial tension and also stabilizes the drops from coalescing. The advantage of using a single flow microfluidic (figure 2.1a) device is the mixing of protein and precipitant is performed off-chip, where one can obtain high accuracy in the chemical composition of protein and precipitant mix.

To produce drops at high supersaturations, we mix the protein and precipitant on chip using a co-flow microfluidic device (figure 2.1b). Since the number of nucleation events is proportional to volume, there is a supersaturation limit at which nucleation starts occurring in off-chip mixing at pipette-able volumes, $\simeq 10\ \mu$L, while in on-chip mixing the drops produced are at much smaller volumes, $\simeq 100$ pL, therefore one can achieve higher supersaturations. We adjust the flow rates of the streams according

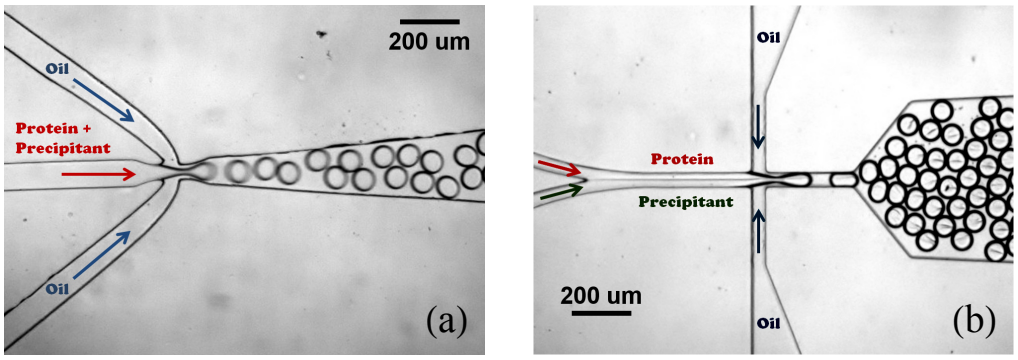


Figure 2.1: Generation of emulsion drops using (a) single flow (b) co-flow microfluidic devices. Notice that, in the co-flow microfluidic devices, the protein and precipitant streams are separated by a visible interface due to the refractive index difference between the protein and precipitant solutions.

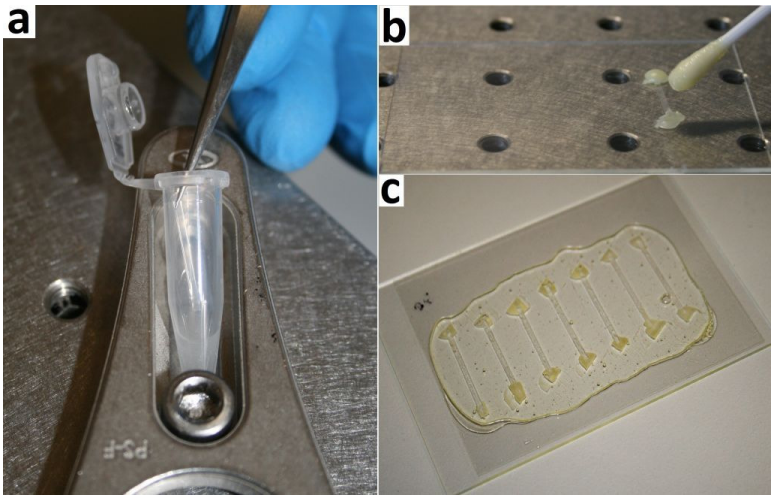


Figure 2.2: (a) Emulsion drops being filled in a rectangular capillary. To fill a capillary make a gentle contact between the capillary and the cream. (b) Application of VALAP (c) Air-tight seal with 5 min epoxy.

the required drop size and surface properties of microfluidic device and collect the drops only after the flows are stabilized.

Droplet Storage

The emulsion drops from the microfluidic device are collected in a 0.5 mL eppendorf tube. The emulsion creams to the top of the tube because, the density of the fluorinated oil ($\rho \approx 1.6$ g/ml) is greater than the aqueous medium ($\rho \approx 1$ g/ml). The emulsion drops are stored in rectangular capillaries, purchased from VitroCom. Figure 2.2a shows a rectangular capillary being filled with drops by capillary forces.

The dimensions of capillaries are chosen such that droplets are packed in a hexagonal monolayer. The emulsion filled capillaries are sealed with VALAP[29] (figure 2.2b), a mix of equal parts of VAseline, LAnolin and Paraffin wax with low melting temperature. After the application of VALAP, the extra wax is scrapped off using a razor blade. The capillaries are sandwiched between a microscope slide (75×50 mm) and a cover slip (48×65 mm) which are held together by 5 min epoxy, (purchased from Amazon Inc.) and are sealed air-tight. The sealed sample is shown in figure 2.2c. To minimize the difference between set and measured temperatures of the sample, we insulate the sample from the surroundings, using an air chamber as shown in figure 2.3. The sample and a same sized microscope (75×50 mm) slide are pressed together with the rectangular poly-siloxane (Product No. 3788T24, McMaster-Carr) washer, outer dimensions 75×50 mm, in between to create a sealed air chamber.

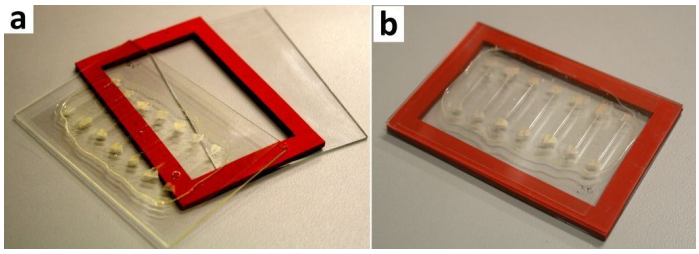


Figure 2.3: Creating an air-chamber using a rectangular washer.

The homogeneity in composition of protein and precipitant mixture in the drops is confirmed by measuring the cloud point. We have experimentally confirmed that all the drops are identical in chemical composition within a few percent of variation. See Appendix A for more discussion. We built a robotic stage, in association with Olin college, which can scan and acquire images of capillaries with an accuracy of 6 μm. The robotic stage is capable of acquiring images in both brightfield and fluorescent modes. Figures 2.4b and 2.4c show the illumination arms for brightfield and fluorescent modes. The stage is equipped with two Thermo Electric Coolers (TECs) to

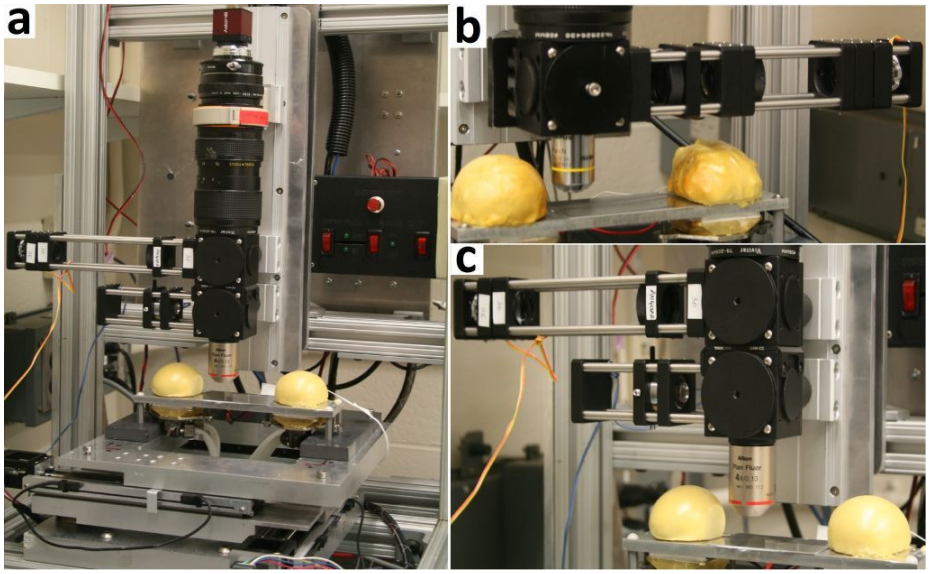


Figure 2.4: (a) Robotic stage for scanning and acquiring images, the stage is also equipped with TECs for temperature control. (b) Illumination arm for brightfield image acquisition. (c) Illumination arm for brightfield and fluorescent image acquisition.

control the temperature with a working range of -4 ℃ to 40 ℃ and are independently controlled. The robotic stage and the temperature are controlled using LabVIEW interface. The sample is mounted on the stage with the cover slide in direct contact with the thermal stage (figure 2.5), which further minimizes the difference between set and measured temperature of the sample.

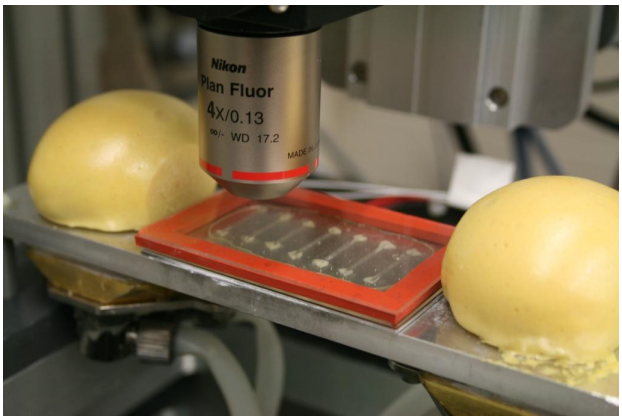


Figure 2.5: Sample mounted on the stage.

2.2.2 Image Analysis

In order to obtain good statistics, we scan multiple capillaries each containing large number of drops (> 2000) using the robotic stage. At the end of an experiment, we obtain large number of images containing crystallizing drops taken at regular intervals of time. To calculate the fraction of drops with no crystals, one needs to analyze large sets of images which is a laborious and time-taking process. Therefore we developed MATLAB programs to automate the process of detecting drops with and without crystals. The drop and crystal detection is performed in two steps, (1) Drop detection (2) crystal(s) detection in the corresponding drop. In the following section we briefly list the steps to detect drops and crystals.

Drop Detection

Using the following drop detection routines, more than 95% of the drops are usually detected. However, the efficiency of drop detection depends on the parameters chosen for the drop detection routines, illumination uniformity and drops in focus.

- *Adaptive histogram equalization* to correct for any non-uniform illumination and we used `adapthisteq`, a built in MATLAB function.

- *Inetnsity threshold* to binarize the image. The threshold value depends on the settings of the illumination source (usually LED) and camera. We used `imthresh`, a function written in MATLAB.

- *Erosion* is a morphological operation in image processing to erode away the boundary pixels of foreground objects i.e. objects containing white pixels in a binary image. This operation disconnects the clustered-objects such that drops and other individual features can be treated as independent objects. Erosion operation is performed using `imerode`, a built in MATLAB function.

- Removing small objects, noise pixels and objects touching to image boundary using `bwareaopen, bwclearborder`, built in MATLAB functions.

- *Convex hull* to obtain the pixels on the boundary of a drop or any feature. A *convex hull* is a polygon which completely encloses the object such that line joining any two pixels chosen from the object lies inside the polygon. To obtain convexhull, we used `regionprops`, a built in MATLAB function to obtain the properties of connected objects.

- The vertices of the convex hull polygon are fitted to a circle to obtain the circular edge of a drop. The fitting is performed using `fitCircle`, written in MATLAB.

Figure 2.6 shows the images at various steps in detecting the drops.

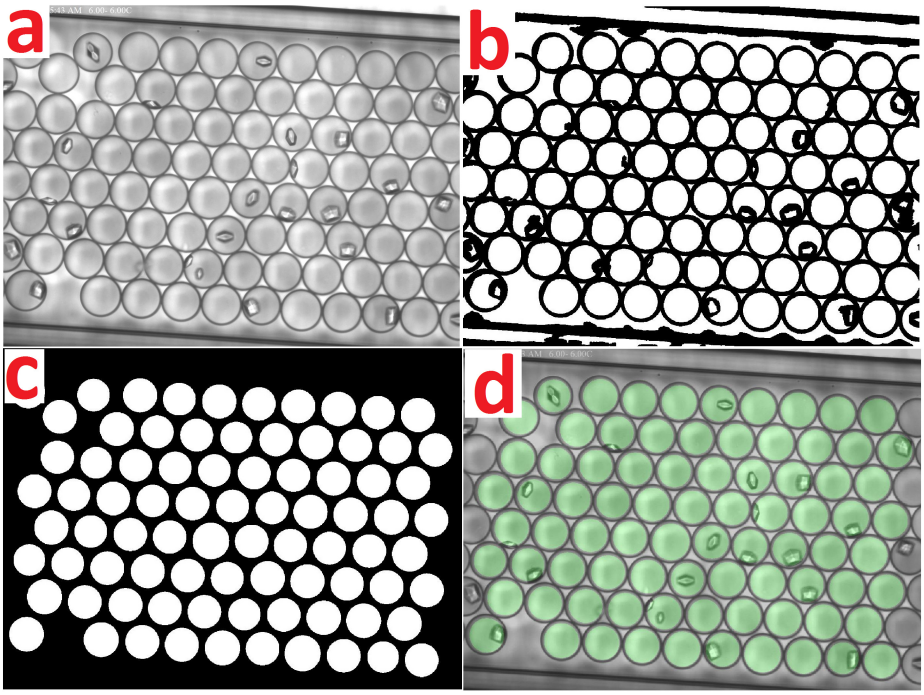


Figure 2.6: (a) Original image (b) Thresholded (c) Detected drops and (d) Detected drops overlaid on original image.

Crystal Detection

We used *Canny* edge detection method to detect the crystals. Figure 2.7a shows the detected crystal using Canny method. The Canny algorithm finds intensity gradients in an image. Usually only 70-80% of the crystals are detected. We found that the efficiency of the crystal detection depends on crystals formed at the edge of a drop (usually 30-40% of crystals are formed at the edge), crystals whose edges are not in focus (usually 10-20% of crystals are out of focus) and presence of unwanted objects (e.g. tiny droplets) in the sample. In such cases, we manually count the drops containing crystals.

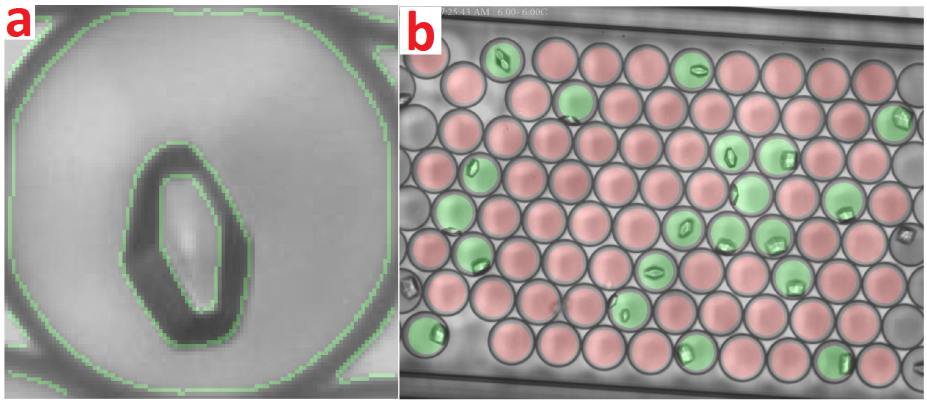


Figure 2.7: (a) Crystal detection using Canny edge detection. (b) Detected drops and crystal overlaid on the original image. Crystallized drops highlighted in green and non-crystallized drops highlighted in red.

Chapter 3

A Microfluidic based Technique to Measure Nucleation Rates

Abstract

We present a microfluidic technique to measure the nucleation rates of protein crystals using emulsion drops containing supersaturated protein solution. The technique involves nucleating large numbers of independent nano-liter drops at a constant temperature and counting the number of drops which have not nucleated as a function of time. Assuming nucleation is a Poisson process, at constant temperature the probability that an emulsion drop has no crystals decays exponentially with time and the decay constant is proportional to the nucleation rate and the volume of the drop. In the present work, we describe the technique in detail and present our analysis of the measured nucleation rates of lysozyme crystals within the context of Classical Nucleation Theory.

3.1 Introduction

Determining the structure of proteins is an important step in biology and biotechnology. Crystallization, which is necessary for the X-ray diffraction, remains a bottleneck in the structure determination of proteins. Many researchers around the world employ robotic arms to fill crystallization trays in order to exhaustively scan the protein-precipitant phase space in search of the crystallization conditions. The underlying assumption guiding this arduous routine is that crystallization is a phase transition and it is necessary to find the physical-chemical conditions of the equilibrium crystal phase. But crystallization is also an activated process and knowing the conditions for the equilibrium does not guarantee that crystallization will occur; the correct kinetic path must also be found. Crystallization involves nucleation and growth of the crystalline phase from the supersaturated solution phase. *Nucleation* is the process of formation of the nucleus, the smallest ordered form of the macroscopic crystalline phase spontaneously emerging from the supersaturated bulk phase. *Growth* is the subsequent growth of the nucleus into a macroscopic crystal. Classical Nucleation Theory (CNT) attempts to explain the thermodynamics of nucleation, but the applicability of CNT to protein molecules is an on-going field of research due to the large size and orientational interactions of protein molecules. In 1950s, Turnbull[9, 10, 11, 12, 13] characterized nucleation in supercooled liquid metals using an emulsion method. Since then, there have been efforts towards characterizing and understanding the nucleation in proteins by several research groups[17, 20, 21, 22, 25] around the world. Vekilov et al[17], developed a method to measure nucleation rates using sample volumes ≈ 1 μL and the Fraden[22] and Veesler[25] groups have extended the method to smaller volumes using microfluidics.

We adopted Turnbull's technique to measure nucleation rates of protein crystals

using drop volumes of < 1 nL. Turnbull's method involves supersaturating a large number of independent drops at constant temperature and counting the number of drops without crystals as a function of time. Under the assumption that nucleation is a Poisson process, the number of drops without crystals decays exponentially with time with the decay constant is proportional to the nucleation rate and the volume of the drop. We report the nucleation rates of lysozyme measured using the technique and present our analysis based on classical theory of nucleation.

3.2 Background

Nucleation is an activated process. Assuming that nucleation is a Poisson process, the probability that a sample has nucleated in time interval τ, is proportional to τ. Therefore, at constant nucleation rate J and sample volume v, the probability to nucleate is, $Jv\tau$. At times $t \gg \tau$, the probability, q that the sample has *not* nucleated is e^{-Jvt}. If we have N of such *identical* samples, then the fraction of samples which have *not* crystallized at time t is

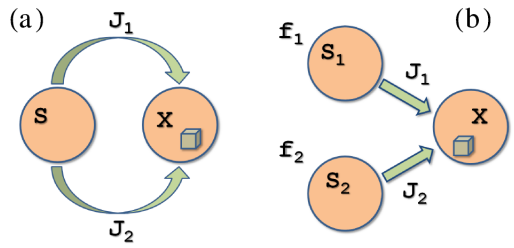


Figure 3.1: (a) Nucleation in a system with two nucleating pathways from a system of identical drops. (b) Nucleation in a system with two nucleating pathways from a system of two populations of drops.

$$f_\phi = \frac{N_\phi(t)}{N} = e^{-Jvt} \tag{3.1}$$

where, $N_\phi(t)$ is the number of samples which do not have crystals at time t. Consider a system of identical drops S, each of volume v, as shown in Figure 3.1a, in which two nucleation pathways exist (for example, homogeneous and heterogeneous pathways)

with rates, J_1 and J_2. If we assume that these two pathways are independent random process, then the chemical rate equation describing the process is,

$$\frac{\mathrm{dS}}{\mathrm{dt}} = -(J_1 + J_2)v\mathrm{S}$$

Solving the rate equation for the fraction of drops which have *not* crystallized at time t yields the following form with effective nucleation rate as the sum of existing nucleation rates J_1 and J_2.

$$f_\phi = e^{-(J_1+J_2)vt} \tag{3.2}$$

The fraction of non-crystallized drops is a single exponential despite there being multiple nucleation pathways.

In a system, where all the samples are *not identical* i.e., two(or more) populations of drops exist then the fraction of samples which have *not* crystallized is no more a single exponential decay. The system shown in figure 3.1b has two populations of samples S_1 and S_2 with fractions f_1 and f_2 which crystallize via nucleation pathways with rates J_1 and J_2 respectively. The chemical rate equations describing such a system are,

$$\frac{\mathrm{dS_1}}{\mathrm{dt}} = -J_1 v\mathrm{S_1}$$
$$\frac{\mathrm{dS_2}}{\mathrm{dt}} = -J_2 v\mathrm{S_2}$$

Solving the chemical rate equations for the fraction of samples which have *not* crystallized at time t gives,

$$f_\phi(t) = f_1 e^{-J_1 vt} + f_2 e^{-J_2 vt}$$

In general,

$$f_\phi(t) = \sum_i f_i e^{-J_i v t} \tag{3.3}$$

Where, f_i and J_i are the fraction of population i and the corresponding nucleation rate. Therefore, in a system with multiple populations of drops, the fraction of non-crystallized drops as a function of time is a multi-exponential with each population nucleating with the corresponding rate.

3.3 Experimental Methods and Data Analysis

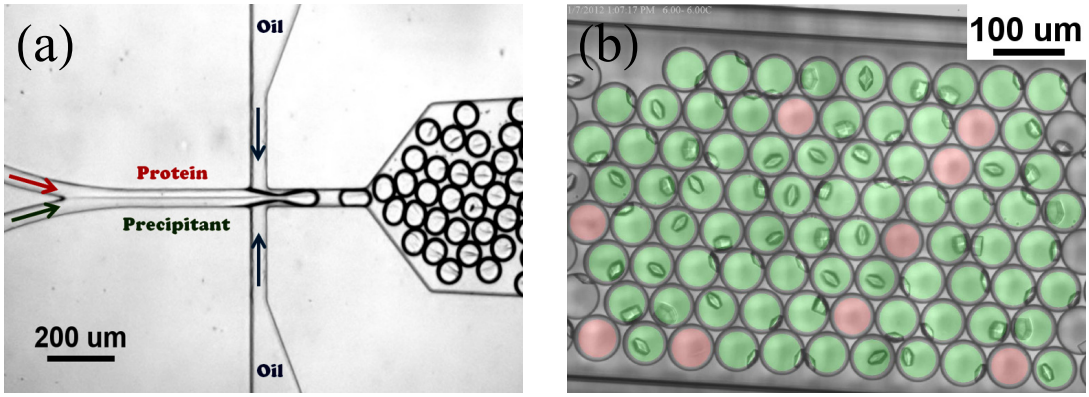


Figure 3.2: (a) Droplet generation using a co-flow microfluidic device, the protein and precipitant are mixed on-chip to avoid any nucleation before starting the experiment. The stream labelled 'Protein' contains lysozyme + 12.5% w/v PEG 8kD and the stream labelled 'Precipitant' contains 12.5% w/v PEG 8kD + 10% w/v NaCl. (b) Detected drops with crystals highlighted in green and without crystals highlighted in red.

We produce emulsion drops using a flow-focusing microfluidic device. Figure 3.2a shows a photograph of the droplet formation using a flow-focusing nozzle. To avoid nucleation before starting the experiment, the protein and precipitant are mixed on-chip just before making drops. For example, to produce emulsion drops containing 30 mg ml^{-1} lysozyme, 12.5% w/v PEG 8kD and 5% w/v NaCl, we followed the protocol given below.

1. 1:1 mix of 120 mg ml^{-1} lysozyme and 25% w/v PEG 8kD solutions. The final concentrations in the resulting mixture are 60 mg ml^{-1} lysozyme and 12.5% w/v PEG 8kD. Without any filtration, this solution is used for 'Protein' stream.

2. 1:1 mix of 20% w/v NaCl and 25% w/v PEG 8kD solutions. The final concentrations in the resulting mixture are 10% w/v NaCl and 12.5% w/v PEG 8kD. Without any filtration, this solution is used for 'Precipitant' stream.

3. For on-chip mixing, we used a co-flow microfluidic device shown in figure 3.2a and the 'Protein' and 'Precipitant' streams are pumped at equal flow rates, resulting in a 1:1 mix of streams and producing drops containing 30 mg ml^{-1} lysozyme, 12.5% w/v PEG 8kD and 5% w/v NaCl.

Important: The lysozyme/PEG mixture obtained in step 1, would *never* crystallize, but when lysozyme/PEG mixture (step 1) and PEG/NaCl (step 2) are mixed at equal proportions in a 500 μL eppendorf, lysozyme crystals form in seconds.

The homogeneity in composition of protein and precipitant mixture in the drops produced using a co-flow microfluidic device is confirmed by measuring the cloud point. We have experimentally confirmed that all the drops are identical in chemical composition within a few percent of variation. See Appendix A for more discussion. The emulsion drops are then loaded in a rectangular capillary and the ends of capillary are sealed with VALAP[29], a mix of equal parts of VAseline, LAnolin and Paraffin wax with low melting temperature. The capillaries are scanned at regular intervals using a home built robotic stage which can scan and acquire images of capillaries with an accuracy of 6μm. The stage is equipped with two Thermo Electric Coolers (TECs) to control the temperature with a working range of -4℃ to 40 ℃ and are independently controlled. The robotic stage and the temperature are controlled using LabVIEW interface. The images of capillaries scanned at regular intervals are processed semi-

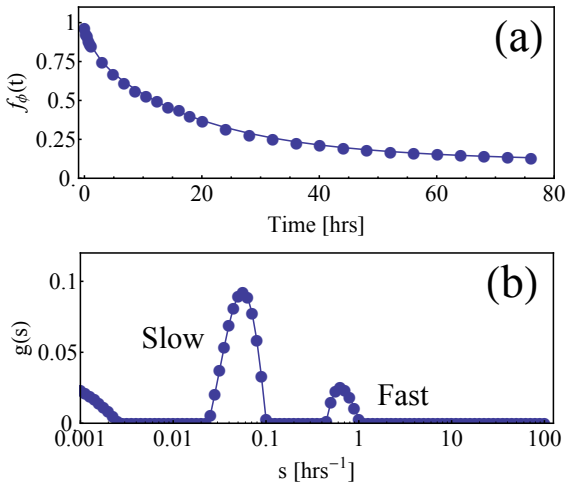


Figure 3.3: (a) Fraction of non-crystallized drops, $f_\phi(t)$ measured from the experiment. (b) Normalized decay distribution $g(s)$ obtained by inverse Laplace transforming $f_\phi(t)$.

automatically using MATLAB for droplet and crystal detection (figure 3.2b) to obtain $f_\phi(t)$. Figure 3.3a is the plot of fraction of drops with no crystals versus time, $f_\phi(t)$ and the fit. Mathematically, $f_\phi(t)$ is the Laplace transform of the decay time distribution, $g(s)$ (equation (3.4)). Hence, we obtained the fit by inverse Laplace transforming $f_\phi(t)$ using a CONTIN[31, 32] like algorithm written in MATLAB. Figure 3.3b shows the corresponding normalized decay time distribution, $g(s)$. See appendix B for more discussion on inverse Laplace transform. The calculation of the decay time distribution, $g(s)$, involves numerical computation of inverse Laplace transform of $f_\phi(t)$, which is an ill-posed problem. These kind of mathematical problems are solved using *Tikhonov regularization*[30] methods. The presence of two peaks in the decay distribution, $g(s)$, is not an artifact due to the choice of regularization parameter, α. Figure 3.4 is the plot of decay rate distribution, $g(s)$ as a function of regularization parameter and the two peaks are present over a wide range of α. Note that, the peak in $g(s)$, at small s is due to the inaccuracy in baseline measurement of $f_\phi(t)$. We

obtain two nucleation rates corresponding to the two peaks in $g(s)$.

$$f_\phi(t) = \mathcal{L}g(s) = \int_0^\infty g(s)\mathrm{e}^{-ts}\,\mathrm{d}s \tag{3.4}$$

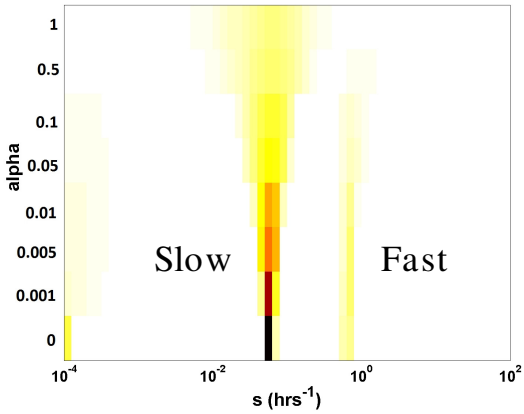


Figure 3.4: Presence of two peaks in the decay rate distribution as a function of regularization parameter. This confirms that the presence of two peaks is not due to an artifact caused by the choice of α.

When multiple decay modes exist in a process, ILT is the most general way of obtaining the decay mode distribution. However, only two rates exist in our system therefore we employed a much simpler model to obtain decay rates as proposed by Pound and La Mer[14]. The model, shown in figure 3.5, consists of large number of drops containing an average number, m, of nucleation sites per drop randomly distributed. These nucleation sites are due to the presence of *impurities* in the sample and serve as heterogeneous nucleation centers. Nucleation from the bulk solution of the volume of a drop, v, occurs at a rate, k_s and nucleation from a single nucleation site occurs at rate k_f. For example, a drop containing p nucleation sites nucleates with rate, $(k_s + pk_f)$, i.e. nucleation can occur from the bulk solution as well as from

the sites. The fraction of non-crystallized drops as a function of time is given by,

$$f_\phi(t) = e^{-m} e^{-k_s t} e^{m e^{-k_f t}} \tag{3.5}$$

When $m \ll 1$, equation (3.5) reduces to a two exponential process as shown in

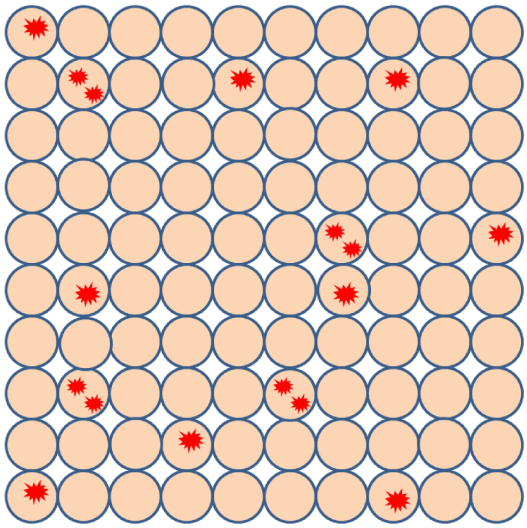


Figure 3.5: La Mer Model

figure 3.1b. In all that follows, we have obtained the 'fast', k_f and 'slow', $k_s = J_s v$ nucleation rates by fitting the experimentally measured fraction of drops with no crystals as a function of time to equation (3.5). One important distinction to address is the fact that the number of nucleation events occurring in a drop of volume v have different volume dependence for 'slow' and 'fast' nucleation. The number of nucleation events due to nucleation from bulk solution scales linearly with volume while the number of nucleation events due to nucleation from a site is independent of volume. We performed Monte-Carlo simulations using the La Mer model to account and estimate the errors in the developed method. Our simulation results suggest that one needs to perform at least 500 experiments to obtain nucleation statistics within 10% confidence range. When crystals grow slowly, the time at which a crystal is

large enough to be observed is significantly later than the time at which the crystal has nucleated. One can calculate the corrected fraction of drops with no crystals by accounting for the characteristic growth time, t_g due to slow growth rates as $f_\phi(t - t_g)$. Under the studied crystallization conditions, $t_g \approx 0.1$ hrs, which is less than the characteristic time scales for nucleation, $1/k_s$ and $1/k_f$. Therefore correcting the data for t_g did not improve our results. See Appendix G for more discussion on Monte-Carlo simulation results.

3.4 Results and Discussion

3.4.1 Measurement of Nucleation rates

According to Classical Nucleation Theory (CNT), the nucleation rate, $J(\sigma)$ is an exponential function of supersaturation, σ. The measured fraction of drops without crystals is fitted to equation (3.5) to obtain $J_s = k_s/v$, k_f and m. As explained in the previous section, k_s and k_f have different functional dependence on the volume of the drop therefore the measured nucleation rates, at a constant temperature, are fitted to the following equations.

$$J_s(\sigma) = \frac{k_s(\sigma)}{v} = A \cdot C \cdot \mathrm{e}^{-B/\sigma^2} \tag{3.6}$$

$$k_f(\sigma) = A' \mathrm{e}^{-B'/\sigma^2} \tag{3.7}$$

Where, $\sigma = \Delta\mu/k_B T \approx \ln(\mathrm{C}/\mathrm{C_S})$, C_S is the solubility of lysozyme and $A \cdot C(A')$, is the kinetic pre-factor associated with 'slow'('fast') process. $B(B')$ is related to the barrier height, ΔG^* of the 'slow'('fast') nucleation process and the interfacial tension

γ between the crystal nucleus and the solution phase as,

$$\frac{B}{\sigma^2} = \frac{\Delta G^*}{k_B T} = \frac{1}{k_B T} \frac{16\pi}{3} \frac{\Omega^2 \gamma^3}{\Delta \mu^2} \tag{3.8}$$

where, $\Omega = 3 \times 10^{-20}$ cm^3, is the lysozyme molecular volume. Figure 3.6 shows the plot of two sets of measured nucleation rates and average number, m, of nucleation sites per drop as a function of supersaturation. The measured value of $m < 1$, indicates that only a fraction of drops contain the impurities, similar to the two exponential process shown in figure 3.1b.

Note that, the definition of supersaturation, $\Delta\mu/k_B T = \ln(\mathrm{C/C_S})$ is an approximation of an ideal protein solution. To estimate the correction due to non-ideality of the protein solution, consider the expression[34] for $\Delta\mu$,

$$\sigma = \Delta\mu/k_B T = \ln(\mathrm{C/C_S}) + 2B_2 M(C - C_S) \tag{3.9}$$

where, B_2 is the second virial coefficient and M, is the molecular weight of lysozyme. We estimate[33] $B_2 = -4.85 \times 10^{-4}$ mL mol g^{-2}, an extrapolated value for the second virial coefficient for the data shown in figure 3.6. See Appendix C for more discussion on corrections for non-ideal solution behaviour. For example, the fitting parameters for the 'slow' nucleation data shown in figure 3.6 using $\sigma = \ln(\mathrm{C/C_S})$ are $A = 6.2 \times 10^7$ mg^{-1}s^{-1} and $B = 293.2$. We corrected σ using equation (3.9) taking the non-ideality of the protein solution into account and obtained $A = 1.2 \times 10^{12}$ mg^{-1}s^{-1} and $B = 354.1$. Due to the lack of measured second virial coefficients, we analyzed our data using $\Delta\mu/k_B T = \ln(\mathrm{C/C_S})$.

The 'slow' nucleation rates are fitted to equation (3.6). One can obtain details about the nucleation barrier, ΔG^*, and interfacial tension, γ, from B and the nature of nucleation and growth of the crystal from A. For brevity, we will present our

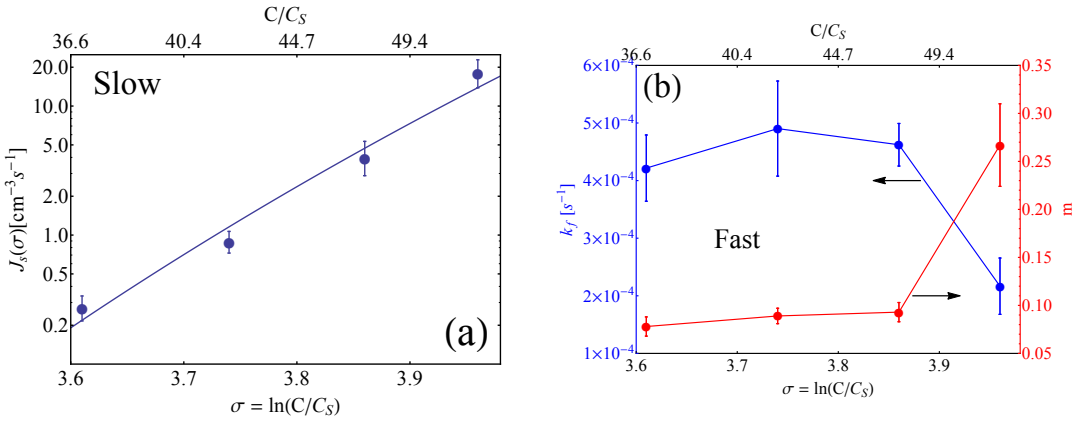


Figure 3.6: (a) Slow nucleation rates, $J_s = k_s/v$, corresponding to the rate per volume of solution (b) Fast nucleation rates, k_f, corresponding to the rate per impure site and average number, m, of nucleation sites per drop measured at 21-30 mg/ml Lysozyme, 5% w/v NaCl and 12.5% PEG 8kD at 9 ℃. Each data point is obtained from more than 2000 drops, each of volume < 1 nL.

analysis on nucleation barriers in this chapter and discuss the kinetic pre-factor in chapter 4. As mentioned earlier, the robotic stage has two thermoelectric coolers (TECs) which can be operated independently in the temperature range -4 ℃ to 40 ℃. By operating the TECs at two different temperatures, we scan the temperature vs. protein concentration phase space along the temperature axis. We measured the nucleation rates in the temperature range 7.2 ℃ to 12 ℃ at different concentrations of lysozyme. In the following sections we discuss the 'slow' and 'fast' nucleation processes separately in detail.

3.4.1.1 Slow Nucleation Process

Nucleation is an activated process and the rate at which nucleation occurs is proportional to the Boltzmann weight, $e^{-\Delta G^*/k_B T}$, where ΔG^* is the activation energy or the barrier height. Figure 3.7 are the plots of 'slow' nucleation rates, J_s vs. temperature at different supersaturations of lysozyme. The plots in figure 3.7 show that, as

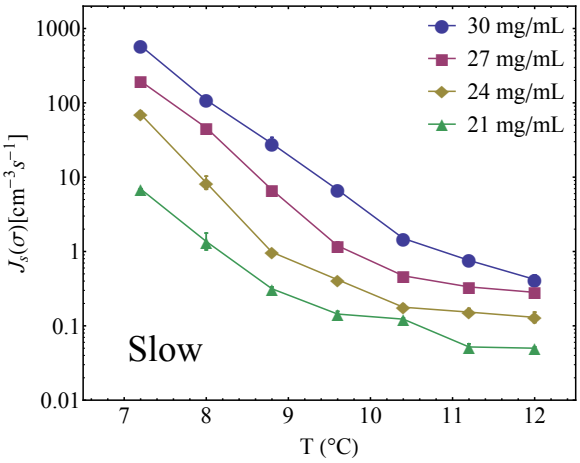


Figure 3.7: 'Slow' nucleation rates vs. temperature. The crystallization conditions are lysozyme, 12.5% w/v PEG 8kD, 5% w/v NaCl in 0.1M NaAc buffer at pH 4.8.

temperature is lowered nucleation rates increase, indicating an increase in Boltzmann weights. We calculate the barrier height, ΔG^*, using equation (3.8). Figure 3.8a is the plot of ΔG^* as a function of temperature. The measured nucleation rates in figure 3.7 increase by four orders of magnitude as temperature is lowered. We have also calculated ΔG^* for the corrected supersaturation, σ using equation (3.9). We observed qualitatively similar trend in ΔG^* as a function of temperature, however there is a $\simeq$ 50% increase in ΔG^* calculated using the corrected supersaturation (Appendix C). The nucleation rate, J_s (equation (3.6)), is a product of two terms, the kinetic pre-factor, $A \cdot C$ and the Boltzmann weight, e^{-B/σ^2} associated with the activation barrier. As temperature is lowered, fits of J_s to equation (3.6) reveal that the barrier increases and correspondingly the Boltzmann weight decreases by five orders of magnitude therefore to satisfy equation (3.6), kinetic pre-factor must increase nine orders of magnitude in order to fit the experimental rate measurements of figure 3.7. Thus the kinetic pre-factor and the activation energy work in opposite ways with the kinetic pre-factor dominating the nucleation rate. Inspection of figure 3.7 reveals that at constant temperature, the change in nucleation rates, with

respect to supersaturation, is larger at low temperatures than at high temperatures. This observation indicates that ΔG^* is larger at lower temperature than at higher temperature. Figure 3.8a and figure 3.8b are the plots of barrier heights, ΔG^* and

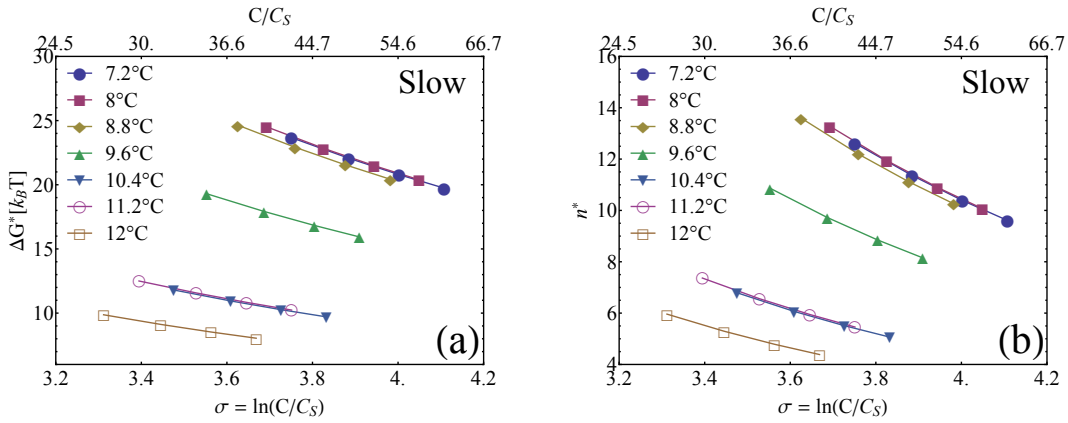


Figure 3.8: (a) ΔG^* vs. σ (b) n^* vs. σ measured at temperatures ranging from 7.2 ℃ to 12 ℃. Lines are drawn to guide the eye.

number of protein molecules in a critical cluster, n^* as a function of supersaturation at different temperatures. We have calculated n^* using the expression from CNT relating ΔG^* and n^* as $\Delta G^* = \frac{n^* \Delta \mu}{2}$. As expected from Classical Nucleation Theory,

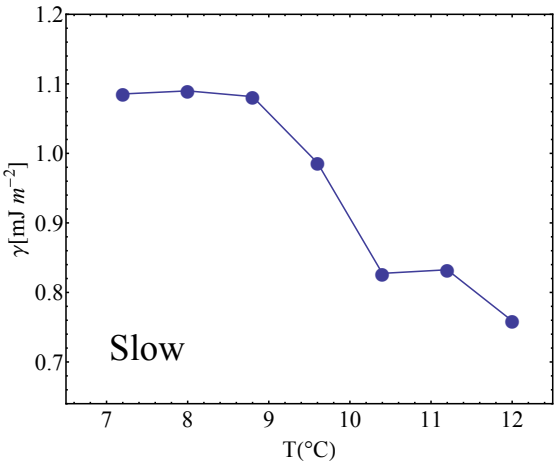


Figure 3.9: Interfacial tension vs. Temperature.

both decrease as supersaturation increases. As temperature increases, ΔG^* and n^* decrease and within the context of CNT via equation (3.6), this is attributed to the

decrease in interfacial tension, γ with an increase in temperature. Figure 3.9 shows the fitted decrease in interfacial tension as temperature increases. The range of interfacial tensions measured, $\simeq 0.7 - 1.1$ mJ m^{-2} are consistent with the values reported in literature[17, 22].

3.4.1.2 Fast Nucleation Process

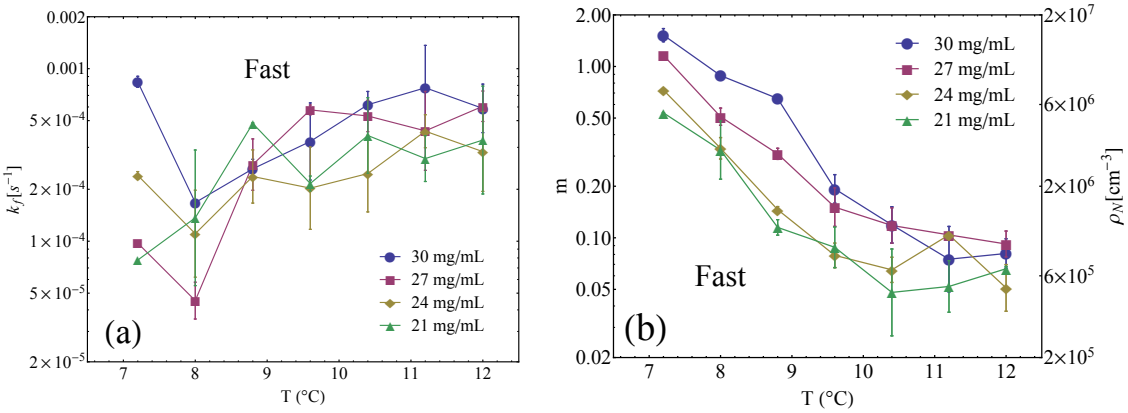


Figure 3.10: (a) 'Fast' nucleation rates vs. temperature. (b) Average number of nucleation sites, m and the number density, ρ_N, of nucleation sites vs. temperature. The crystallization conditions are Lysozyme, 12.5% w/v PEG 8kD, 5% w/v NaCl in 0.1M NaAc buffer at pH 4.8.

Figure 3.10a shows 'fast' nucleation rates measured as a function of temperature at different concentrations of lysozyme. The nucleation rates do not show any systematic variation with respect to supersaturation and temperature. Since, $\Delta G^* \propto \frac{\partial \ln J}{\partial \sigma}$[36] we conclude that the nucleation barrier, $\Delta G^* = 0$ for the 'fast' nucleation process. As temperature is lowered the average number of nucleation sites per drop, m increases (figure 3.10b) which suggests creation of nucleation sites at lower temperatures. We speculate that these nucleation sites are disordered aggregates of lysozyme monomers.

3.4.2 Validation of Technique

Galkin and Vekilov[17], developed a drop based method to measure nucleation rates of protein crystals which involves nucleating the supersaturated protein solution for a quench period of Δt at T_1 and subsequently growing the nucleated crystals at $T_2 > T_1$. Under the assumption that the time scale for growth is much longer than the quench period, only nucleation occurs at temperature T_1 while at temperature T_2, no nucleation occurs but growth of the nucleated crystals is favorable, i.e. nucleation and growth are decoupled. The average number of crystals formed per drop, $\langle N \rangle = N_{hetero} + Jv\Delta t$, where N_{hetero} is the number of crystals per drop measured at $\Delta t = 0$, J is the nucleation rate and v is the volume of the drop. Galkin and Vekilov attributed the intercept to heterogeneous nucleation occurring on impurities present in the drops and attributed the crystals that formed at later time, $\Delta t > 0$, to be due to homogeneous nucleation. Note that the supersaturation does not decrease significantly due to the nucleated crystals from heterogeneous nuclei to suppress further nucleation from bulk solution because of the slow crystal growth. Galkin and Vekilov claim that heterogeneous nucleation primarily occurs at the droplet-oil interface, but we speculate dis-ordered lysozyme aggregates which are formed at higher supersaturations serve as heterogeneous nucleation centers. See Appendix D for more discussion on heterogeneous nucleation at the droplet-oil interface. Galkin and Vekilov[18] also observed an increase in N_{hetero} with increasing supersaturation, consistent with our speculation of the formation of lysozyme aggregates at higher supersaturations.

We have performed nucleation rate measurements at crystallization conditions similar to those of Galkin and Vekilov. Once again, we fitted the fraction of drops which do not have crystals to equation (3.5) and obtained two nucleation rates, namely, 'fast' and 'slow'. The 'slow' nucleation rate is in agreement with the reported nucleation rates. Figure 3.11 shows the nucleation rate vs. supersaturation

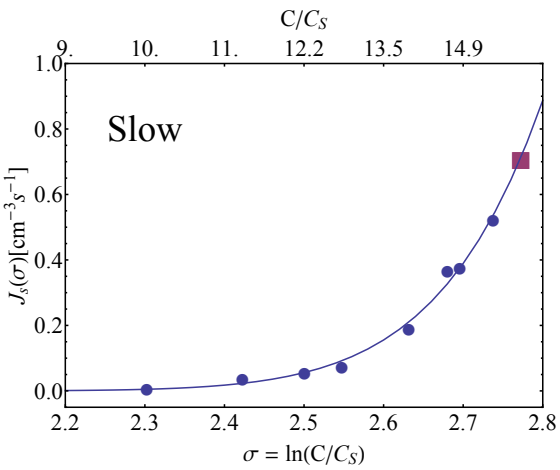


Figure 3.11: Measurement of Nucleation rate under conditions mentioned in reference[17]. The crystallization conditions are Lysozyme, 2.5% w/v NaCl in 0.05 M NaAc at pH 4.5, T = 12.6 ℃. ■ is the nucleation rate obtained using the method described in this paper.

obtained by Galkin and Vekilov. The solid points (•) are the measurements by Galkin and Vekilov and the continuous line is the fit to nucleation rate described by Classical Nucleation Theory. The solid point '■' is the 'slow' nucleation rate measurement performed using Turnbull's method of measuring $f_\phi(t)$, confirming the validity of the method. The measured average number, m of impurities per sample is $\simeq 0.12$ per drop, confirming the argument that the offset in Vekilov experiments is indeed due to a *few* impurities per drop. Galkin and Vekilov[18] estimated the average number N_{hetero} of impurities per drop to be 0.2, which is in agreement with the measured value m from our experiments.

3.5 Conclusion

We have developed an emulsion based technique for the measurement of nucleation rates and applied the technique to measure the nucleation rates of lysozyme crystals. The emulsion drops are produced using a flow-focusing microfluidic device. We

observe two nucleation rates within the drop population. The measured nucleation rates, namely 'slow' and 'fast', are analyzed according to Classical Nucleation Theory (CNT). The Pound and La Mer model[14] ascribes the fast nucleation rate to impurities present in a small subset of the drop population, but makes no prediction on whether or not the slow nucleation rate is due to homogeneous or heterogeneous nucleation. A study of the kinetic pre-factor, presented in chapter 4, will address this question. From the CNT analysis, we extracted the barrier heights and size of critical nuclei associated with the nucleation processes. As the temperature is lowered, we observed an increase in the nucleation rates. One possibility to account for the increase in nucleation rates is that the nucleation barrier is lowered with decreasing temperature. However, counter-intuitively, the CNT analysis reveals that the barrier heights increase as the temperature is lowered. A detailed analysis[8] of the kinetic pre-factor suggests that, the pre-factor increases 9 orders of magnitude as the temperature is lowered by 5 ℃, which was surprising to us as often the kinetic pre-factor is considered to be a weak function of temperature[37]. Even though the barrier heights increase as the temperature is lowered, the increase in the number of nucleation events, resulting from exponential increase in kinetic pre-factor, account for the increase in the nucleation rates. We have also successfully validated our technique by comparing it to the experiments of Galkin and Vekilov[17].

Chapter 4

Determining the Nature of Nucleation: Homogeneous or Heterogeneous

Abstract

Classical Nucleation Theory (CNT) predicts that the nucleation rate varies exponentially with supersaturation, with the exponent characterizing the activation energy of nucleation and the pre-exponential term (kinetic pre-factor) describing the growth kinetics of a nucleus. We have measured nucleation rates at different temperatures and supersaturations of lysozyme using an emulsion based technique described in chapter 3. We obtained the activation energy and the kinetic pre-factor by analyzing the measured nucleation rates according to the predictions of CNT. In this chapter, we have presented our analysis of the kinetic pre-factors as a means to determine the nature of nucleation, as suggested by Sear[42].

4.1 Introduction

Classical Nucleation Theory (CNT) describes the kinetics of nucleation and predicts the functional form of nucleation rate in terms of the thermodynamic quantities involved, such as supersaturation and temperature. *Nucleation* is a stochastic process by which embryos of the stable phase appear. The embryo or the nucleus, is the smallest ordered form of the macroscopic crystalline phase that spontaneously emerges from the supersaturated bulk liquid phase. Nucleation is an activated process and the activation barrier arises as a result of the competition between the energy gain in transferring the molecules from solution to the interior of the nucleus and the energy cost in creating the interface between the solution and the nucleus. Therefore, the total change in free energy in creating a nucleus from bulk phase is given by,

$$\Delta G = \Delta G_v + \Delta G_s$$

Where $\Delta G_v = -\frac{4\pi r^3}{3\Omega}\Delta\mu$ is the volume energy and $\Delta G_s = 4\pi r^2\gamma$ is the surface energy of a spherical cluster of radius r. Ω is the molecular volume, $\Delta\mu$ is the difference in chemical potential between a molecule in the solution phase and the crystalline phase and γ is the interfacial tension between the newly formed solid phase and the bulk solution phase. Figure 4.1a, shows the development of the barrier, ΔG^*, as a result of competition between the volume and the surface free energies of a cluster.

$$\Delta G = -\frac{4\pi r^3}{3\Omega}\Delta\mu + 4\pi r^2\gamma \tag{4.1}$$

We obtain the activation energy, ΔG^* by setting $\frac{\partial \Delta G}{\partial r}\big|_{r^*} = 0$,

$$\Delta G^* = \frac{16\pi}{3}\frac{\Omega^2\gamma^3}{\Delta\mu^2}$$

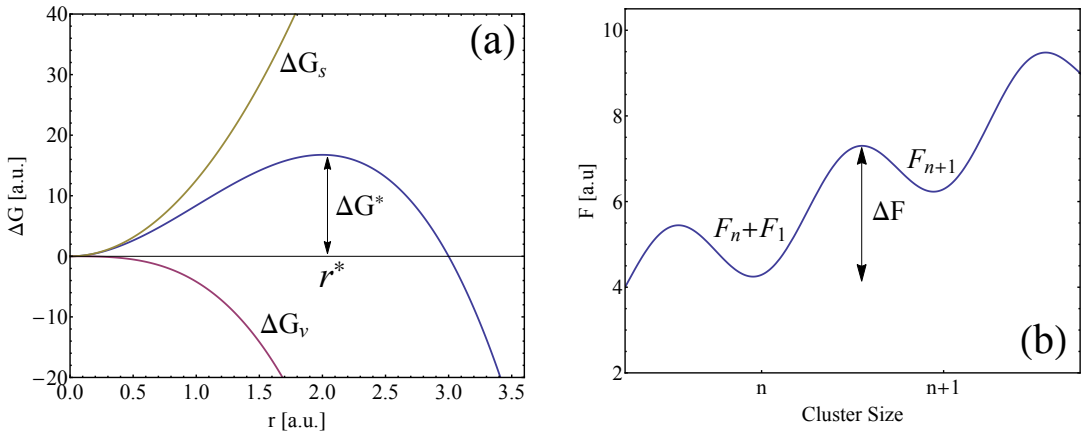


Figure 4.1: (a) Energy barrier, ΔG^*, for nucleation which is a result of competition between volume free energy, ΔG_v and surface free energy, ΔG_s. (b) Activation energy required, ΔF, for the growth of a cluster of size n to $n+1$.

where, r^* is the radius of the critical cluster. Note that, in heterogeneous nucleation because of the presence of a surface of lower energy for the protein, the effective surface tension is reduced, but a barrier still remains. The rate at which nucleation occurs is proportional to the Boltzmann weight associated with the activation energy as is given by,

$$J \propto e^{-\Delta G^*/k_B T} \tag{4.2}$$

The proportionality factor in equation (4.2) is $\rho_N Z j$, known as the kinetic pre-factor. Where ρ_N is the number density of nucleation sites, Z is the Zeldovich factor, which is the probability for the critical nucleus to grow, and j is the rate at which individual molecules attach to a cluster. Thus the rate of nucleation is given by,

$$J = \rho_N Z j e^{-\Delta G^*/k_B T} \tag{4.3}$$

The rate at which molecules add on to the nucleus j can further be expressed as,

$$j \simeq 4\pi\rho D R^* e^{-\Delta F/k_B T}$$

Where, ρ is the number density of monomers, D is the diffusion constant of monomers, R^* is the radius of the critical cluster and ΔF is the activation energy (figure 4.1b) for the addition of a monomer to an existing cluster. The physical understanding of j is as follows. $4\pi\rho DR^*$ is the diffusion limited in-flux of protein molecules towards a critical cluster of radius R^* and $e^{-\Delta F/k_BT}$ is the probability that a collision of a monomer and a critical nucleus will result in the monomer joining the cluster. We assumed that ΔF to be independent of the nature of nucleation since ΔF is the change in free energy associated with a molecular attachment to a cluster of protein molecules. The Zeldovich factor, $Z \simeq (n^*)^{-2/3}$[37], is a slow varying function of n^*, therefore the following estimates for homogeneous and heterogeneous nucleation rates are calculated with $Z = 0.2$ even though the number of molecules in a critical cluster for homogeneous and heterogeneous nucleation is different. In the case of homogeneous nucleation, $\rho_N = \rho$ since nucleation can occur from any of the individual molecules but in case of heterogeneous nucleation, the number density of nucleation sites can be much smaller, $\rho_N \ll \rho$. Hence the nucleation rates for homogeneous and heterogeneous nucleation rates are,

$$J_{hom} \simeq 4\pi\rho^2 DR^* Z e^{-\Delta F/k_BT} e^{-\Delta G^*_{hom}/k_BT} \tag{4.4a}$$

$$J_{het} \simeq 4\pi\rho_N \rho DR^* Z e^{-\Delta F/k_BT} e^{-\Delta G^*_{het}/k_BT} \tag{4.4b}$$

We can estimate the value of kinetic pre-factor as follows. For typical lysozyme (molecular weight $\sim$ 14,700 gm mol^{-1}) crystallization trials, the concentration of lysozyme is $\sim$ 30 mg ml^{-1} which corresponds to $\rho = 10^{18}$ cm^{-3}, $D = 10^{-6}$ cm^{2}s^{-1}. For a critical cluster of 12 molecules, $R^* = 3.5$ nm and $Z \simeq (n^*)^{-2/3} = 0.2$.

$$J_{hom}[\mathrm{cm}^{-3}\mathrm{s}^{-1}] \simeq 10^{23} e^{-\Delta F/k_BT} e^{-\Delta G^*_{hom}/k_BT} \tag{4.5a}$$

$$J_{het}[\mathrm{cm}^{-3}\mathrm{s}^{-1}] \simeq 10^{5} \rho_N e^{-\Delta F/k_BT} e^{-\Delta G^*_{het}/k_BT} \tag{4.5b}$$

Therefore in order to characterize the nature of nucleation, one needs to experimentally obtain the kinetic pre-factors.

4.2 Results and Discussion

We have measured the nucleation rates of lysozyme crystallization using an emulsion based method described in chapter 3. The method involves counting the fraction of drops with no crystals, f_ϕ as a function of time. The measured $f_\phi(t)$ is analyzed according to a model proposed by Pound and La Mer[14]. The model, shown in figure 4.2, consists of large number of drops containing an average number, m, of nucleation sites per drop that are randomly distributed. These nucleation sites are due to the presence of *impurities* in the sample and serve as heterogeneous nucleation centers. Therefore, by definition the 'fast' process is *heterogeneous*. Nucleation from bulk solution occurs at a rate, k_s and nucleation from a single heterogeneous nucleation site occurs at rate k_f. For example, a drop containing p nucleation sites nucleates with rate, $(k_s + pk_f)$, i.e. nucleation can occur from the bulk solution as well as from the sites. The fraction of non-crystallized drops as a function of time is given by,

$$f_\phi(t) = e^{-m} e^{-k_s t} e^{m e^{-k_f t}} \tag{4.6}$$

In all that follows, we have obtained the 'fast', k_f and 'slow', $k_s = J_s v$ nucleation rates by fitting the experimentally measured fraction of drops with no crystals as a function of time to equation (4.6). Figure 4.3 shows the measured nucleation rates, k_s and k_f as function of temperature at different supersaturations. One important distinction to note is that k_s is the number of nucleation events per time per drop and so scales linearly with drop volume, v, while k_f is the nucleation rate per heterogeneous site and therefore is independent of drop volume. Therefore, equation (4.3) is re-written

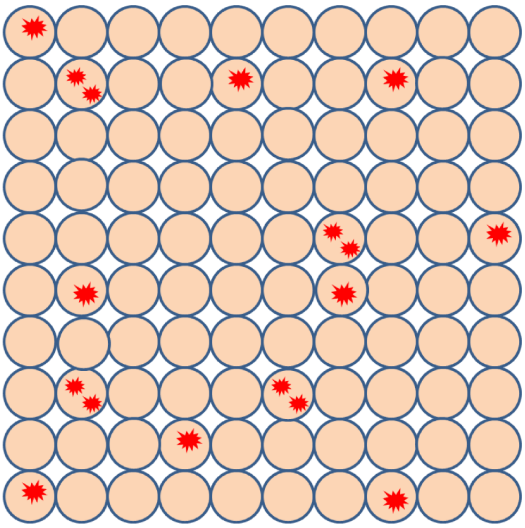


Figure 4.2: Schematic of La Mer Model

for k_s and k_f as,

$$J_s(\sigma) = \frac{k_s(\sigma)}{v} \approx 4\pi\rho_N\rho D R^* Z \mathrm{e}^{-\Delta F/k_B T}\mathrm{e}^{-\Delta G_s^*/k_B T} \tag{4.7a}$$

$$k_f(\sigma) \approx 4\pi\rho D R^* Z \mathrm{e}^{-\Delta F/k_B T}\mathrm{e}^{-\Delta G_f^*/k_B T} \tag{4.7b}$$

Experimentally, we obtained the pre-factor by fitting the nucleation rates at different supersaturations to equations (4.7) expressed in a simpler form as,

$$J_s(\sigma) = A \cdot C \cdot \mathrm{e}^{-B/\sigma^2} \tag{4.8a}$$

$$k_f(\sigma) = A'\mathrm{e}^{-B'/\sigma^2} \tag{4.8b}$$

Where, $A(A')$ and $B(B')$ are fitting parameters, C is the concentration of protein and $\sigma = \Delta\mu/k_B T$ is the supersaturation. C_S is the solubility of the protein. Approximating the protein solution to ideal solution, we set the chemical potential $\sigma = \Delta\mu/k_B T = \ln(\mathrm{C}/\mathrm{C_S})$.

The plots shown in figure 4.4 are the pre-factors measured at different temperatures. We analyzed (discussed in sections 4.2.1 and 4.2.2) the kinetic pre-factors

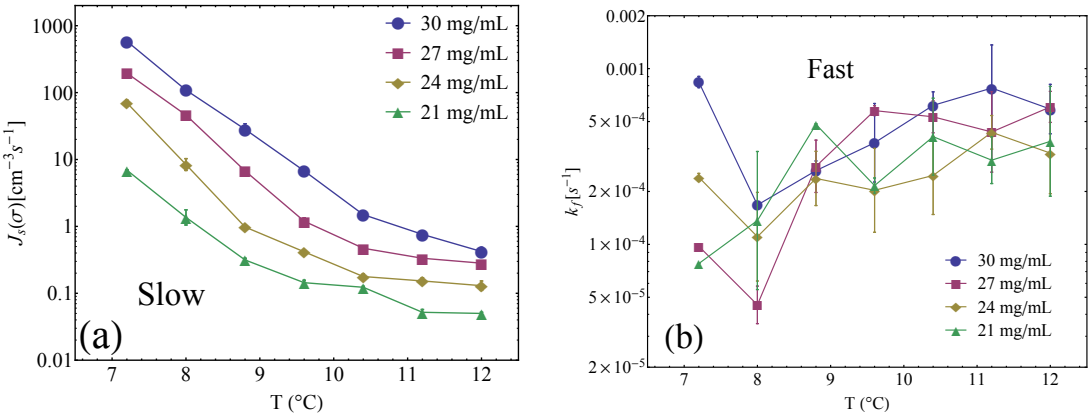


Figure 4.3: (a) Slow and (b) Fast Nucleation Rates as a function of temperature at different supersaturations. The crystallization conditions are Lysozyme, 12.5% w/v PEG 8kD, 5% w/v NaCl in 0.1M NaAc buffer at pH 4.8.

in the context of CNT and found that both the observed nucleation processes are *heterogeneous* in nature. The impurities causing heterogeneous nucleation could be dust particles, aggregates of protein of the interest or of other proteins. In the 'slow' process, the pre-factors vary 9 orders of magnitude from $\sim 10^{-2} - 10^{7}$ cm^{-3}, over 5 ℃. The variation could result from variation in ρ_N or $\Delta F/k_BT$ or both. Classical Nucleation Theory assumes that the form of the pre- and post-critical nucleus is the same. Therefore the growth kinetics of a critical cluster should be identical. Logically if the variation in the kinetic pre-factor id due to $\Delta F/k_BT$, then a similar effect should be observed in the growth rates of the post-critical cluster because ΔF, the activation barrier for the addition of a molecule to an existing cluster, is the rate limiting step in determining the growth rate of a cluster. We measured growth rates (figure 4.5b) of crystals at the same crystallization conditions as the nucleation rate experiments were performed and the growth rates do not vary significantly with temperature suggesting that ΔF is independent of temperature and the variation in pre-factor arises completely due to an increase in the number density of nucleation sites as temperature is lowered. However, another possibility is that the assump-

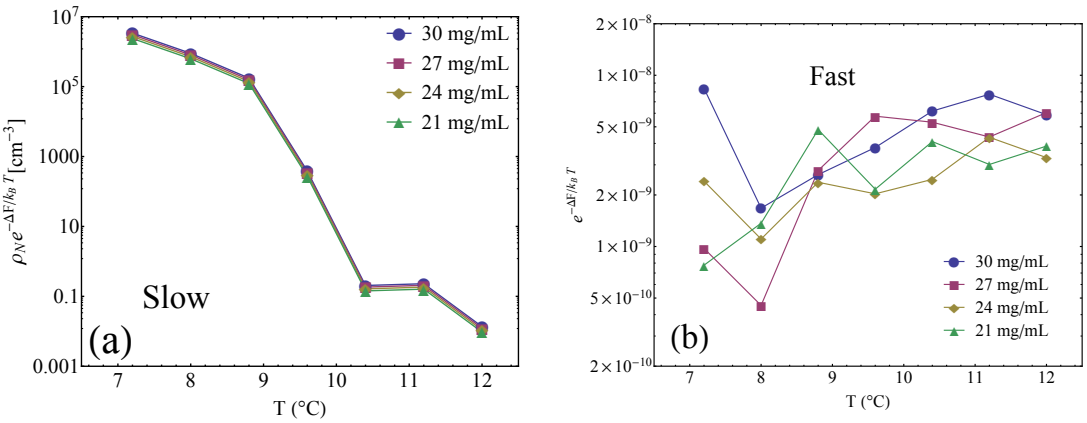


Figure 4.4: Pre-factors obtained by fitting the measured nucleation rates to equations (4.8) at every temperature. The crystallization conditions are Lysozyme, 12.5% w/v PEG 8kD, 5% w/v NaCl in 0.1M NaAc buffer at pH 4.8.

tions of the CNT are wrong and that the growth kinetics of the pre-critical nucleus and post-critical nucleus are different, which is very likely to be true. In such case, there could be a large ΔF for addition a monomer to a pre-critical nucleus and a $\Delta F \approx 0$ for addition to a post-critical nucleus. Although, we observe two nucleation pathways leading to crystallization, the growth kinetics of post-critical nucleus are identical in both nucleation processes. The increase in the number density of sites does not necessarily happen by creating new aggregates but can also occur through an increase in the number of active sites on the aggregates as shown schematically in figure 4.5a. Nucleation can occur at any site on the aggregate, but once a crystal has nucleated further nucleation is suppressed due to small sample volume($\simeq$1 nL) because monomers diffuse from one end of the drop to the growing crystal faster than the rate at which new crystals are nucleated. Therefore, once one crystal has nucleated the concentration of the entire drop rapidly decreases thereby suppressing any subsequent nucleation. This negative feedback assures that each drop will have at most one crystal. In the following sections, we discuss the nature of impurities that account for the 'slow' and 'fast' nucleation processes in separate sections.

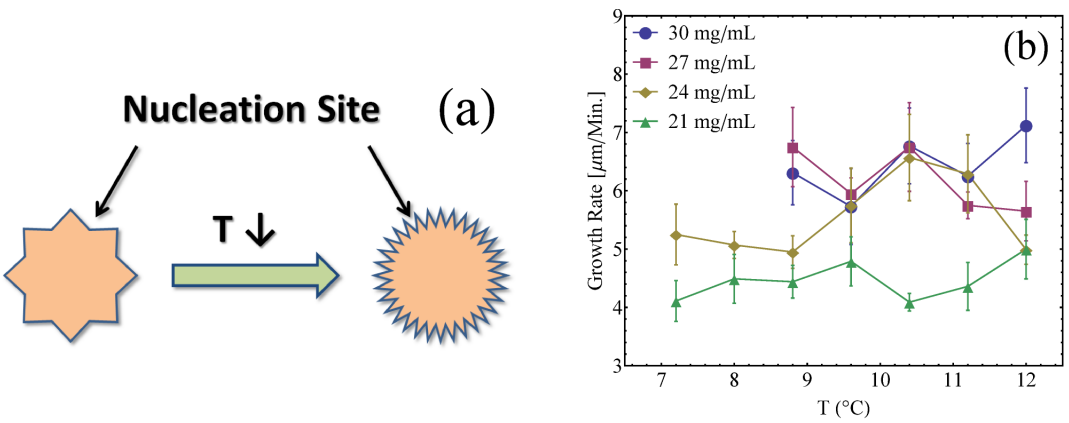


Figure 4.5: (a) Creation/Activation of nucleation sites by lowering the temperature. (b) Measured growth rates vs. Temperature. The growth rates do not considerably change as function of temperature suggesting that ΔF is constant.

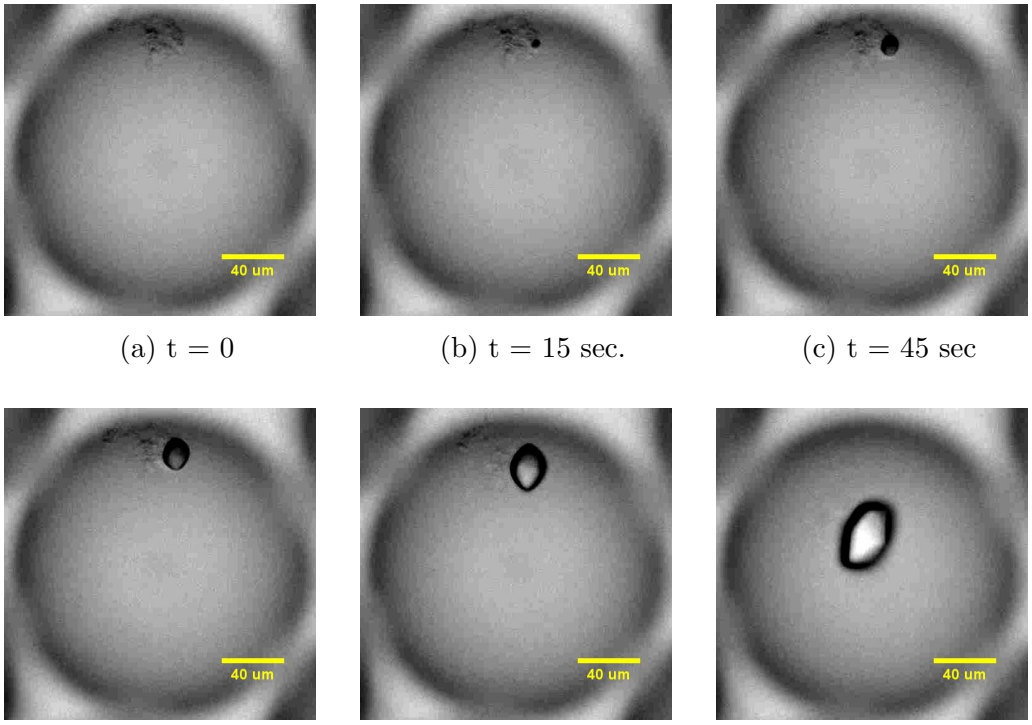
4.2.1 Impurities causing Slow Process

The impurities causing the slow process are protein aggregates and are formed due to the depletion interaction induced by Poly-Ethylene Glycol (PEG) molecules and the size of the aggregates range from $0.1-10$ μm. The presence of the aggregates was confirmed by Dynamic Light Scattering (DLS) measurements and also using Differential Interference Contrast (DIC) and fluorescence microscopy techniques. See Appendix E for more discussion on characterizing protein aggregates. In all our experiments, we used a $1:1$ mixture of $80-120$ mg ml^{-1} protein and 25% w/v PEG 8kD to comprise the 'Protein' stream (discussed in section 3.3). Soon after equal proportions of protein and PEG solutions are mixed in 500 μL eppendorf, protein molecules in some regions of the lysozyme/PEG solution are strongly depleted to form the lysozyme aggregates due to the inhomogeneous mixing. After attaining a homogeneous mixture of lysozyme/PEG solutions, the mixture is filtered through 0.22 μm cellulose acetate filters to remove the aggregates. After filtration, the aggregates do not form, even after 48 hrs, indicating that they are *not* in a thermodynamic equilibrium with the lysozyme monomers but are in a dynamically arrested state. This observation sug-

gests that the depletion interaction is not strong enough to form lysozyme aggregates once the equilibrium concentrations are attained. We also confirmed the absence of protein aggregates using DLS and optical microscopy. As the temperature is lowered, the aggregates do not undergo any measurable growth. The lysozyme concentration in the filtered solutions did not change by a measurable amount compared to the unfiltered samples indicating that the total mass of the protein in the aggregates is very small. Using the filtered 'Protein' stream, we produced the emulsion drops and performed nucleation rate measurements on the filtered samples.

Important: All experimental results reported in this research work are obtained with un-filtered 'Protein' stream, except for the following measurement.

Once again, we obtain two nucleation rates, 'slow' and 'fast' however, the 'slow' rates are at least 4 times smaller than the 'slow' rates obtained from unfiltered samples. This measurement independently confirms that 'slow' nucleation is *heterogeneous*, as one cannot explain the decrease in homogeneous nucleation rates by filtration. Also, the average number of impurities per sample, m resulting in 'fast' nucleation in the filtered samples does not change considerably from the unfiltered samples. This suggests that the impurities causing 'fast' nucleation are different in size from the large, visible protein aggregates and are not filterable. We observe that the nucleation occurs on one of the protein aggregate and as the crystal grows, the protein aggregate is consumed. Figure 4.6 shows the nucleation and growth of a crystal from a protein aggregate. We have noted the crystals that appear at later times ($\geq$ 2 hrs) are mostly from the protein aggregates, which are the source of 'slow' nucleation process observed in our experiments. Figure 4.4a shows the measured pre-factors as a function of temperature at different concentrations of lysozyme. Note that continuous increase in the pre-factor as temperature lowered, which we suggest is a result of the activation of more nucleation sites as depicted in figure 4.5a.



(a) t = 0 (b) t = 15 sec. (c) t = 45 sec

(d) t = 75 sec (e) t = 120 sec (f) t = 300 sec

Figure 4.6: Nucleation and Growth of a crystal from a protein dense aggregate/gel. The aggregate dissolves as the crystal grows.

We also calculated the number density, ρ_N, of impurities causing the slow process as a function of temperature using an estimated $\Delta F \simeq 19 k_B T$ (estimation explained in section 4.2.2). Figure 4.7 is the plot of ρ_N as a function of temperature calculated at different supersaturations of lysozyme. At all temperatures, $\rho_N \ll \rho$, which is consistent with the claim that the slow process is *heterogeneous*.

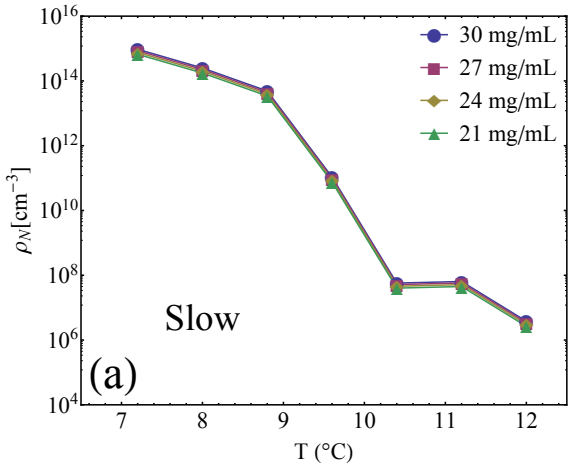


Figure 4.7: Number density, ρ_N, of nucleation sites as a function of temperature calculated at different supersaturations of lysozyme. Calculated assuming $\Delta F = 19 k_B T$.

4.2.2 Impurities causing Fast Process

Figure 4.3b shows the 'fast' nucleation rates measured as a function of temperature at different concentrations of lysozyme. The nucleation rates don't show any systematic and significant variation with respect to supersaturation and temperature. Therefore we conclude that the nucleation barrier, ΔG^* associated with the 'fast' nucleation process is zero. One can estimate the number density of impurities by measuring the average number of impurities per drop, m, resulting in fast nucleation as a function of volume. Figure 4.8a shows m as a function of volume of the drop measured at two different crystallization conditions. The plot indicates that the number of impurities

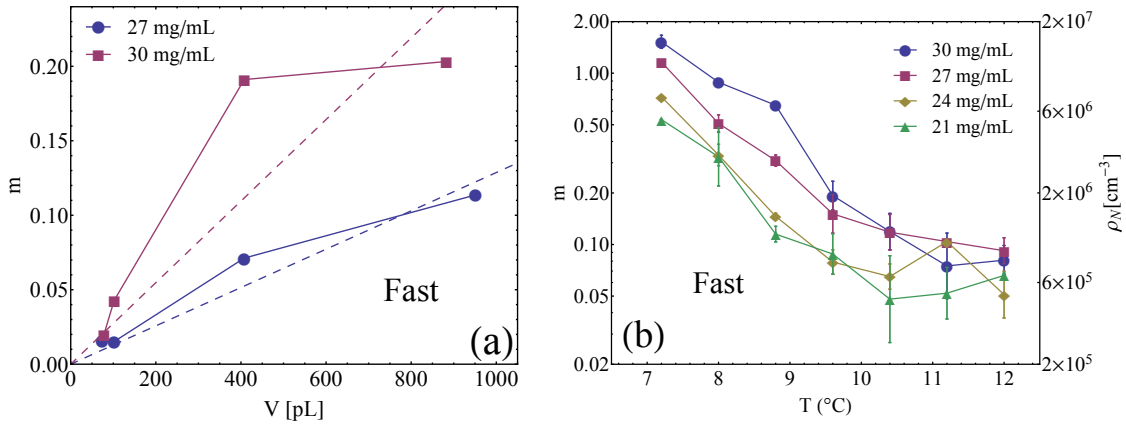


Figure 4.8: (a) Average number of nucleation sites per drop, m as a function of volume at 9 ℃. The measured nucleation rates are $J_s = 17.46 \pm 3.41$ cm^{-3}s^{-1}, $k_f = (0.17 \pm 0.02) \times 10^{-3}$ s^{-1} for 30 mg ml^{-1} and $J_s = 3.41 \pm 0.7$ cm^{-3}s^{-1}, ($k_f = 0.77 \pm 0.3) \times 10^{-3}$ s^{-1} for 27 mg ml^{-1}. (b) Average number, m of nucleation sites per drop and number density, ρ_N, of nucleation sites as a function of temperature at different supersaturations.

per drop is less than 1 and vanishes when the volume of the drop is sufficiently small. When the number of impurities per drop of volume V is $\simeq 1$, one would expect the probability to find an impurity in volume $v < V$ decreases linearly to zero as volume, v, decreases. The dotted lines represent the expected behaviour in m as volume decreases. The kinetic pre-factor for the fast nucleation process from equation (4.7)b is $\rho D R^* Z e^{-\Delta F/kT}$ and the measured value of pre-factor is $\approx 0.5 \times 10^{-3}$ s^{-1}. Using typical estimates of $\rho = 10^{18}$ cm^{-3}, $D = 10^{-6}$ cm^{2}s^{-1}, $R^* = 3.5 \times 10^{-7}$ cm^{-1} and $Z = 0.2$ for lysozyme crystallization trials, we estimate $\Delta F \simeq 19 k_B T$. If the structures of the pre-critical nuclei of the 'fast' and 'slow' processes are the same, then the $\Delta F \approx 19 k_B T$ is the same as well.

As the temperature is lowered, we observe an increase in the average number of nucleation sites per drop, m as shown in figure 4.8b, which indicates that at lower temperatures the probability to find an impurity increases which can happen only through the creation of more impurities. We speculate that the created impurities

could be gel-like clusters of lysozyme[28, 26, 27].

4.3 Conclusion

We have measured nucleation rates of lysozyme crystals using an emulsion based technique. We obtain two nucleation rates at every measurement, both of which we argue are due to the presence of heterogeneous nucleation sites. One set of nucleation sites is rare or fewer per drop and these impurities cause 'fast' nucleation. The other set of nucleation sites, causing the 'slow' nucleation, are protein aggregates and are higher in numbers per drop. The two nucleation rates, described as 'slow' and 'fast', are fitted to theoretical expression of nucleation rates predicted by Classical Nucleation Theory. A detailed analysis of the kinetic pre-factor allowed us to determine the nature of nucleation of both the slow and fast processes to be *heterogeneous*. We have also estimated the activation energy required for a molecule to add on to an existing pre-critical cluster, is $\Delta F \approx 19 k_B T$. For the 'slow' nucleation process the impurities involved are large lysozyme aggregates of order 10μm and are formed due to depletion induced aggregation by the polyethylene glycol molecules. The aggregates are not in a thermodynamic equilibrium with the monomers but are in a dynamically arrested state. We have observed nucleation and growth of crystals from the lysozyme aggregates and as the crystal grows, it consumes the hosting aggregate. As the temperature is lowered, the analysis suggests that more nucleation sites are created as seen from the pre-factor analysis for the 'slow' process. When the large aggregates were removed by filtration, the slow nucleation rate slowed by a factor of four. This rules out the possibility that the slow rate could be homogeneous nucleation, as all monomers pass through the filter. As temperature is lowered, there is an exponential increase in the average number of impurities per sample, m. We speculate that these

impurities are gel-like aggregates of lysozyme monomers that serve as heterogeneous nucleation centers.

Chapter 5

Effect of Bio-Glass on Nucleation of Crystal Polymorphs

Abstract

The role of nucleants in promoting protein crystal nucleation is an on-going field of research. We used amorphous mesoporous bioactive gel-glass[48, 49], $CaO.P_2O_5.SiO_2$ (known as bio-glass) with $2-10$ nm pore-size diameter distribution as a nucleant. The pores act as *heterogeneous* nucleation centers and are claimed[46] to enhance the nucleation rates by molecular confinement. For the protein lysozyme there are multiple polymorphs and we demonstrate that bio-glass preferentially enhances nucleation of spherulite polymorph. Preliminary studies are presented in which the fraction of spherulite crystals shows interesting behaviour suggesting that nucleation of spherulite crystals proceeds via secondary nucleation pathways, or a time-lag in nucleation. The crystal fractions for the different polymorphs are obtained through optical microscopy studies of thousands of independent crystallization samples of water-in-oil emulsion drops produced using microfluidics.

5.1 Introduction

The spherulite, or "sea urchin" crystal morphology is considered a failure in crystallization trials, as they do not yield good quality X-ray diffraction data. Spherulites are bundles of needle-like crystals growing radially outward from a common nucleation center. Nucleation of spherulite crystals is observed[43] in lysozyme/NaCl solutions nearthe liquid-liquid phase boundary[43]. Upon L-L phase separation, sub-micron sized drops of protein-rich phase appear in a solution of protein-poor phase[43]. In the protein-rich drops, multiple nucleation events occur resulting in a cluster of multiple crystals which compete to grow in the protein-poor phase. Due to the high density of crystals at the center of the crystal cluster, the protein is depleted at much faster rates than the diffusion limited in-flux of protein molecules. Therefore the crystals grow radially outward, where protein is available, resulting in a "sea urchin" morphology. The growth of spherulites is controlled by surface kinetics[45] rather than by volume diffusion and varies linearly with time[44, 45]. In the present work, we study the nucleation kinetics of tetragonal and spherulite crystals in the absence and presence of bio-glass particles.

5.2 Experimental Procedure

Lysozyme from chicken egg white is purchased from Sigma Aldrich (Product Number L6876). Without further purification, we dissolved the protein in 0.05 M sodium acetate buffer at pH 4.5. In all our experiments the protein solution is centrifuged for half an hour at $\simeq$10,000g and filtered through 0.22 μm cellulose acetate filters. Lysozyme concentration is measured using Thermo Scientific nano UV-VIS spectrophotometer with lysozyme extinction co-efficient $\epsilon = 2.64$ mL mg^{-1}cm^{-1} at 280 nm. Stock solution of 20% w/v NaCl is also prepared in 0.05 M sodium acetate buffer

at pH 4.5 and filtered through 0.22 μm cellulose acetate filters. Bio-glass (BG) particles are obtained from Prof. Naomi Chayen of Imperial College, London. Suspension of BG particles is prepared in 0.05 M NaAc buffer at pH 4.5 and the suspension is carefully fractionated to obtain BG particles of average size approximately 0.5 μm. The fractionation involves two steps (1) centrifuging the suspension at ** rpm for 10 min and (2) collecting a supernatant layer such that the average particle size is approximately 0.5 μm. Note that, the un-disturbed supernatant contains a suspension with varying particle size. We characterized the particle size using Dynamic Light Scattering (DLS) measurements. The emulsion drops are produced using a

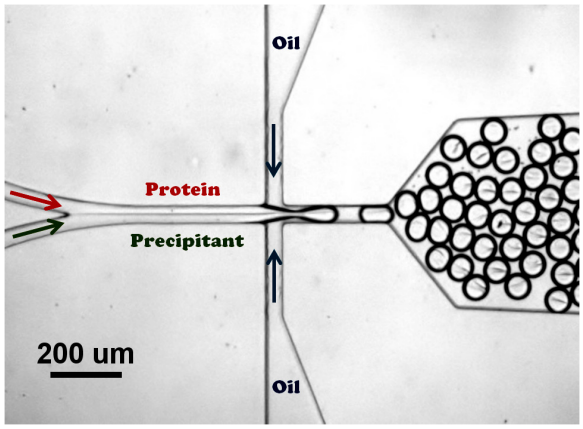


Figure 5.1: Droplet generation using a microfluidic device, the protein and precipitant are mixed on-chip to avoid any nucleation before starting the experiment. The stream labelled 'Protein' contains lysozyme and the stream labelled 'Precipitant' contains 6 % w/v NaCl

flow-focusing microfluidic device as shown in figure 5.1. To avoid nucleation before starting the experiment, the protein and precipitant are mixed on-chip just before making drops. In figure 5.1, the stream labelled 'Protein' contains lysozyme and the stream labelled 'Precipitant' contains 6% w/v NaCl. The suspension of BG particles is added to the precipitant stream. The emulsion drops are then loaded in a capillary and the ends of capillary are sealed with VALAP[29], a mix of equal parts of VAseline, LAnolin and Paraffin wax with low melting temperature. The capillaries are scanned

at regular intervals using a home built robotic stage which can scan and acquire images of capillaries with an accuracy of 6μm. The stage is equipped with two Thermo Electric Coolers (TECs) to control the temperature with a working range of -4℃ to 40 ℃ and are independently controlled. The robotic stage and the temperature are controlled using LabVIEW interface. The images of capillaries scanned at regular intervals are processed manually for droplet and crystal counting.

5.3 Results and Discussion

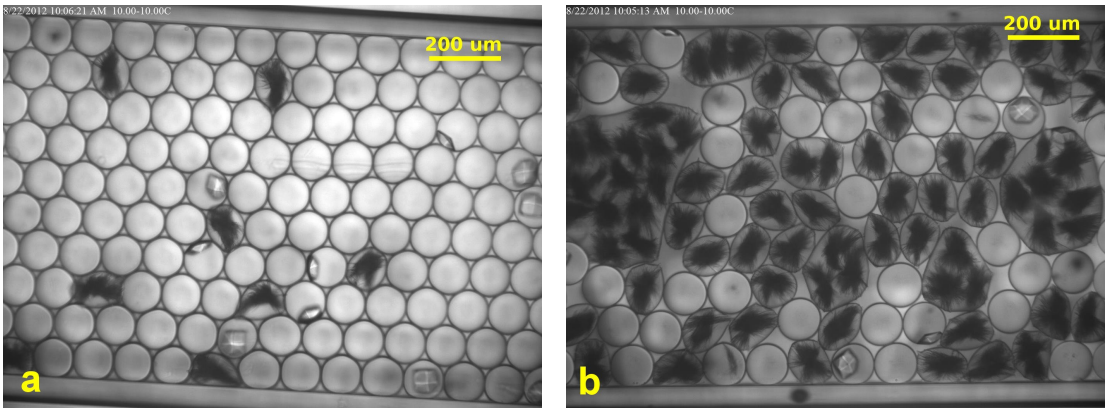


Figure 5.2: Presence of tetragonal and spherulite crystal polymorphs in lysozyme after 40 hrs. in (a) absence and (b) presence of BG particles. The crystallization conditions are Lysozyme 60 mg ml^{-1}, 3% w/v NaCl in 0.05 M NaAc buffer at pH 4.5 and T = 10 ℃. We verified using optical microscopy that each protein drop has more than one BG particles.

Figures 5.2a and 5.2b show the protein drops crystallizing in absence and presence of BG particles at 10 ℃. Under these crystallization conditions, lysozyme exhibits two crystal polymorphs, *tetragonal* and *spherulite*. Notice the enhancement of spherulite crystals in figure 5.2b due to the presence of BG particles after 40 hrs. We measured the crystal fraction of tetragonal and spherulite crystal forms as a function of time (figure 5.3). The kinetic profiles of crystal fractions of tetragonal and spherulite crystals are qualitatively different indicating different nucleation pathways. We discuss

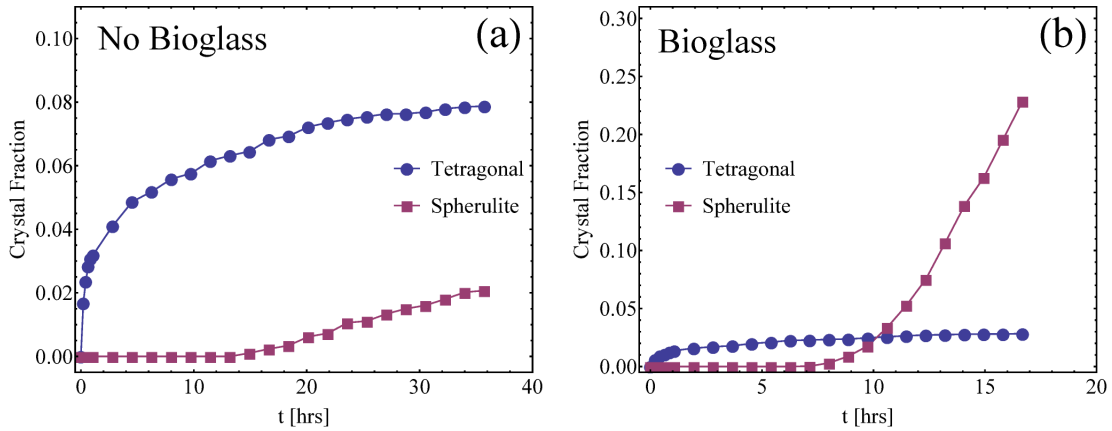


Figure 5.3: Tetragonal and spherulite crystal fractions of lysozyme in (a) absence and (b) presence of BG particles. The crystallization conditions are Lysozyme 60 mg ml^{-1}, 3% w/v NaCl in 0.05 M NaAc buffer at pH 4.5 and T = 10 ℃.

the possible nucleation mechanisms which would corroborate with the experimental observations. In the following discussion, we assumed that nucleation of both tetragonal and spherulite crystals occurs via heterogeneous pathways.

1. **Crystallization of spherulites via secondary nucleation pathways**. Consider

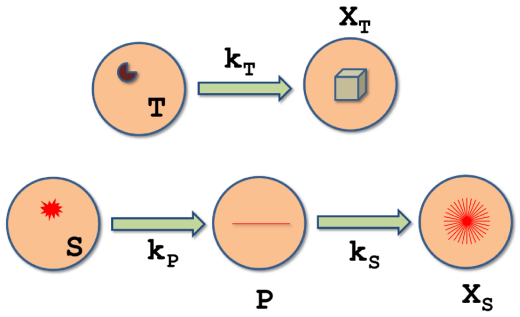


Figure 5.4: Proposed model to understand the nucleation mechanism of tetragonal and spherulite crystals, in which nucleation of spherulites proceeds via secondary nucleation pathways.

a scenario in which, two kinds of impurities exist with each kind promoting *only* one of the crystal polymorphs (figure 5.4). The impurities are distributed randomly among the drops such that a small fraction of drops contain a few tetrag-

onal promoting nucleation sites and rest of the drops contain many spherulite promoting nucleation sites. The tetragonal impurities induce rapid nucleation, but once all of drops containing the tetragonal nucleation sites have crystallized, nucleation of tetragonal crystal form ceases and the tetragonal crystal fraction stops growing with time (figure 5.3). In case of spherulite crystals, nucleation proceeds via primary and secondary nucleation pathways. During primary nucleation, we speculate that invisible primary crystals nucleate and after a long maturation time act as nucleation sites for secondary nucleation. Since our measurements are taken at low magnification, we do not detect the primary crystals until spherulite crystals form via secondary nucleation, which explains the lag time in the kinetic profile of spherulite crystal fraction. Upon adding BG glass particles the nucleation of spherulites is greatly enhanced, while the nucleation of tetragonal crystals does not get significantly altered. However, the underlying nucleation mechanism for the crystallization of tetragonal or spherulite crystals in the absence and presence of BG particles is qualitatively the same in the sense that both forms are present with and without BG particles.

An alternative scenario explaining the cessation of the nucleation of tetragonal crystals at a drop fraction of few percent is that each drop contains two classes of the nucleants, a spherulite nucleant and a tetragonal nucleant. It is possible that an invisible primary spherulite crystal somehow fouls the tetragonal nucleants. This seems improbable and we prefer the scenario of there being two classes of nucleants; a rare tetragonal nucleants present in a few drops and a common spherulite nucleant present in most of the drops.

2. **Time-lag in the nucleation of spherulite crystals**. This model (figure 5.5) is identical to the model discussed above except that, nucleation in spherulites

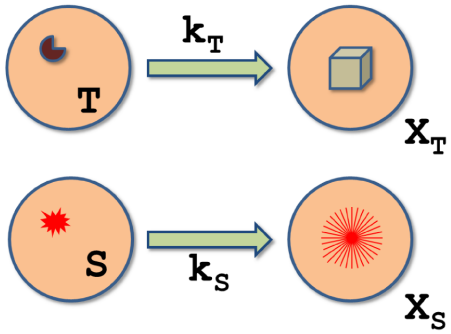


Figure 5.5: Proposed model to understand the nucleation mechanism of tetragonal and spherulite crystals, in which nucleation of spherulites proceeds with a time-lag.

occurs with a time-lag which is the time needed to establish a steady-state distribution of clusters of sizes up to the critical spherulite cluster.

Chapter 6

Conclusion

The principal aim of this research work is to understand the underlying nucleation kinetics of protein crystallization. *Nucleation* is the spontaneous occurrence of embryos of the stable crystalline phase from a bulk metastable supersaturated protein solution. The driving force for nucleation is the chemical potential difference between the stable phase and metastable phase. Nucleation is *homogeneous* when it occurs from the bulk phase and *heterogeneous* if it occurs in contact with a foreign object. Nucleation is an activated process and the kinetics are determined by the activation barrier and the number of nucleation sites. Therefore, besides finding the thermodynamic conditions for the existence of stable crystalline phase, one must follow the right kinetic path to crystallize a protein. One must also be aware of kinetic pathways which lead to the occurrence of gelation and the liquid-liquid phase transition, as these processes deter proteins from crystallization. Being an activated process, nucleation is stochastic, which means one must perform a large number of independent experiments to characterize nucleation. An additional constraint in the case of proteins is the limited availability of purified protein. Microfluidics offers the solution of producing and storing a large number of supersaturated protein drops and requires

only small amounts of protein. Therefore, we developed a microfluidic based method to measure nucleation rates. The method involves crystallizing a large number of supersaturated protein drops at constant temperature. Assuming that nucleation is a Poisson process, the probability for a drop to *not* nucleate decays exponentially with the decay constant proportional to nucleation rate and volume of the drop. We successfully applied the developed technique to measure the nucleation rates of lysozyme crystals as a function of temperature at different supersaturations. Surprisingly, we observe two nucleation rates, 'fast' and 'slow', at every measurement. Pound and La Mer[14] observed the same phenomenon while characterizing the nucleation of molten tin drops and proposed a simple model to explain the presence of two nucleation rates. The model assumes the presence of un-filterable impurities in only a fraction of drops. The drops containing the impurities nucleate via heterogeneous pathways resulting in 'fast' nucleation rate. Classical Nucleation Theory (CNT) predicts that the nucleation rate varies exponentially with supersaturation, with the exponent characterizing the activation energy of nucleation and the kinetic pre-factor multiplying the exponential describing the growth kinetics of a critical nucleus. Therefore, we analyzed the measured nucleation rates according to CNT. Our analysis suggests two important conclusions (1) both the 'slow' and 'fast' nucleation processes are heterogeneous and (2) the kinetic pre-factor, which is usually considered as a weak function of temperature, controls the nucleation kinetics instead of the activation energy. The impurities causing the 'slow' nucleation process are $0.1 - 10$ μm lysozyme aggregates formed due to the depletion attraction mediated by the non-adsorbing polymer Poly-Ethylene Glycol (PEG). These irreversibly aggregated clusters of lysozyme serve as heterogeneous nucleation centers. Nucleation occurs in/on aggregates and as the crystal grows it consumes the hosting aggregate. The impurities causing the 'fast' nucleation are un-filterable by 0.22 μm cellulose acetate filters. At lower temperatures, the number

of impurities causing the 'fast' nucleation increases i.e, the probability to find the 'fast' nucleating impurity in a drop increases which could only happen by creating new impurities. Though the crystallization conditions are far from the liquid-liquid phase separation of lysozyme, we speculate that the impurities are gel-like clusters of lysozyme which serve as heterogeneous nucleation centers at low temperatures. We calculated the change in free energy associated with molecular attachment to an existing pre-critical cluster, $\Delta F \simeq 19 k_B T$. Physically, $e^{-\Delta F/k_B T}$ is the sticking probability of a monomer to an existing pre-critical cluster. We found that ΔF to be independent of temperature.

Appendix A

Chemical Homogeneity in the Drops

To avoid nucleation of protein crystals before starting the experiment, we make drops using a co-flow microfluidic devices. These devices allow us to mix protein and precipitant solutions on chip prior to making the drops and to make protein drops at high supersaturation levels. Since the mixing is done on chip, we need to ensure that all drops are identical. We selected the cloud point or liquid-liquid phase transition as our probe and our conditions are such that the variation of cloud point is sensitive enough to measure a difference of less than a 10% variation in composition. To verify any chemical variation in drop making, the chemical conditions are chosen such that the protein does not crystallize during off-chip mixing (in bulk), which allows us to compare the mixing on-chip vs. off-chip. In both cases, the final composition in the drops is 50 mg/ml lysozyme, 5% w/v PEG 8kD, 0.5 M NaCl in 0.2 M Phosphate buffer at pH 6.2. The drops are collected only after the flows are stabilized. The temperature of the drops is lowered in steps of 0.5℃ until we reach below the cloud point temperature and then raised in steps of 0.5℃ until the drops are clear. At each tem-

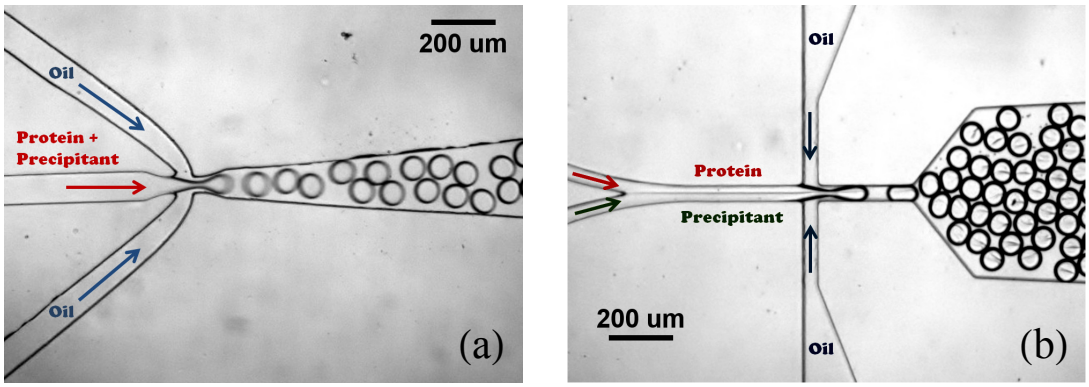


Figure A.1: Generation of emulsion drops using (a) off-chip mixing (b) on-chip mixing microfluidic devices.

perature step, the drops are imaged. Figure A.2a is the plot of the average normalized intensity of the drops produced on-chip and off-chip as a function of temperature. The observed hysteresis in the intensity is due to presence of a finite energy barrier for the liquid-liquid transition while lowering temperature and absence of energy barrier while raising the temperature. We determined the cloud point temperature to be the mean temperature of $T^C_{\downarrow}$ and $T^C_{\uparrow}$, where $T^C_{\downarrow}$ (or $T^C_{\uparrow}$) is the temperature at which the drops are half as cloudy as they could get while lowering (or raising) the temperature of the sample. The overlapping of hysteresis curves for the drops produced on-chip and off-chip indicates that on-chip mixing is as good as off-chip mixing.

To measure the sensitivity of the cloud point measurement to changes in concentration, we performed another experiment in which two sets of drops were produced with a 10% difference in protein concentration using a co-flow microfluidic device. Figure A.2b shows the cloud point hysteresis measured in two sets of drops with a 10% difference in protein concentration. The measured cloud points are $\approx$ 1°C apart indicating that our measurement is sensitive to changes in composition of a few percent.

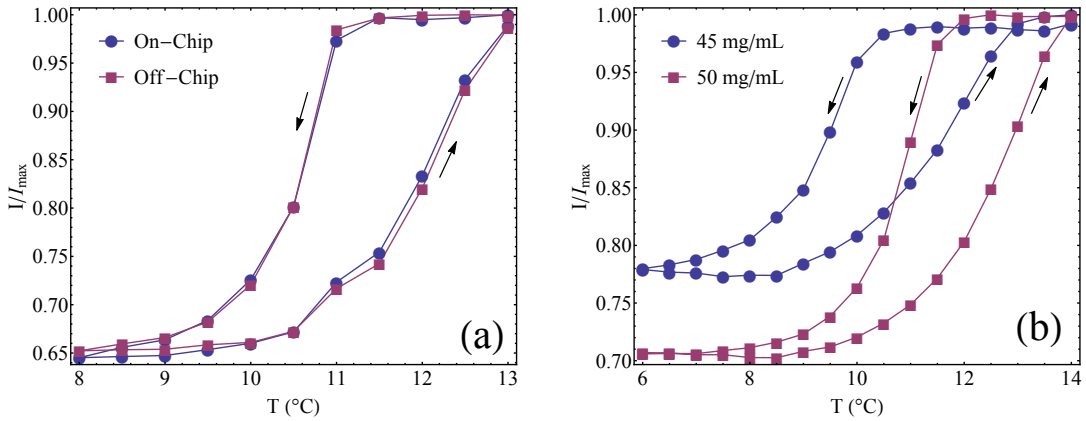


Figure A.2: Average fractional intensity of drops while lowering and raising temperature. (a) Comparison between on vs off-chip mixing of protein and precipitant (b) Sensitivity of the system to cloud point temperature. In both cases, the final composition in the drops is 50 mg/ml lysozyme, 5% w/v PEG 8kD, 0.5 M NaCl in 0.2 M Phosphate buffer at pH 6.2.

Appendix B

Inverse Laplace Transformation

Consider a system of drops nucleating with a distribution of rates $g(s)$, then the fraction of drops which do not have crystals, $f_\phi(t)$ is the Laplace transform of the distribution function $g(s)$. Mathematically,

$$f_\phi(t) = \mathcal{L}g(s) = \int_0^\infty g(s)\mathrm{e}^{-ts}\,\mathrm{d}s \tag{B.1}$$

Since we measure $f_\phi(t)$ and are interested in the decay distribution function, we find the inverse Laplace transform (ILT) of $f_\phi(t)$. The inverse Laplace transform is computed numerically using *Tikhonov Regularization*[30]. This method involves the minimization of the following function.

$$g(s,\alpha) = \min_{g(s)} \left\{ \|f_\phi(t) - \int_0^\infty g(s)\mathrm{e}^{-ts}\,\mathrm{d}s\|^2 + \alpha\|g(s)\|^2 \right\} \tag{B.2}$$

Where, α is the regularization parameter. We used `fmincon`, a MATLAB function to perform the constrained minimization of equation (B.2). Figure B.2a, shows the fractions of drops with no crystals, $f_\phi(t)$ and the corresponding fits obtained using ILT at various temperatures. Calculating $g(s)$ using ILT is a regression analysis involving

infinite number of parameters (since $g(s)$ is a continuous function), hence it is not surprising that ILT yields the best results. Therefore, to affirm the existence of two decay rates we have fitted our data to sum of two exponentials. Figure B.2b is the plot of fractions of drops with no crystals and the fits to sum of two exponentials with the following functional form.

$$f_\phi(t) = a\mathrm{e}^{-bt} + (1-a)\mathrm{e}^{-ct}$$

where, b and c are the decay rates with weights a and $1-a$ respectively. The decay rates b and c correspond to the two peaks observed in the decay rate distribution, $g(s)$ (figure B.1). The weights a and $1-a$ are the areas under the peaks at b and c in the decay rate distribution. Figures B.3 show the plots b's and c's obtained using ILT and two exponential fits. The dotted lines represent a perfect scenario in which both fits yield same fitting parameters. The ILT and two exponential fits yielded somewhat identical b's, however the c's have large discrepancy. Currently, we don't have any understanding regarding the discrepancy in c's. Note that the fitting parameters b, c corresponding to 'slow' nucleation rate, k_s and 'fast' nucleation rate, k_f in our measurements.

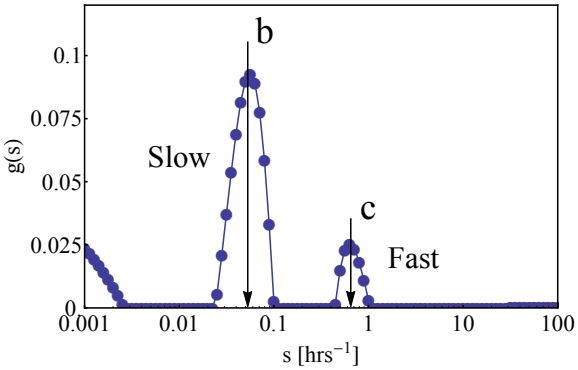


Figure B.1: The two decays rates b and c in equation (B) correspond to the two peaks observed in the decay rate distribution, $g(s)$obtained from the decay rate distribution $g(s)$.

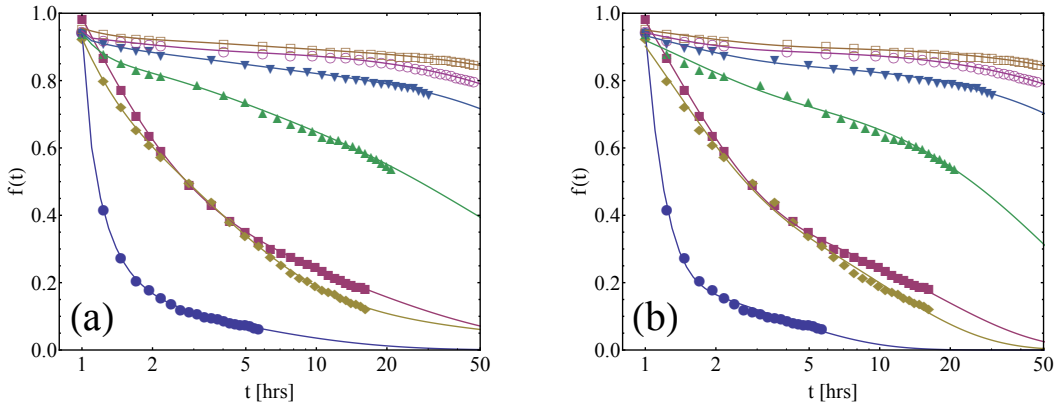


Figure B.2: (a) Data analysis using Inverse Laplace Transform method, (b) sum of two exponentials. The crystallization conditions are 30 mg/mL Lysozyme, 12.5% w/v PEG 8kD, 5% w/v NaCl in 0.1M NaAc at pH 4.8 at different temperatures.

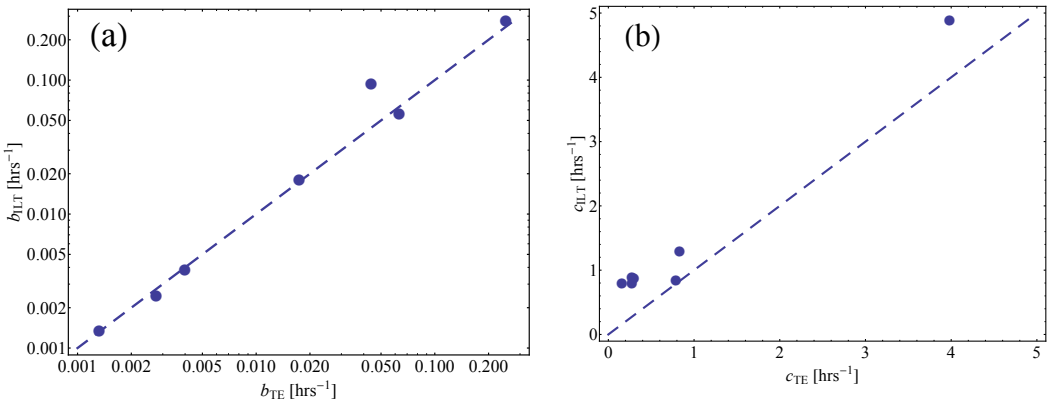


Figure B.3: (a) Slow and (b) fast decay rates obtained from Inverse Laplace Transformation and sum of two exponential fit. The dotted lines represent a perfect scenario in which both fits yield same fitting parameters.

Appendix C

Corrections due to Non-Ideal Solution Behaviour

The definition of supersaturation $\Delta\mu/k_BT = \ln(\mathrm{C}/\mathrm{C_S})$ is an approximation of an ideal protein solution. To estimate the correction due to non-ideality of the protein solution, consider the expression[34] for $\Delta\mu$,

$$\sigma = \Delta\mu/k_BT = \ln(\mathrm{C}/\mathrm{C_S}) + 2B_2M(C - C_S) \tag{C.1}$$

where, B_2 is the second virial coefficient, M, is the molecular weight of lysozyme and C_S is the lysozyme solubility. Table C.1 lists the extrapolated values for the second virial coefficient for crystallization conditions 21-30 mg/ml lysozyme, 12.5% w/v PEG 8kD, 0.5 M NaCl in pH 6.2 phosphate buffer at temperatures 7.2-12 ℃. However, the studied crystallization conditions are at a concentration of 0.86 M NaCl and at a solution pH 4.8. At pH 4.8, the net charge on lysozyme is greater than or equal to the net charge on lysozyme at pH 6.2, which results in stronger Coulombic repulsive interaction between lysozyme molecules at pH 4.8. In other words, the second virial coefficient, B_2, is more positive as a result of stronger Coulombic repulsive forces at

T (℃)	B_2 (mL mol g^{-1})
7.2	-5.03×10^{-4}
8	-4.95×10^{-4}
8.8	-4.87×10^{-4}
9.6	-4.79×10^{-4}
10.4	-4.71×10^{-4}
11.2	-4.63×10^{-4}
12	-4.55×10^{-4}

Table C.1: Estimated[33] second virial coefficients for crystallization conditions: lysozyme 21-30 mg ml^{-1}, 12.5% w/v PEG and 5% w/v NaCl in 0.1 M NaAc at pH 4.8.

pH 4.8. At 0.86 M NaCl, the lysozyme molecules are screened more effectively against the Coulombic repulsive interactions due to higher ionic strength causing B_2 to be more negative. Therefore, decrease in pH and increase in NaCl concentration act in opposite ways. Hence, we expect the estimated B_2 values to be a reasonable for the studied crystallization conditions. It may be noted that lysozyme has a net positive charge of $8.5e^-$ at pH 7.0 and its isoelectric point is pH 11.2[39].

Using the estimated B_2's and equation (C.1), we corrected σ for the non-ideality of the crystallization conditions. The measured 'slow' nucleation rates are fitted to the equation (C.2) using the corrected σ.

$$J_s(\sigma) = A \cdot C \cdot e^{-B/\sigma^2} \tag{C.2}$$

We calculated the barrier height ΔG^* and the kinetic pre-factor as follows.

$$\frac{\Delta G^*}{k_B T} = \frac{B}{\sigma^2} = \frac{1}{k_B T} \frac{16\pi}{3} \frac{\Omega^2 \gamma^3}{\Delta \mu^2}$$

where, $\Omega = 3 \times 10^{-20}$ cm^3, is the lysozyme molecular volume and γ is the interfacial tension between the solution phase and the crystalline phase.

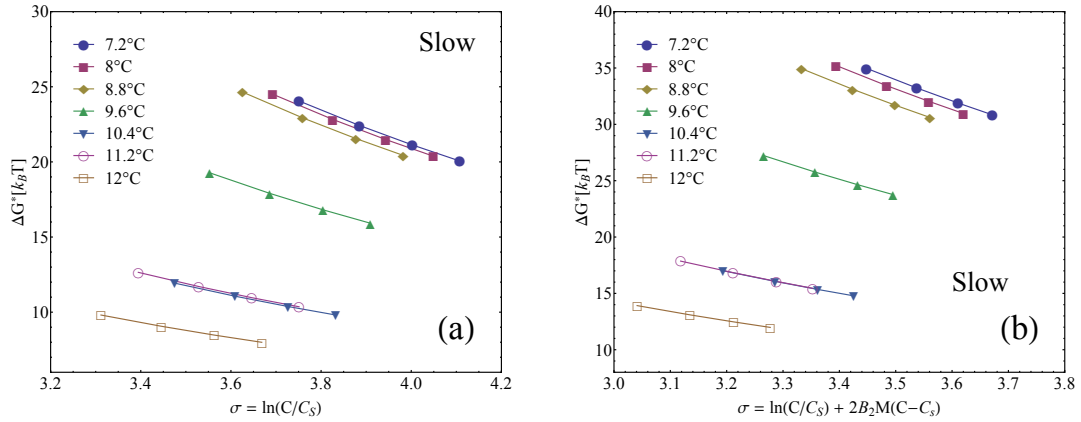


Figure C.1: The barrier heights obtained using (a) σ for ideal solutions and (b) σ for non-ideal solutions (equation (C.1)). Note $\simeq 50\%$ increase in the barrier heights, ΔG^*, obtained using the corrected σ (equation (C.1)).

Conclusion 1: The barrier heights, ΔG^*, obtained (figure C.1) using the corrected σ show qualitatively the same trend as the non-corrected σ although there is a $\simeq 50\%$ increase in the former.

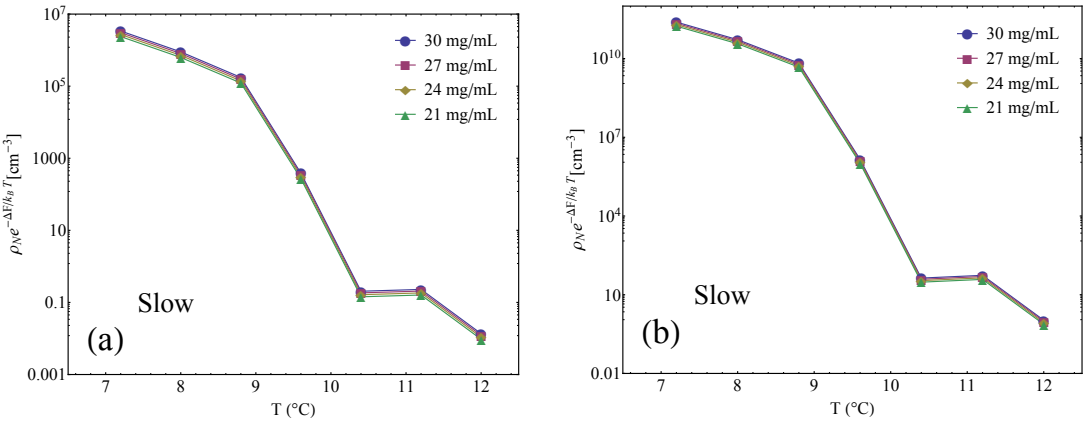


Figure C.2: The kinetic pre-factors obtained using (a) σ for ideal solutions and (b) σ for non-ideal solutions (equation (C.1)). Note 50% increase in the barrier heights, ΔG^*, obtained using the corrected σ (equation (C.1)).

Conclusion 2: Unlike that barrier heights, the kinetic pre-factors (figure C.2) differ by approximately four orders of magnitude, however qualitatively similar trend is observed.

Appendix D

Understanding Galkin & Vekilov Experiments

Galkin and Vekilov[17] attributed the the number of crystals formed at $\Delta t = 0$ to heterogeneous nucleation occurring primarily at the protein droplet-oil interface. Although we agree to their observation, we would like to state predictions of such an assumption. The model shown in figure D.1a, is a protein drop surrounded by oil medium. The protein molecules diffuse and stick to the droplet-oil interface. The protein molecules stuck at the interface de-nature at rate k to form denatured protein molecules which serve as heterogeneous nucleation centers at the interface. Then, the number of denatured proteins at the interface is $k(T_i + \Delta t)$, where T_i is the incubation time which is the time taken until the start of experiment. Note that we define the start of experiment as the time at which target supersaturation is achieved or the target temperature is reached. Assuming that nucleation from heterogeneous nucleation center occurs instantaneously, the rate at which crystals appear at the interface is also $k(T_i + k\Delta t)$, because de-naturation is the rate limiting step in nucleating the crystals. Galkin & Vekilov observe 1 or 2 heterogeneous nucleation events at $\Delta t = 0$,

which suggests that there has been a formation of 1 or 2 denatured proteins during the incubation time, T_i. Nucleation from bulk solution occurs at a rate J and the number of crystals formed in a drop of volume V is $JV\Delta t$. Therefore at any time, $\Delta t > 0$, the number of crystal observed in the drop is $\langle N \rangle = kT_i + (k + JV)\Delta t$ (figure D.1b). Note that, $N_{hetero} = kT_i$ is the number of crystals observed at $\Delta t = 0$. The model has the following predictions.

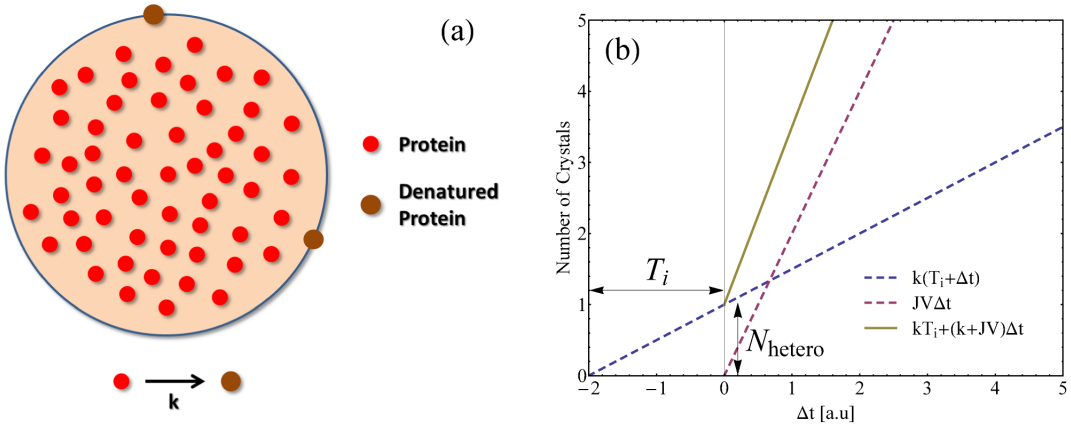


Figure D.1: (a) Proposed model to understand the resulting offset in number of crystals per drop due to heterogeneous nucleation occurring at the protein droplet-oil interface, (b) The number of crystals per drop as a function of quench period Δt according to the proposed model.

1. The average number of crystals N_{hetero} per drop at $\Delta t = 0$ increases with the incubation time. In other words, aging a sample results in increase in N_{hetero}.

2. The average number of crystals N_{hetero} per drop at $\Delta t = 0$ increases (decreases) with increasing (decreasing) drop size.

Appendix E

Characterization of Protein Aggregates

We have verified the presence of protein aggregates using two independent techniques namely, Optical microscopy and Dynamic Light Scattering on Lysozyme and PEG mixtures. The size of the aggregates range from $\simeq$0.1-10 μm. We have used 1-Anilino-8-Napthalene Sulfonic acid (1,8-ANS), a fluorescent dye, which binds specifically to the hydrophobic regions of the protein, to observe the presence of aggregates in fluorescence microscopy. Briefly we discuss our findings in the following two sections.

E.0.1 Using Dynamic Light Scattering (DLS)

Dynamic Light Scattering (DLS) or Quasi-Elastic Light Scattering (QELS) is a technique to determine the hydrodynamic size (or size distribution) of particles in the sub-micron range. The scattered intensity from the diffusing particles is used to calculate the intensity auto-correlation as a function of lag time. For a system with mono-disperse particles, the intensity auto-correlation function decays exponentially with decay rate proportional to the diffusion constant. Using the Stokes-Einstein

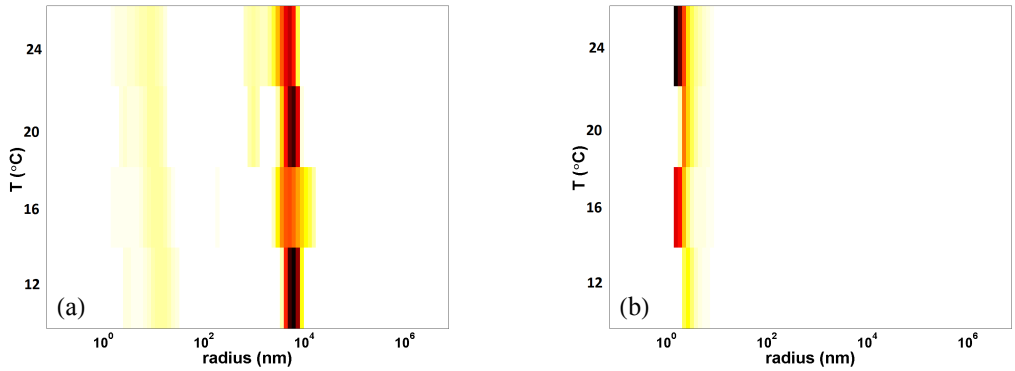


Figure E.1: Presence of aggregates in 50 mg/ml Lysozyme + 12.5% w/v PEG 8kD in 0.1 M NaAc at pH 4.8. (a) Intensity weighted particle size distribution (b) Mass weighted particle size distribution.

relation, the particle size is calculated once diffusion constant, viscosity and temperature of the suspension are known. Figures E.1a and E.1b show the intensity and mass weighted particle size distributions of Lysozyme, PEG mixtures as function of temperature with fraction of particle size mapped on to an intensity colormap (inverted hot). In the limit of Rayleigh scattering, the scattered intensity is $\propto R^6$, where R is the radius of the particle. We can obtain the mass weighted distribution by simply dividing the intensity distribution by the corresponding R^3 at each point along the radius axis, but the Rayleigh limit breaks down at particle sizes comparable to the wave length of light. The particle size at which the Rayleigh limit breaks down is $\rho \approx 2$, where ρ is a reduced variable defined as $\rho = \frac{4\pi(n-1)R}{\lambda}$, n is the refractive index of the particle, R is the radius of the particle and λ is the wavelength of probing light[40]. In our case, we have used $\lambda = 633$ nm, the corresponding particle size at which Rayleigh limit breaks down is R $\simeq$ 250 nm. Therefore particles of radius 250 nm and bigger scatter (approximately) as if they are particles of radius 250 nm. We have taken this into account and adjusted the intensity weighted particle size distribution to obtain the mass weighted distribution. The peak at $\sim$ 10 μm vanishes in the mass weighted plot indicating that the amount of protein mass in the

aggregates is negligible. Also, a strong peak at $\sim$ 1 nm in the mass weighted plot further confirms the fact that most of the protein is in monomeric form. Mathematically, the calculation of particle size (or decay time) distribution involves numerical computation of inverse Laplace transform of the intensity auto-correlation function, which is an ill-posed problem. These kind of mathematical problems are solved using *Tikhonov regularization* methods. We confirmed that the existence of small peaks in the particle size distributions is highly dependent on the regularization parameter and are not trust worthy. Figure E.2 is the plot of particle size distribution as a function of regularization parameter. Note the disappearance of smaller peaks as the regularization parameter, α increases suggesting that α acts as low pass filter.

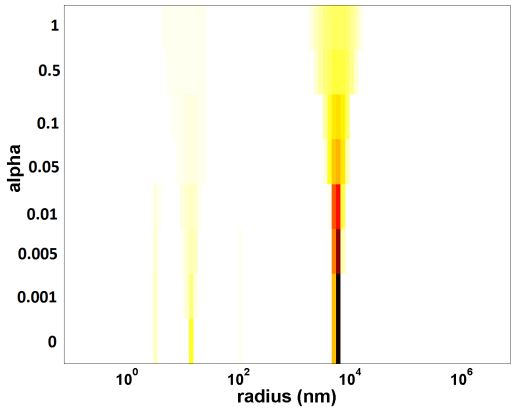


Figure E.2: Intensity weighted particle size distribution of 50 mg/ml Lysozyme + 12.5% w/v PEG 8kD in 0.1 M NaAc at pH 4.8, T = 12 ℃, solution. Disappearance of a smaller/weaker peaks as the regularization parameter, α is increased.

E.0.2 Using DIC and Fluorescence Microscopy

DLS measurements show that the Lysozyme/PEG mixtures contain aggregates of sizes $0.1 - 10$ μm, which can be certainly observed in Optical Microscopy. We have used Differential Interference Contrast (DIC) and Fluorescence microscopy techniques to observe the aggregates. We have used 100 μM 1-Anilinonaphthalene-8-Sulfonic

Acid (1,8-ANS) as the fluorescent dye, which fluoresces only when present in hydrophobic environment. In a protein aggregate (or crystal) the fluorescent dye preferentially partitions into the hydrophobic regions. The dye has excitation at 370 nm and emission at 480 nm. Figure E.3 shows the aggregates in DIC and fluorescence microscopy.

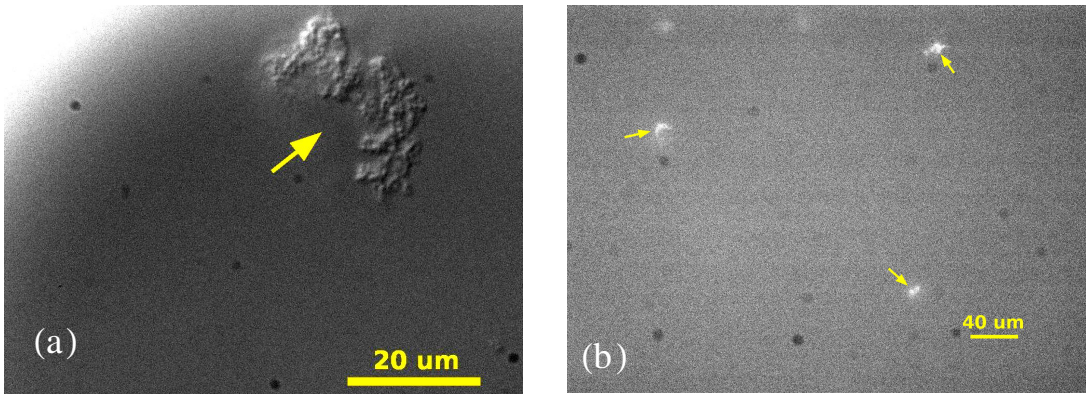


Figure E.3: Presence of aggregates in two different trials of 50 mg/ml Lysozyme + 12.5% w/v PEG 8kD in 0.1 M NaAc at pH 4.8, in (a) DIC and (b) Fluorescence microscopy. 100 μM 1-Anilinonaphthalene-8-Sulfonic Acid as the fluorescent dye which fluoresces only when present in the hydrophobic regions of protein.

Appendix F

Growth Rate Measurements

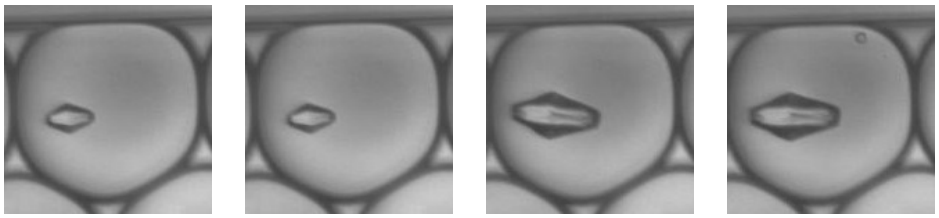
We measured the length of the longest dimension in a crystal as a function of time. Figures F.1 show the images taken at different times during the growth of a crystal. We fit the length of the crystal as a function of time to the following equation.

$$L(t) = a\left(1 - e^{-b(t-t_0)}\right) \tag{F.1}$$

We calculate the growth rate, γ as the growth rate when crystal size is negligibly small. Mathematically,

$$\gamma = \left.\frac{dL}{dt}\right|_{t=t_0} = ab \tag{F.2}$$

Figure F.2 shows the length of crystal as a function of time and the fit to equation (F.1).



(a) t = 0 (b) t = 14 min. (c) t = 28 min (d) t = 42 min

Figure F.1: Measuring the growth of the longest dimension.

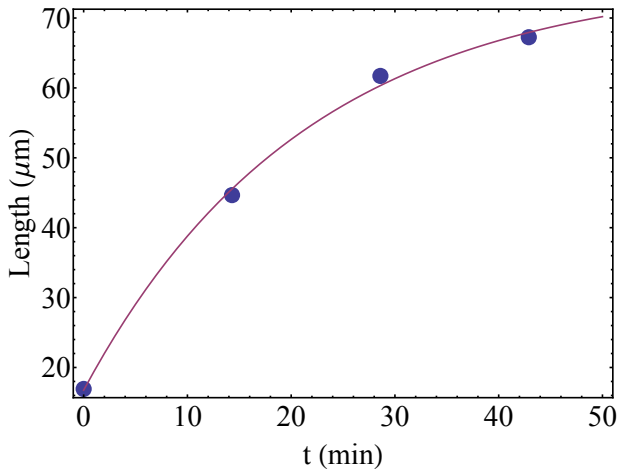


Figure F.2: Length of the longest dimension as a function of time and the fit to equation (F.1).

Appendix G

Monte-Carlo Simulation of Nucleating Drops

We performed simple Monte-Carlo simulations of nucleating drops to understand and estimate the errors associated with scan time which is the time taken by the robotic stage to acquire images of all the drops, and lag time which is the characteristic time for a nucleus to grow to a detectable size (≈ 5 μm). The simulation is based on a model proposed by Pound and La Mer[14]. The model, shown in figure G.1a, consists of large number of drops containing an average number of nucleation sites per drop, m randomly distributed. Nucleation from the drops which *don't* have any nucleation sites occurs at a rate, k_0 and nucleation from a drop containing p number of nucleation sites proceeds with rate, $(k_0 + pk)$, i.e. nucleation can occur from the solution as well as from the sites, where k is the rate of nucleation from a *single* nucleation site. For example, drops which contain 2 nucleation sites nucleate at a rate of $(k_0 + 2k)$. A subtle assumption in Pound and La Mer original model[14] is that nucleation from a drop containing p number of nucleation sites proceeds with rate kp, i.e. nucleation can only occur from a site *not* from solution while in our simulation, nucleation can

occur from the solution as well as from the sites. Assuming that the nucleation sites are distributed according the Poisson distribution, the probability, s, that a drop contains p number of nucleation sites is,

$$s(p) = \frac{e^{-m}m^p}{p!} \tag{G.1}$$

Then, the fraction of drops which do not have crystals as a function of time is given by the following expression.

$$f_\phi(t) = \sum_{p=0}^{\infty} s(p)e^{-(k_0+pk)t} = \sum_{p=0}^{\infty} \frac{e^{-m}m^p}{p!}e^{-(k_0+pk)t} = e^{-m}e^{-k_0t}e^{me^{-kt}} \tag{G.2}$$

The image, shown in figure G.1b, is the simulation model in which each pixel is

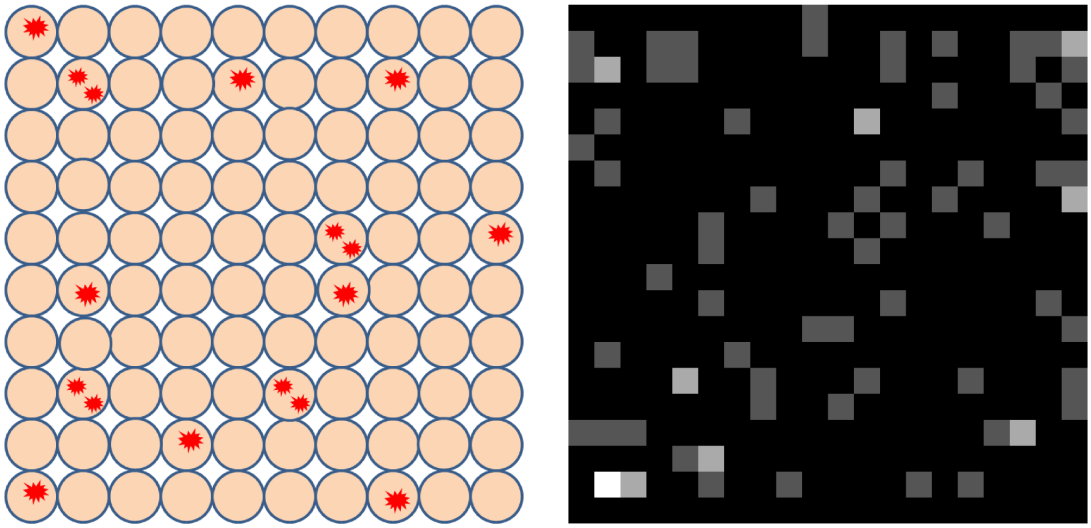


Figure G.1: (a) Schematic of Pound and La Mer Model (b) Simulation model, where each pixel is equivalent to a drop and the number of nucleation sites in a drop is mapped onto the gray scale of a pixel, for this image $m = 0.2$.

equivalent to a drop. The gray-scale value of each pixel is mapped onto the number of nucleation sites associated with that pixel. At a time step dt the probability for nucleation to occur in a randomly chosen pixel is given by $(k_0 + pk)dt$. The resulting

decay in the number of dark pixels (non-crystallized drops) follows equation (G.2). The time delay, t_g in detecting the crystal due to slow growth rates is simulated as the time required for a pixel to acquire a threshold value i.e. for a pixel to be considered as crystallized, the pixel value must be greater than a threshold value. Figure G.2b shows a screen shot of nucleating drops, in which 'black' pixels did not crystallize, 'white' pixels crystallized and any other colored pixel is nucleated, but the crystal has not yet grown to the detectable size. The measured time delay, $t_g \sim 0.1$ hrs, corresponds to the experimental value for lysozyme at the conditions we studied. Figure G.3 shows the fraction of non-crystallized drops as a function of time from an experiment and the corresponding simulation. In a typical experiment, 1000s

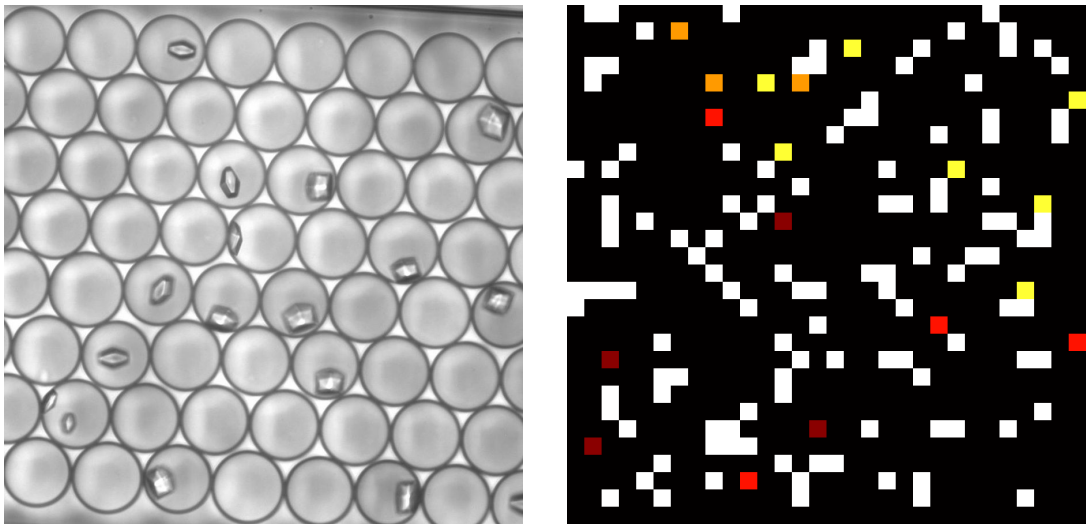


Figure G.2: (a) Nucleating drops in an experiment. (b) Nucleating drops in simulation. 'Black' and 'White' pixels corresponding to non-crystallized and crystallized drops respectively and any other colored pixel is a nucleated drop but the crystal has not yet grown to the detectable size.

of drops are stored in a rectangular capillary and are scanned at regular intervals ($\sim 0.25 - 2$ hrs) using a home-built robotic stage. The scan time, t_s is the required for scanning and acquiring images of $\sim$ 4000 drops is $\sim$ 0.1 hrs. In obtaining the fraction of non-crystallized drops, we assumed that time delay, $t_g = 0$ (instantaneous

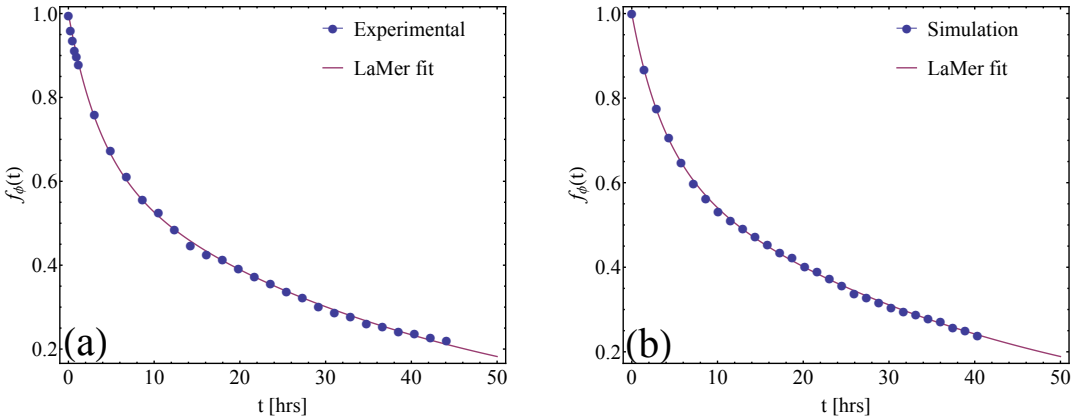


Figure G.3: (a) Experimentally obtained fraction of non-crystallized drops and the fit to (G.2) from an experiment with $k_0 = 0.02504$ nL^{-1} hr^{-1}, $k = 0.2024$ hr^{-1} and $m = 0.4506$. (b) Fraction of non-crystallized and the fit to (G.2) from the Monte-Carlo simulation also with $k_0 = 0.02504$ nL^{-1} hr^{-1}, $k = 0.2024$ hr^{-1} and $m = 0.4506$. The parameters obtained from fit to equation (G.2) are $k_0 = 0.02539$ nL^{-1} hr^{-1}, $k = 0.1814$ hr^{-1} and $m = 0.4509$. Data not corrected for delay due to slow growth.

growth) and $t_s \ll 1/k$, where $1/k$ is the characteristic time for nucleation. Under these assumptions, we estimated the errors as t_s and t_g are varied. In the following simulations, unless stated, we used $k_0 = 0.018$ nL^{-1} hr^{-1}, $k = 0.72$ hr^{-1} and $m = 0.2$ which are typical values from our experiments.

Conclusion 1: From the simulations using different number of drops, we concluded that ~ 500 independent drops are necessary to obtain statistics within 10% confidence range. All our experimental results are obtained with $\sim 1200-4000$ drops.

Conclusion 2: When t_s is much smaller than the characteristic nucleation times, the errors are minimal, but when t_s is comparable to the characteristic nucleation times the errors are large. This effect is observed in figures G.4, G.5 and G.6 as $t_s > 1/k$. Note that, we define characteristic nucleation time as inverse of nucleation rate. From our experiments, the scan time, $t_s = 0.1$ hrs which is much smaller compared to the typical characteristic nucleation times ~ 56 hrs and ~ 1.4 hrs corresponding

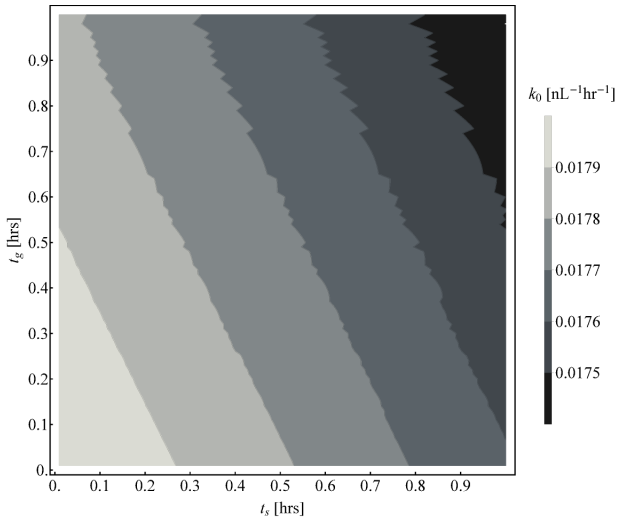


Figure G.4: The simulation is performed with $k_0 = 0.018$ nL^{-1} hr^{-1}, $k = 0.72$ hr^{-1}, $m = 0.2$. Nucleation rate k_0 as a function of scan time, t_s and delay time, t_g. Data not corrected for delay due to slow growth.

to $k_0 = 0.018$ nL^{-1} hr^{-1} and $k = 0.72$ hr^{-1} respectively. Therefore, we expect to have $< 10\%$ error in our measurement of 'fast' nucleation rate.

Conclusion 3: For finite t_g (slow growth), the error can be minimized by correcting for the offset in $f_\phi(t)$ as $f_\phi(t - t_g)$. In our experiments the correction did not improve our errors since $t_s \sim t_g$.

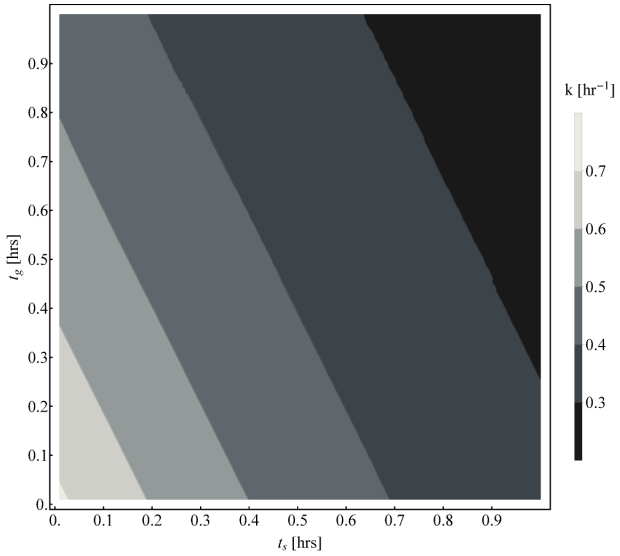


Figure G.5: The simulation is performed with $k_0 = 0.018$ nL^{-1} hr^{-1}, $k = 0.72$ hr^{-1}, $m = 0.2$. Nucleation rate k as a function of scan time, t_s and delay time, t_g. Data not corrected for delay due to slow growth.

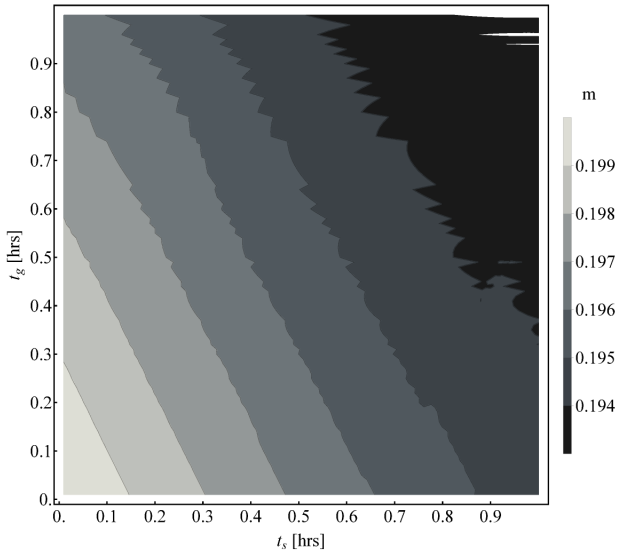


Figure G.6: The simulation is performed with $k_0 = 0.018$ nL^{-1} hr^{-1}, $k = 0.72$ hr^{-1}, $m = 0.2$. Average number, m of impurities per drop as a function of scan time, t_s and delay time, t_g. Data not corrected for delay due to slow growth.

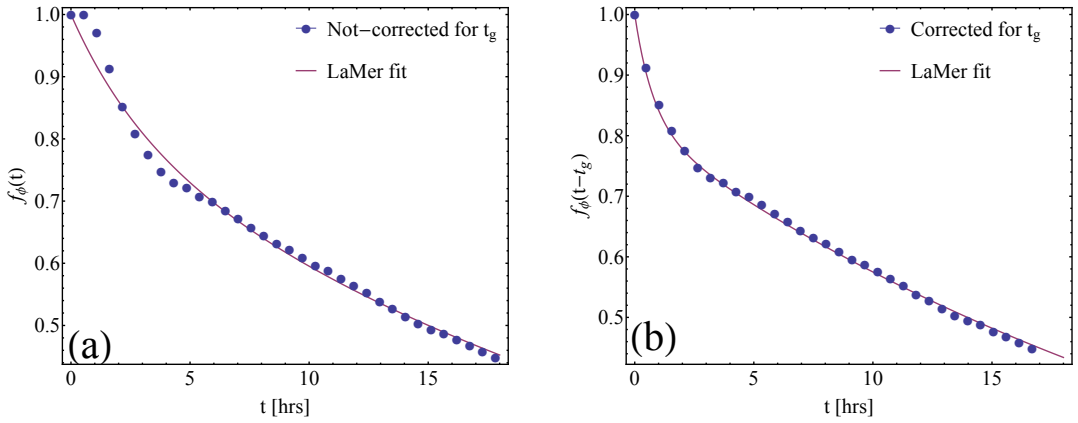


Figure G.7: The simulation is performed with the number of drops $N = 1600$, $k_0 = 0.036$ nL^{-1} hr^{-1}, $k = 1.44$ hr^{-1}, $m = 0.2$, $t_g = 1.2$ hrs and $t_s = 0$ hrs. (a) fraction of drops with no crystals vs. time, $f_\phi(t)$ and the La Mer fit with fit parameters $k_0 = 0.03268$ nL^{-1} hr^{-1}, $k = 0.265$ hr^{-1} and $m = 0.2065$. (b) Corrected fraction of drops with no crystals vs. time, $f_\phi(t - t_g)$ and La Mer fit, with fit parameters $k_0 = 0.0352 \pm 0.0002$ nL^{-1} hr^{-1}, $k = 1.43 \pm 0.0504$ hr^{-1} and $m = 0.2 \pm 0.0036$.

Appendix H

Nucleation rate measurements under no Poly-Ethylene Glycol (PEG)

We demonstrated that the protein aggregates, which are formed when PEG and Salt are added to lysozyme solution, act as nucleation centers and are the source of 'slow' nucleation rate. To understand the role of aggregates in observing two nucleation rates, we have measured nucleation rates under crystallization conditions without PEG. We still obtain two nucleation rates and figure H.1 shows the 'fast' and 'slow' nucleation rates measured as a function of supersaturation, $\sigma = \ln(C/C_0)$. The 'slow' nucleation rates are fitted to the following equation.

$$J_s(\sigma) = A \cdot C \cdot e^{-B/\sigma^2} \tag{H.1}$$

We have obtained the barrier height, $\Delta G^* \approx 9k_BT$, $\Delta F = 19k_BT$ and $\rho_N \approx 10^7$ cm^{-3}. The measured number density of nucleation sites, ρ_N, is much smaller than one would expect for *homogeneous* nucleation ($\approx 10^{18}$ cm^{-3}), which suggests the nucleation is

heterogenous.

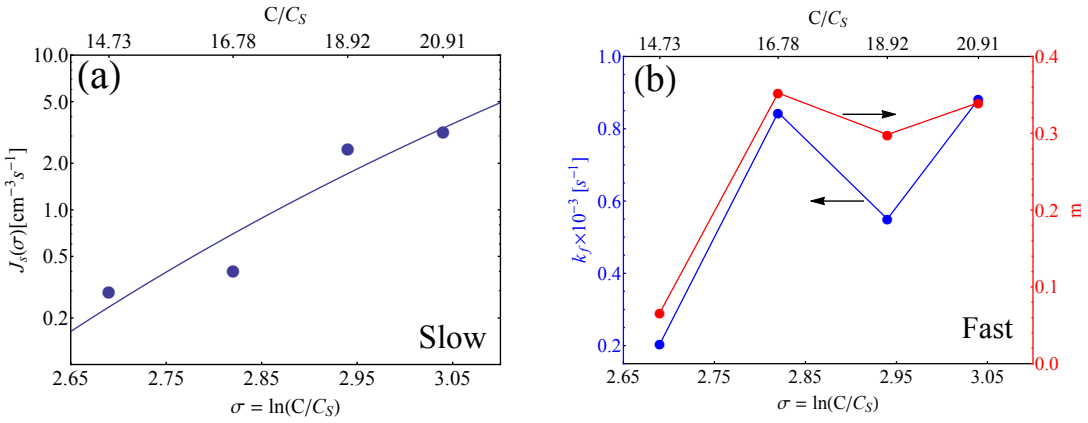


Figure H.1: (a) Slow nucleation rate (b) Fast nucleation rate and the average number, m, of nucleation sites per drop vs. supersaturation. The crystallization conditions are Lysozyme 42-60 mg ml^{-1}, 2.5% NaCl in 0.05 M NaAc buffer at pH 4.5 and T = 8 ℃

Bibliography

[1] P.A. Darcy, J.M. Wiencek, *Acta Crystallogr. D*, 54 **1998**, 1387.

[2] Shanfeng Jiang and Joop H. ter Horst, *Crystal Growth and Design*, **2011**, 11, 256-261.

[3] Joop H. ter Horst et. al, *Crystal Growth and Design*, **2013**, 11, 2435-2440.

[4] P. Laval et al. *Journal of Crystal Growth*, **2007**, 303, 622-628.

[5] Jean-Baptiste Salmon et. al, *Langmuir*, **2009**, 25, 1836-1841.

[6] N.M. Dixit, A.M. Kulkarni, C.F. Zukoski, *Colloids Surf. A*, **2001**, 190, 47-60.

[7] Sathish V. Akella, Aaron Mowitz, Michael Heymann and Seth Fraden, submitted to *Crystal Growth and Design*, **2013**.

[8] Sathish V. Akella, Michael F. Hagan and Seth Fraden, submitted to *Crystal Growth and Design*, **2013**.

[9] J.C. Fisher, J.H. Hollomon and D. Turnbull, *Journal of Applied Physics*, **1948** , 19, 775-784.

[10] D. Turnbull; J.C. Fisher, *J. Chem. Phys.*, **1949**, 17, 71-73.

[11] J.C. Fisher, J.H. Hollomon and D. Turnbull, *Science*, **1949**, 109, 168-169.

[12] D. Turnbull, *Journal of Applied Physics*, **1949**, 20, 817.

[13] D. Turnbull, *J. Appl. Phys.*, **1950**, 1022-1028.

[14] Guy M. Pound, Victor K. La Mer, *J. Am. Chem. Soc.*, **1952**, 74, 2323-2332.

[15] V. G. Taratuta, A. Holschbach, G. M. Thurston, D. Blankenstein and G. B. Benedek, *J. Phys. Chem*, **1990**, 94, 2140-2144.

[16] P. Jolles and J. Berthou, *FEBS Lett.*, **1972**, 23, 21-24.

[17] Oleg Galkin; Peter G. Vekilov, *J. Phys. Chem. B*, **1999**, 103, 10965-10971.

[18] Oleg Galkin; Peter G. Vekilov, *Journal of Crystal Growth*, **2001**, 232, 63-76.

[19] Oleg Galkin; Peter G. Vekilov, *J. Mol. Biol.*, **2004**, 336, 4359.

[20] Venkateswarlu Bhamidi et al, *Crystal Growth and Design*, **2002**, 395-400.

[21] Jean-Pierre Astier, Stephane Veesler, *Crystal Growth and Design*, **2008**, 8, 4215-4219.

[22] Seila Selimovic, Yanwei Jia and Seth Fraden, *Crystal Growth and Design*, **2009**, 9, 1806-1810.

[23] Seila Selimovic, Frederic Gobeaux and Seth Fraden, *Lab on a Chip*, **2010**, 10, 1696-1699.

[24] Richard D. Dombrowski et al, *Am. Inst. Chem. Engg.*, **2010**, 56, 79-91.

[25] M. Ildefonso; N. Candoni; S. Veesler, *Crystal Growth and Design*, **2011**, 11, 1527-1530.

[26] Avanish S. Parmar, Paul E. Gottschall and Martin Muschol, *Biophysical Chemistry*, **2007**, 129, 224-234.

[27] Emmanual Saridakis et al, *Acta Cryst.*, **2002**, D58, 1597-1600.

[28] Anna Stradner et al, *Nature*, **2004**, 432, 492-495.

[29] Shinya Inoue and Kenneth R. Spring, *Video Microscopy: The fundamentals*, Plenum Press, 1997, New York.

[30] A.N. Tikhonov, V.Y. Arsenin, *Solutions of Ill-Posed Problems*, Winston, 1977.

[31] Stephen. W. Provencher, *Computer Physics Communications*, **1982**, 27, 213-227.

[32] Stephen. W. Provencher, *Computer Physics Communications*, **1982**, 27, 229-242.

[33] J. Bloustine et al, *Physical Review Letters*, **2006**, 96, 087803.

[34] B. Guo; S. Kao; H. McDonald;, A. Asanov; L. L. Combs and W. W. Wilson, *J. Crystal Growth*, **1999**, 196, 424-433.

[35] J. D. Schmit and K. Dill, *J. Am. Chem. Soc.*, **2012**, 134, 3934-3937.

[36] Dimo Kashchiev, *J. Chem. Phys.*, **1982**, 10, 5098-5102.

[37] R. P. Sear, *J. Physics:Condensed Maatter*, **2007**, 19, 033101.

[38] Pablo G. Debenedetti, *Metastable Liquids: Concepts and Principles*, Princeton University Press, 1996.

[39] J.D. Gunton; A. Shiryayev and D.L. Pagan, *Protein Condensation: Kinetic Pathways to Crystallization and Disease*, Cambridge University Press, 2007.

[40] H. C. Van de Hulst, *Light Scattering by Small Particles*, Dover Publications, Inc., 1981, New York.

[41] Richard P. Sear, *Crystal Growth and Design*, **2013**, 13, 1329-1333.

[42] Richard P. Sear, *J. Phys. Chem. B*, **2006**, 110, 21944-21949.

[43] Muschol, M. and Rosenberger, F. *J. Chem. Phys.*, **1997**, 107, 1953-1962.

[44] Keith, H.D. and Padden, F.J, *J. Appl. Phys.*, **1963**, 34, 2409-2421.

[45] Chow, P. Liu, X.; Zhang, J.; Tan, R. *Appl. Phys. Lett.*,**2002**, 81, 1975-1977.

[46] Amanda J. Page and Richard P. Sear, *Phys. Rev. Lett.*, **2006**, 97, 065701.

[47] Chayen, N.E. et al., *J. Mol. Biol.*, **2001**, 312, 591595.

[48] Chayen, N.E. et al., *Proc. Natl. Acad. Sci., U. S. A.*, **2006**, 103, 597601.

[49] Emmanuel Saridakis and Naomi E. Chayen, *Trends in Biotechnology*, **2008**, Vol.27, No.2, 99-106.

[50] Samuel I. A. Cohen, Michele Vendruscolo, Christopher M. Dobson and Tuomas P.J. Knowles, *J. Mol. Biol.*, **2012**, 421, 160-171.

[51] C. Holtze et. al, *Lab on a Chip*, **2008**, 8, 1632-1639.