CRYSTAL GROWTH & DESIGN

Article

pubs.acs.org/crystal

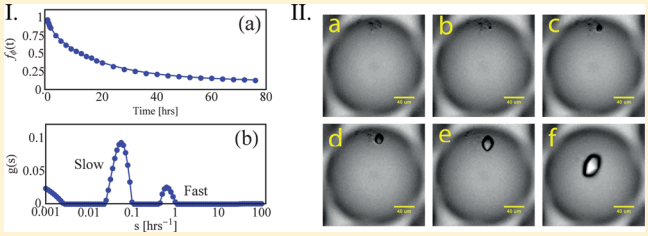
Emulsion-Based Technique To Measure Protein Crystal Nucleation Rates of Lysozyme

Sathish V. Akella,[†] Aaron Mowitz,[†] Michael Heymann,[†,‡] and Seth Fraden*[,†]

[†]Martin A. Fisher School of Physics and [‡]Department of Biochemistry, Brandeis University, Waltham, Massachusetts 02454, United States

Supporting Information

ABSTRACT: We measured the nucleation rates of lysozyme protein crystals using microfluidically produced emulsion drops containing supersaturated protein solution. The technique involves quenching several thousand independent nanoliter drops by rapidly lowering the temperature and then counting the number of drops that have not nucleated as a function of time at constant temperature. We fit the number distribution to a theoretical model developed by Pound and La Mer (*J. Am. Chem. Soc.* **1952**, *74*, 2323−2332) for heterogeneous nucleation and extract two nucleation rates and the number of nucleation sites per drop. We describe the technique in detail and present our analysis of the measured nucleation rates within the context of Classical Nucleation Theory, which adequately describes our observations. Of the two nucleation rates, one is a slow rate that varies with temperature and one is a fast rate independent of temperature. The nucleation barrier and kinetic prefactors are obtained for each rate. Notably, there is no detectable barrier for the fast rate. Both rates are inconsistent with the process of homogeneous nucleation and are consistent with heterogeneous nucleation.



1. INTRODUCTION

Determining the structure of proteins is an important step in biology and biotechnology. Crystallization, which is necessary for structure determination using X-ray diffraction, remains a bottleneck in this endeavor. At the present, for proteins whose structure is unknown, there is no theoretical guidance for whether a crystalline phase exists. Furthermore, even if a crystal phase does exist, protein solutions often have a metastable liquid−liquid phase or form gels and other precipitates when purified, preventing the formation of the underlying crystalline phase. For the foreseeable future, protein crystallization will remain a trial and error process. However, that does not mean that all crystallization methods are equivalent; there are sound principles dating back to Gibbs' work in 1876 that provide guidance on how to systematically approach crystallization.[2−8] The first step in the crystallization process is to scan the protein-precipitant phase space in search of promising crystallization conditions, if any. The underlying assumption is that crystallization is a phase transition and it is necessary to find the physical−chemical conditions of the equilibrium crystal phase. However, crystallization is also an activated process and therefore even formulating a protein solution under the exact conditions of the equilibrium crystal phase is not sufficient for crystallization to occur. It is necessary to supersaturate the protein solution to induce crystallization, and the temporal variation of supersaturation, or supersaturation kinetics, is crucial for optimizing conditions that will produce crystals most suitable for X-ray crystallography.[4]

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 transformation of the nucleus into a macroscopic crystal. Classical Nucleation Theory (CNT) attempts to explain the statistical mechanics and thermodynamics of nucleation, but the applicability of CNT remains controversial. Important questions that we address are to what extent is nucleation homogeneous or heterogeneous and does nucleation proceed according to CNT.[9−11]

Our aim is to develop a method for measuring nucleation rates that is accurate, sensitive, simple, and applicable to a variety of proteins. In the 1950s, Turnbull[12−16] characterized nucleation in supercooled liquid metals using an emulsion method in which drops of liquid metals were dispersed in a continuous phase of oil and stabilized against coalescence by a surfactant. Since then, there have been efforts based on Turnbull's method toward characterizing and understanding nucleation in proteins.[9,17−21] Vekilov's group[9] developed a method to measure nucleation rates using sample volumes ≈ 1 μL, and our group[20,22,23] and the Veesler[21] group have extended the method to smaller volumes using microfluidics. The work presented here continues in this vein, and we have

Received: April 23, 2014
Revised: July 17, 2014
Published: July 28, 2014

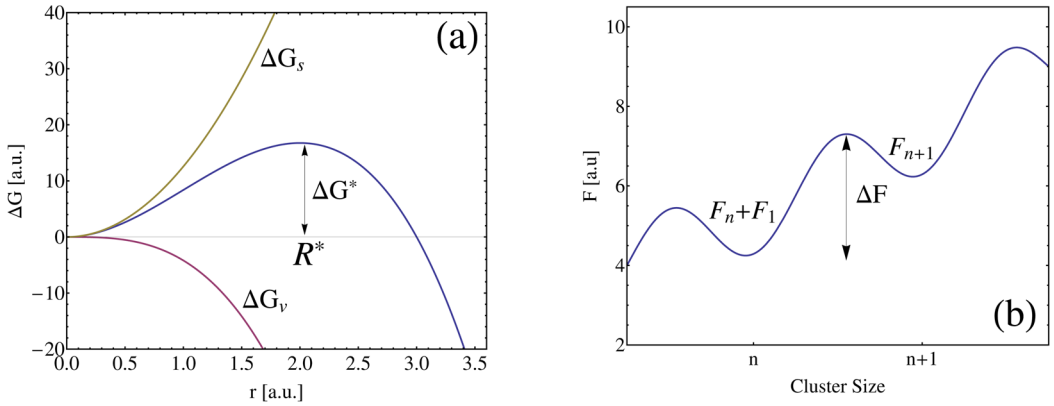


Figure 1. (a) Free energy as a function of radius. The energy barrier, ΔG^*, for nucleation is a result of the competition between the volume free energy, ΔG_v and surface free energy, ΔG_s. (b) Activation energy required, ΔF, for the growth of a cluster of size n to $n + 1$.

adopted Turnbull's technique to measure nucleation rates of protein crystals using drop volumes of less than 1 nL. This method involves supersaturating large numbers of identical and 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 and that all the drops are identical, the number of drops without crystals is predicted to decay exponentially with time with a decay constant that is proportional to the nucleation rate and the volume of the drop. We present measurements on nucleation rates of the model protein lysozyme, which was chosen because we could validate our method by comparing our results against previous experiments. The measured nucleation rates are analyzed in the context of CNT. The analysis yielded values for the nucleation barriers and suggests that the nucleation is heterogeneous. We identify the heterogeneous sites as being aggregates of the protein using multiple techniques, including optical microscopy.

The determination of nucleation rates requires measuring only the fraction of drops without crystals. However, a consequence of our particular experimental protocol was that on achieving equilibrium each drop contained only one crystal. This is not a necessary condition for our measurement method, which works equally well for conditions under which more than one crystal is nucleated per drop. However, one crystal per drop is the optimal condition for serial crystallography in which diffraction data is collected from a series of crystals, one at a time, using noncryocooled, randomly oriented crystals. It is worthwhile to mention the efforts of other research groups[24−26] in accomplishing the task of performing X-ray crystallography on crystals grown in glass capillaries and microfluidic chips. Our group is also actively working on designing a microfluidic chip where the background radiation is minimized by making the chip thin.[27] Serial crystallography has several technical advantages over conventional methods. First, the crystals can be small, which increases the potential for growing crystals in the first place. Second, it avoids the 10−100-fold increase in crystal mosaicity typically encountered during cryoprotection.[28] Third, it eliminates the need to search for cryoprotectant conditions. Although noncryoprotected crystals suffer radiation damage at a roughly hundred times higher rate than that of cryoprotected crystals, it is relatively easy to merge single diffraction frames taken from many crystals of random orientations into a single, complete diffraction set that covers the Ewald sphere. Serial crystallography was first pioneered at X-ray free-electron lasers,[29,30] where diffraction data is collected before appreciable radiation damage occurs, but serial crystallography may also be fruitfully employed at synchrotron beamlines,[31,27] if the technology to produce the necessary crystals of sufficient quality and quantity is simple and inexpensive. It turns out that our emulsion-based nucleation rate measurement method achieves this objective.

2. BACKGROUND

2.1. Nucleation Rates. 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, as well as kinetic factors, such as the monomer diffusion constant and rates of monomer incorporation into the growing precritical nuclei. 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. 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 $\Delta G_v = -(4\pi r^3/3)\rho_c \Delta\mu$ is the volume energy and $\Delta G_s = 4\pi r^2\gamma$ is the surface energy of a spherical cluster of radius r. ρ_c is the number density of the crystalline phase, $\Delta\mu$ 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 1a shows the development of the barrier, ΔG^*, as a result of competition between the volume and the surface free energies of a cluster. From the total free energy

$$\Delta G = -\frac{4\pi r^3}{3}\rho_c \Delta\mu + 4\pi r^2\gamma \quad (1)$$

we obtain the activation energy, ΔG^*, by setting $(\partial\Delta G/\partial r)|_{R^*} = 0$

$$\Delta G^* = \frac{16\pi}{3}\frac{\gamma^3}{(\rho_c\Delta\mu)^2} \quad (2)$$

where $R^* = 2\gamma/\rho_c\Delta\mu$ is the radius of the critical cluster. Clusters smaller that R^* are unstable and will lower their free energy by shrinking, whereas clusters larger than R^* will lower their free energy by growing. Heterogeneous nucleation is described by a similar expression. The presence of a surface of lower energy for the protein means that the surface tension is

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487−4509

reduced, but a barrier still remains, so in eq 2 γ is replaced with γ_{hetero}.

The rate at which nucleation occurs is proportional to the Boltzmann weight $e^{-\Delta G^*/k_BT}$, associated with the activation energy

$$J = \rho_N Z j\, e^{-\Delta G^*/k_BT} \tag{3}$$

Here, $\rho_N\, e^{-\Delta G^*/k_BT}$ is the number of critical nuclei per unit volume, Z is the Zeldovich factor, which is the probability that the critical nucleus goes forward into the new phase and not back into the metastable phase, and j is the rate at which individual molecules attach to a critical nucleus.[32] The proportionality factor in eq 3, $\rho_N Z j$, is known as the kinetic prefactor, where ρ_N is the number density of nucleation sites.

The rate (j) at which molecules add on to the critical nucleus can be expressed as

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

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 1b) for the addition of a monomer to an existing cluster. The physical understanding of j is that $4\pi\rho D R^*$ is the diffusion limited rate at which protein molecules impinge on a critical cluster of radius R^* and that $e^{-\Delta F/k_BT}$ 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 is independent of whether the nucleation is heterogeneous or homogeneous, and we assumed that ΔF is independent of concentration. These assumptions are based on the picture that a precritical nucleus has the same form as the postcritical nucleus and that the form of the crystal is independent of whether the nucleation mechanism is homogeneous or heterogeneous, as illustrated in Figure 2.

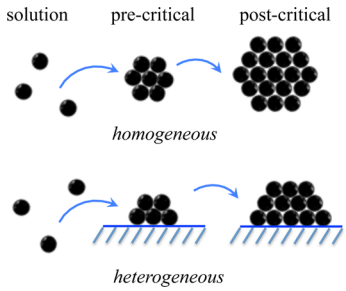


Figure 2. Assumptions of Classical Nucleation Theory. Homogeneous and heterogeneous precritical and postcritical clusters have the same structure. (a) Homogeneous nucleation occurs in bulk. (b) Heterogeneous nucleation occurs on a surface.

The Zeldovich factor, $Z \simeq (n^*)^{-2/3}$,[11] 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 homogeneous nucleation, any of the individual monomers in solution serves as a nucleation site; hence, $\rho_N = \rho$. However, for heterogeneous nucleation, the number density of nucleation sites is typically much smaller, $\rho_N \ll \rho$. The nucleation rates for homogeneous and heterogeneous nucleation rates are similar in form

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

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

We can estimate the value of the kinetic prefactors as follows.[10] For typical lysozyme crystallization trials, the concentration of lysozyme is ~30 mg mL^{-1}, which corresponds to $\rho = 10^{18}$ cm^{-3}, since the molecular weight of lysozyme is $M_w = 14\,700$ g mol^{-1}. The diffusion constant of lysozyme is $D = 10^{-6}$ cm^2 s^{-1}, and for a critical cluster of 12 molecules, $R^* = 3.5$ nm. Hence

$$J_{\text{hom}}\ (\text{cm}^{-3}\,\text{s}^{-1}) \simeq 10^{23}\, e^{-\Delta F/k_BT}\, e^{-\Delta G^*_{\text{hom}}/k_BT} \tag{5a}$$

$$J_{\text{het}}\ (\text{cm}^{-3}\,\text{s}^{-1}) \simeq 10^{5}\rho_N\, e^{-\Delta F/k_BT}\, e^{-\Delta G^*_{\text{het}}/k_BT} \tag{5b}$$

2.2. Nucleation in Heterogeneous Drops. Imagine that at $t = 0$ a drop is instantly quenched so that its nucleation rate increases from zero to J. Further assume that nucleation is a Poisson process. The probability that a single drop containing supersaturated protein has nucleated in the infinitesimal time interval between t and $t + \tau$ is proportional to τ and independent of t. At constant nucleation rate J and sample volume v, the probability of a crystal to nucleate is $Jv\tau$. At times $t \gg \tau$, the probability q that the drop has not nucleated is e^{-Jvt}. If we have N of such identical drops, then the fraction of drops that have not crystallized at time t is

$$f_\phi = \frac{N_\phi(t)}{N} = e^{-Jvt} \tag{6}$$

where $N_\phi(t)$ is the number of drops that do not have crystals at time t. Consider a system of identical drops S, each of volume v, as shown in Figure 3a, in which two nucleation pathways exist

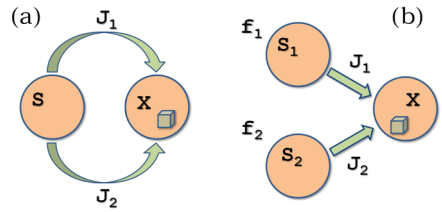


Figure 3. (a) Nucleation in a system with two nucleating pathways from a system of identical drops. (b) Nucleation in a system with two nucleating pathways from a system of two populations of drops.

(for example, homogeneous and heterogeneous pathways) with rates J_1 and J_2. If we assume that these two pathways are independent random processes, then the chemical rate equation describing the process is

$$\frac{dS}{dt} = -(J_1 + J_2)vS$$

Solving the rate equation for the fraction of drops that have not crystallized at time t yields the following form with the effective nucleation rate equal to the sum of the two nucleation rates J_1 and J_2.

$$f_\phi = e^{-(J_1+J_2)vt} \tag{7}$$

The fraction of noncrystallized drops is a single exponential despite there being multiple nucleation pathways.

In a system in which the drops are not identical, i.e., a system containing two or more types of drops, then the fraction of

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487−4509

samples that have not crystallized is no more a single exponential decay. The system shown in Figure 3b has two populations of drops S_1 and S_2 with fractions f_1 and f_2 that crystallize via nucleation pathways with rates J_1 and J_2, respectively. The chemical rate equations describing such a system are

$$\frac{dS_1}{dt} = -J_1 v S_1$$

$$\frac{dS_2}{dt} = -J_2 v S_2$$

Solving the chemical rate equations for the fraction of drops that have not crystallized at time t gives

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

In general

$$f_\phi(t) = \sum_i f_i\, e^{-J_i v t} \tag{8}$$

where f_i and J_i are the fraction of population i and the corresponding nucleation rate, respectively. Therefore, in a system containing multiple types of drops, the fraction of noncrystallized drops as a function of time is a multi-exponential.

The key assumption in this analysis is that the nucleation rate, J, is constant in time. J depends on both temperature and concentration and therefore we experimentally hold the temperature constant using active thermal feedback and hold the protein and precipitant concentrations constant by sealing samples in glass to prevent evaporation of water from the drops. Another factor that could lead to a time-dependent nucleation rate is that proteins can denature or aggregate over time, but the former can be assayed by running gels on the proteins and the latter by following the nucleation process with dynamic light scattering. In practice, with some care and effort, it is possible to find conditions for which the protein solution remains stable over the duration of the nucleation experiment. However, it is unavoidable that once nucleation occurs, monomers in solution accrete onto the growing crystal, thereby reducing supersaturation, which decreases J and thereby violates the assumption of time-independent nucleation. In this study, we circumvent this problem by measuring the probability that drops have not crystallized as a function of time, because that nucleation rate is constant for unnucleated drops.

3. EXPERIMENTAL SECTION

3.1. Formulation and Drop Generation. We produce emulsion drops using a flow-focusing microfluidic device fabricated using PDMS.[33] Figure 4a 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. The crystallization conditions in all of the drops are lysozyme (Sigma-Aldrich, cat. no. L6876) + 12.5% w/v PEG 8kD (OmniPur EMD, cat. no. 6510) + 5% w/v NaCl (Fisher Scientific, cat. no. S271-1) + 0.1 M NaAc (Fisher Scientific, cat. no. S210-500) at pH 4.8 (Thermo Orion pH meter model 330). For example, to produce emulsion drops containing 30 mg mL^{-1} lysozyme, we followed the protocol given below. To prepare samples of different protein concentrations we varied the protein concentration in step 1 below. Note that all the solutions mentioned in the protocol contain 0.1 M NaAc buffer at pH 4.8.

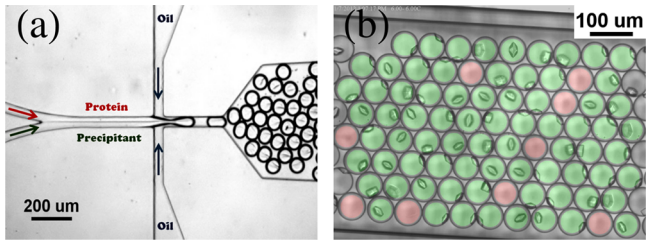


Figure 4. (a) Droplet generation using a coflow microfluidic device fabricated using PDMS; the protein and precipitant are mixed on-chip to avoid any nucleation before starting the experiment. The stream labeled protein contains lysozyme, 12.5% w/v PEG 8kD, and 0.1 M NaAc at pH 4.8, and the stream labeled precipitant contains 12.5% w/v PEG 8kD, 10% w/v NaCl, and 0.1 M NaAc at pH 4.8. See movie S1 in the Supporting Information. (b) Detected drops with crystals highlighted in green and without crystals highlighted in red. See movie S2 in the Supporting Information.

(1) Off-chip, make a 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 constitutes the "protein" stream.

(2) Off-chip, make a 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 constitutes the "precipitant" stream.

(3) For on-chip mixing, we used a coflow microfluidic device shown in Figure 4a, and the protein and precipitant streams are pumped at equal flow rates, ≈300 μL h^{-1}, 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. We used fluorinated oil, HFE-7500 (3M), containing 2% w/w EA surfactant[34] (RainDance Technologies, Inc.). The EA surfactant is a PTFE−PEG block copolymer designed to prevent adsorption of the protein to the drop−oil interface.[34,35] The oil flow rate was set at ≈600 μL h^{-1}.

It is important to note is that the lysozyme/PEG mixture obtained in step 1 was never observed to crystallize by itself, but lysozyme crystals form in seconds when the lysozyme/PEG mixture of step 1 and the PEG/NaCl mixture of step 2 are mixed at equal proportions in a 500 μL Eppendorf tube.

The homogeneity in composition of protein and precipitant mixture in the drops produced using a coflow microfluidic device is assayed by measuring the cloud point, from which we determine that drop-to-drop variation in composition is less than a few percent. See Appendix A for more discussion.

The emulsion drops are then loaded in a rectangular capillary with inner dimensions of width, 1 mm; height, 50/100 μm (VitroCom, Mountain Lakes, NJ, USA), and the ends of capillary are sealed with VALAP,[36] a mixture of equal parts of VAseline, LAnolin, and low melting temperature Paraffin wax. Sealing is very important because the drops are less than 1 nL in volume, and just a minute amount of mass transport will dehydrate the drops.

3.2. Data Acquisition and Treatment. The capillaries are scanned at regular intervals using a homemade robotic stage that can scan and acquire images of capillaries with an accuracy of 6 μm. The stage is equipped with two thermoelectric coolers (TECs) to control the temperature with a working range of −4 to 40 °C. The TECs are independently controlled, allowing the generation of linear temperature gradients across the samples. The robotic stage and the temperature are controlled using LabVIEW interface. The images of capillaries scanned at regular intervals are processed semiautomatically using MATLAB for droplet and crystal detection (Figure 4b) to obtain $f_\phi(t)$, the fraction of drops without crystals, plotted in Figure 5 along with the fit to eq 9.

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487−4509

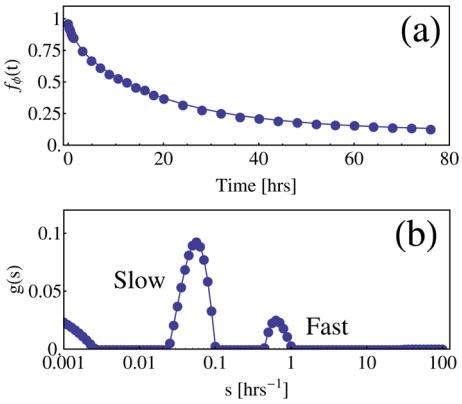


Figure 5. (a) Fraction of noncrystallized drops, $f_\phi(t)$, measured from the experiment. (b) Normalized decay distribution, $g(s)$, obtained by inverse Laplace transformation of $f_\phi(t)$. The data is obtained under crystallization conditions of 30 mg mL^{-1} lysozyme, 12.5% w/v PEG 8kD, and 5% NaCl at 9 °C.

Mathematically, $f_\phi(t)$ is the Laplace transform of the decay time distribution, $g(s)$.

$$f_\phi(t) = \mathcal{L}g(s) = \int_0^\infty g(s)\, e^{-ts}\, ds \tag{9}$$

We obtained $g(s)$ by inverse Laplace transforming $f_\phi(t)$ using a CONTIN-like algorithm written in MATLAB.[37,38] Figure 5b shows the corresponding normalized decay time distribution, $g(s)$. The calculation of $g(s)$ involves numerical computation of the inverse Laplace transform of $f_\phi(t)$, which is an ill-posed problem. See Appendix G for a detailed discussion of the Tikhonov regularization method[39] used to perform the inverse Laplace transform. The peak in $g(s)$ at small rates s, or long times, is due to the inaccuracy in establishing the baseline or long-time behavior of $f_\phi(t)$, as the experiments are stopped after 80 h. A robust feature of our analysis of $f_\phi(t)$ is two nucleation rates corresponding to the two peaks in $g(s)$.

When multiple decay modes exist in a process, the inverse Laplace transform is the most general way of obtaining the decay mode distribution. However, only two rates are present in our system. Therefore, we employed a simpler model to obtain decay rates as proposed by Pound and La Mer.[1] The Pound and La Mer model had also been employed by other research groups[40,41] in analyzing nucleation rate measurements on KNO_3 in microfluidically produced emulsions. The model, illustrated in Figure 6, consists of large number of identical drops of volume v, all of which contain a solution that nucleates at a rate, k_s. Additionally, these drops contain an average number, m, of nucleation sites per drop, which are randomly distributed among the drops. These nucleation sites are due to the presence of impurities in the sample and serve as heterogeneous nucleation centers. Nucleation from the bulk solution 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 noncrystallized drops as a function of time is given by

$$f_\phi(t) = e^{-m}\, e^{-k_s t}\, e^{me^{-k_f t}} \tag{10}$$

See Appendix B for a derivation of eq 10. When $m \ll 1$, eq 10 reduces to a two exponential process, as shown in Figure 3b. In all that follows, we extract the fast , k_f, and slow, k_s, nucleation rates as well as the average number of fast sites per drop, m, by fitting the experimentally measured fraction of drops without crystals as a function of time to eq 10. One important distinction is the nucleation rates have different volume dependencies for slow and fast nucleation. The rate of nucleation from bulk solution, k_s, scales linearly with volume, whereas the rate of nucleation from a single site, k_f, is independent of volume. The average number of impurities, m, scales linearly with volume. Note that the bulk process, characterized by k_s, could arise from either homogeneous or heterogeneous nucleation. In the latter case, there would be an extremely large number of nuclei per drop so that each drop would have close to the average number of nuclei. We performed Monte Carlo simulations using the La Mer model to estimate the errors in our method. Our simulation results, presented in Appendix B, suggest that experiments with 5, 50, 100, and 500 drops yield nucleation statistics with a confidence range of ∼200, ∼20, ∼10, and ∼5%, respectively. All of our experiments were performed with 1000−4000 drops.

When crystals grow slowly, the time interval between nucleation and observation can be significant. One can correct for this by subtracting this growth time, t_g, from the observation time, t, to obtain the distribution of drops for which no crystals have been nucleated as $f_\phi(t - t_g)$. Under the studied crystallization conditions, $t_g \approx 0.1$ h, which is less than the characteristic time scales for nucleation, $1/k_s$ (∼56 h) and $1/k_f$ (∼1.4 h). Therefore, correcting the data for t_g did not improve our results. See Appendix B for more discussion on the sensitivity of the fitted parameters to the growth time of the crystal.

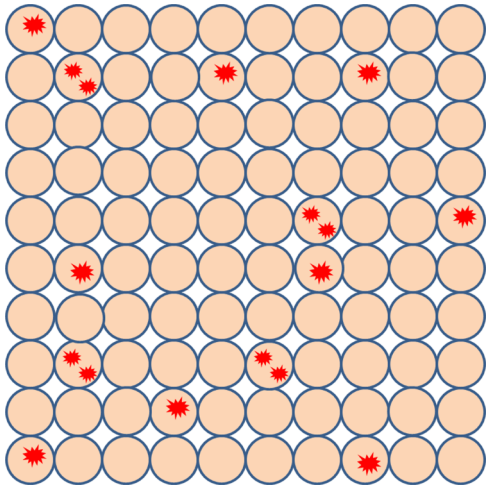


Figure 6. La Mer model.

4. RESULTS AND DISCUSSION

4.1. Measurement of Nucleation Rates. Figure 7 shows the fraction of drops with no crystals, $f_\phi(t)$, as a function of

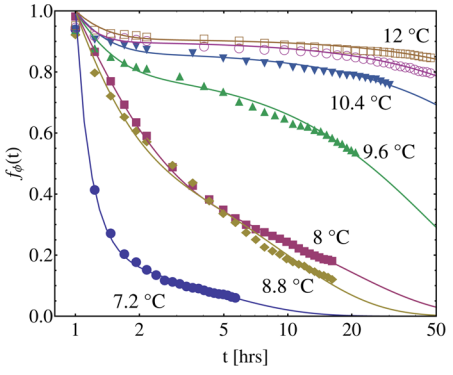


Figure 7. Fraction of drops without crystals, $f_\phi(t)$, as a function of time, t. Measured (colored symbols) and fits to the Pound and La Mer model of eq 10 (colored lines). The crystallization conditions are 30 mg mL^{-1} lysozyme, 12.5% w/v PEG 8kD, and 5% w/v NaCl in 0.1 M NaAc at pH 4.8 at different temperatures.

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487−4509

temperature at a fixed concentration and the corresponding fits to the Pound and La Mer model of eq 10 to obtain k_s, k_f, and m.

As explained in the previous section, k_s, the rate per drop, and k_f, the rate per site, have the same units of rate (t^{-1}), but they have different interpretations as

$$k_s\left(\frac{\text{rate}}{\text{drop}}\right) = v\left(\frac{\text{vol}}{\text{drop}}\right) J_s\left(\frac{\text{rate}}{\text{vol}}\right) \tag{11}$$

$$k_f\left(\frac{\text{rate}}{\text{site}}\right) = \frac{J_f\left(\frac{\text{rate}}{\text{vol}}\right)}{\rho_N\left(\frac{\text{site}}{\text{vol}}\right)} \tag{12}$$

We first state our conclusions based on a series of experiments and model fitting described in detail below. By fitting the measured fraction of drops that have not crystallized, f_ϕ, to the Pound–La Mer model of eq 10, we find that there are two nucleation rates, denoted fast and slow. Each of the rates are heterogeneous, as described by eq 4b. There are only two parameters in eq 4b that differ between the fast and slow processes: the number of nuclei, ρ_N, and the nucleation barrier, ΔG^*. In eq 4b, D and ρ are equal for the two processes, as these are properties of the monomers in solution and not of the nuclei. We argue that the ΔF is equal for the two processes. This is based on two assumptions of the CNT, as illustrated in Figure 2. The first assumption is that the precritical crystal has the same structure as the postcritical crystal. A second assumption of CNT is that the structure and growth kinetics of the fast and slow precritical crystals are identical; their difference is solely in the surface tension between the crystal and the heterogeneous substrate. Testing these assumptions represents daunting experimental challenges, which remain to be addressed. Assuming that R^* and Z being equal for the fast and slow rates is incorrect, however, we expect that the differences between these quantities for the two processes to be small compared to other factors. This allows us to relate parameters extracted from the Pound and La Mer model to parameters for slow and fast rates in the CNT as follows

$$J_s = \frac{k_s}{v} \approx 4\pi\rho_N \rho D R^* Z\, e^{-\Delta F/k_B T}\, e^{-\Delta G_s^*/k_B T} \tag{13a}$$

$$k_f \approx 4\pi\rho D R^* Z\, e^{-\Delta F/k_B T}\, e^{-\Delta G_f^*/k_B T} \tag{13b}$$

Our objective is to extract the nucleation barrier, ΔG^*, and the concentration of nuclei, ρ_N, from the experimentally obtained values J_s and k_f. We rewrite eqs 13a and 13b as a function of supersaturation as

$$J_s(\sigma) = A\cdot e^{\sigma - B/\sigma^2} \tag{14a}$$

$$k_f(\sigma) = A'\cdot e^{\sigma - B'/\sigma^2} \tag{14b}$$

where $A(A')$ and $B(B')$ are fitting parameters, $\sigma = \Delta\mu/k_B T$ is the supersaturation, C (mg mL^{-1}) = ρ (mL^{-1}) × 1000M_w (g mol^{-1})/N_A is the concentration of protein and C_s (mg mL^{-1}) is the solubility of the protein. The crystallization conditions explored in the present work range from 7.2 to 12 °C, over which the lysozyme solubility, $C_s(T)$, is temperature-dependent. At sufficiently high ionic strength, the solubility of lysozyme is independent of PEG concentration;[42,43] therefore, we estimated the lysozyme solubility with PEG based on the reported values of lysozyme solubility[44] measured without PEG at 5% w/v NaCl in 0.1 M NaAc buffer at pH 4.8 at different temperatures. Filled circles (●) in Figure 8 show the variation of lysozyme solubility over the experimental temperature range.

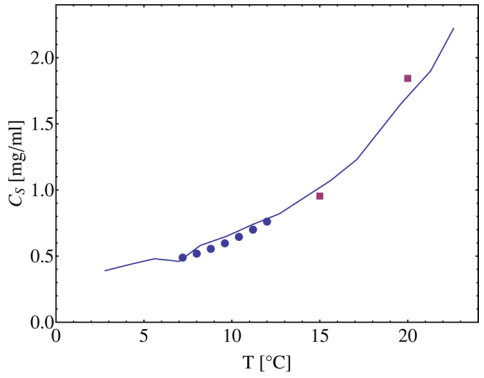


Figure 8. Lysozyme solubility vs temperature. The solid line is the lysozyme solubility measurements by Pusey[44] in 0.1 M NaAc and 5% NaCl at pH 4.8. The filled squares (■) are the measured lysozyme solubility in 0.1 M NaAc, 5% NaCl, and 12.5% PEG 8kD at pH 4.8. These measurements confirm that at sufficiently high ionic strength the solubility of lysozyme is independent of PEG concentration.[42,43]The filled circles (●) represent the interpolated values of lysozyme solubility to the studied crystallization conditions using Pusey's measurements.

We consider the protein solution to be ideal, i.e., we set the chemical potential $\sigma(T) = \Delta\mu/k_B T = \ln(C/C_s(T))$ or, equivalently, $C = C_S(T)\, e^{\sigma}$. The parameters A and A' are

$$A = \frac{4\pi\rho_N N_A C_s(T)\, DR^* Z}{1000 M_w}\, e^{-\Delta F/k_B T} \tag{15}$$

$$A' = A/\rho_N \tag{16}$$

B and B' are related to the barrier height of the slow and fast processes as follows

$$\frac{B}{\sigma^2} = \frac{\Delta G^*}{k_B T} = \frac{1}{k_B T}\frac{16\pi}{3}\frac{\gamma^3}{\rho_c^2 \Delta\mu^2} \tag{17}$$

where ΔG^* is the barrier height, γ is the interfacial tension between the crystal nucleus and the solution phase, and $\rho_c = 4.2 \times 10^{18}$ cm^{-3} is the number density of the lysozyme crystal (calculated using the dimensions of the unit cell).[45] A similar relation holds for B', with γ replaced by γ'. Figure 9 shows the measured nucleation rates, J_s and k_f, and the 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 3b.

4.2. Chemical Potential. The ideal solution approximation for the supersaturation, σ, is $\Delta\mu/k_B T = \ln(C/C_s)$ and is appropriate for dilute solutions. To estimate the correction due to the nonideality of the protein solution, we consider the virial expansion[3] for $\Delta\mu$

$$\sigma = \Delta\mu/k_B T = \ln(C/C_s) + 2B_2 M(C - C_s) \tag{18}$$

where B_2 is the second virial coefficient and M is the molecular weight of lysozyme. We estimate $B_2 = -4.85 \times 10^{-4}$ mL mol g^{-2} for the data shown in Figure 9[46] (see Table 1 in Appendix C). See Appendix C for more discussion on corrections for nonideal solution behavior. For example, the fitting parameters for the slow nucleation data shown in Figure 9 using $\sigma = \ln(C/$

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487–4509

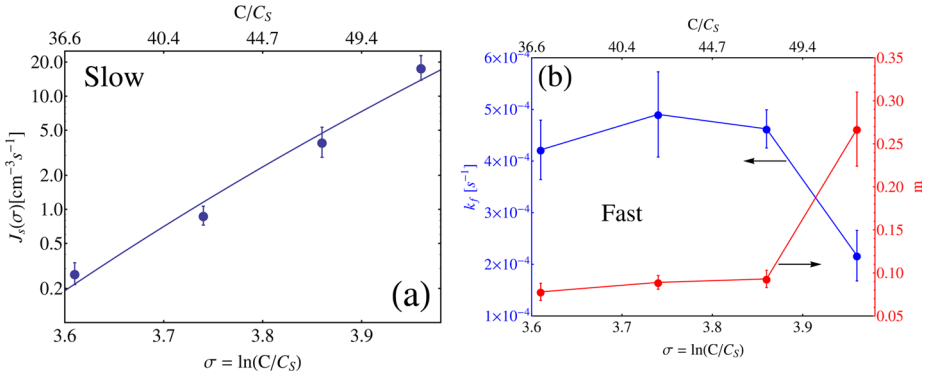


Figure 9. (a) Slow nucleation rate per volume, $J_s = k_s/v$, as a function of the supersaturation ratio, C/C_s and $\Delta\mu/k_BT = \sigma = \ln(C/C_s)$. (b) Fast nucleation rates per impurity, k_f, and average number, m, of nucleation sites per drop as a function of C/C_s and σ. Experimental conditions for both panels a and b are 21−30 mg mL^{-1} lysozyme, 5% w/v NaCl, and 12.5% PEG 8kD at 9 °C in 0.1 M NaAc buffer at pH 4.8. Each data point is obtained from more than 2000 drops, with each drop of volume < 1 nL.

C_s) are $A = 6.2 \times 10^7$ mg^{-1} s^{-1} and $B = 293.2$. When we corrected σ for nonideality using eq 18, we obtained $A = 1.2 \times 10^{12}$ mg^{-1} s^{-1} and $B = 354.1$. Although the kinetic prefactor changes by 4 orders of magnitude, we show in Appendix C that the qualitative behavior is insensitive to whether we treat solution nonideality.

First, we focus on the activation barrier, proportional to fitting parameter B, which is relatively insensitive to solution nonideality. Due to the lack of measured second virial coefficients for our exact conditions, we analyzed our data using the ideal chemical potential $\Delta\mu/k_BT = \ln(C/C_s)$. In Appendix C, we estimate that the nucleation barriers and critical nuclei sizes calculated using the nonideal chemical potential are 50% larger than the barriers estimated using the ideal chemical potential.

4.3. Fitting of Nucleation Rates. The nucleation rates are fitted to eqs 14a and 14b. One can obtain details about the nucleation barrier, ΔG^*, and interfacial tension, γ, from fitting parameter B using eq 17. The character of the nucleation mechanism, i.e., whether nucleation is homogeneous or heterogeneous, and the number density of nucleation sites can be determined from fitting parameter A using eq 15. As mentioned earlier, the robotic stage has two thermoelectric coolers (TECs) that can be operated independently in the temperature range −4 to 40 °C. By operating the TECs at two different temperatures, we generated a linear temperature gradient. We placed a series of identical capillaries oriented perpendicular to the temperature gradient in order to collect nucleation rate data at multiple temperatures. We measured the nucleation rates in the temperature range 7.2−12 °C at different concentrations of lysozyme. In the following sections, we discuss the slow and fast nucleation processes and determine the barrier and concentration of nuclei for each of the processes.

4.3.1. Slow Nucleation Barrier. Nucleation is an activated process, and the rate at which nucleation occurs is proportional to the Boltzmann weight, $e^{-\Delta G^*/k_BT}$, where ΔG^* is the activation energy or the barrier height. Figure 10 are the plots of slow nucleation rates, J_s, vs temperature at different supersaturations of lysozyme. The plots in Figure 10 show that nucleation rates increase by nearly 3 orders of magnitude as temperature is lowered by 5 °C. Our expectation was that higher nucleation rates implied a lower nucleation barrier. Unexpectedly, our analysis indicates that the barrier height

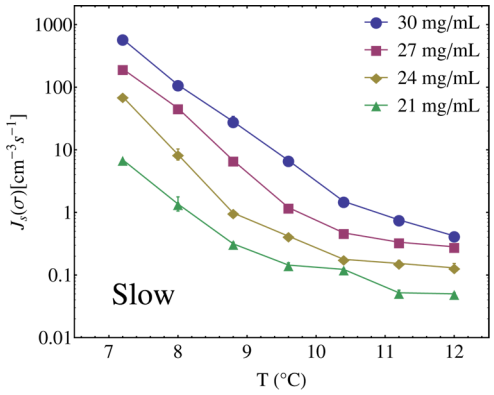


Figure 10. Slow nucleation rates vs temperature. The crystallization conditions are lysozyme, 12.5% w/v PEG 8kD, and 5% w/v NaCl in 0.1 M NaAc buffer at pH 4.8.

increases as temperature is lowered. We calculate the barrier height, ΔG^*, using eqs 15 and 17. Figure 11a is the plot of ΔG^* as a function of the ideal chemical potential for different temperatures. We have also calculated ΔG^* for the nonideal chemical potential, σ, using eq 18. We observed a qualitatively similar trend in ΔG^* as a function of temperature; however, there is a ≃50% increase in ΔG^* calculated using the corrected supersaturation (Appendix C). The nucleation rate, J_s (eqs 13a and 13b), is a product of two terms, the kinetic prefactor, A, and the Boltzmann weight, e^{-B/σ^2}, associated with the activation barrier. As temperature is lowered, fits of J_s to eqs 14a and 14b (see Appendix I for nucleation rates vs supersaturation and the corresponding fits to eqs 14a and 14b) reveal that the barrier increases and correspondingly the Boltzmann weight decreases by 5 orders of magnitude. Therefore, to satisfy eqs 13a and 13b, the kinetic prefactor must increase by 8 orders of magnitude in order to be consistent with the experimental rate measurements of Figure 10. The kinetic prefactor and the activation energy vary in opposite ways as a function of temperature, with the kinetic prefactor dominating the nucleation rate. This is contrary to conventional wisdom because nucleation is normally dominated by the barrier and not the kinetic prefactor.

Figure 11a,b are plots of the barrier heights, ΔG^*, and the 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

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487−4509

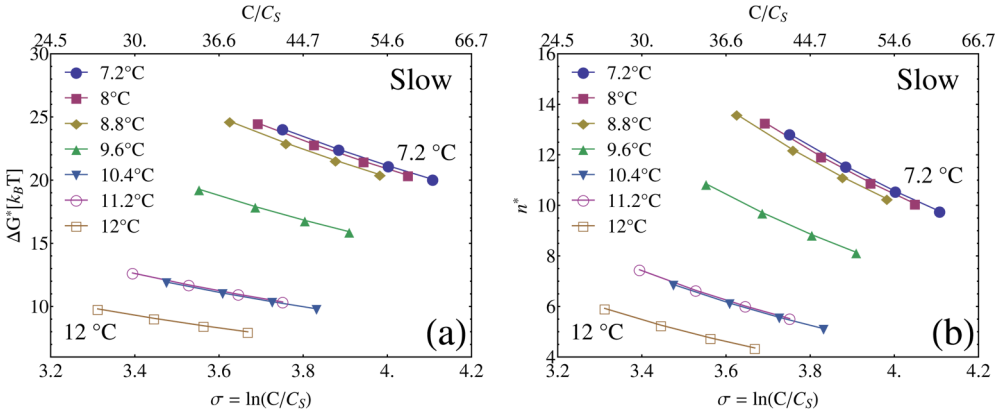


Figure 11. (a) ΔG^* vs σ and C/C_s for different temperatures. (b) n^* vs σ and C/C_s. Nucleation rates used in panels a and b are shown in Figure 10 and are measured at temperatures ranging from 7.2 to 12 °C. Lines are drawn to guide the eye.

n^* as $\Delta G^* = n^*\Delta\mu/2$.[7,47] As expected from CNT (see eq 17), ΔG^* and n^* decrease as supersaturation increases at constant temperature. As temperature increases, ΔG^* and n^* decrease. Within the context of CNT via eq 17, this is attributed to the decrease in interfacial tension, γ, with an increase in temperature. Figure 12 shows the fitted decrease in interfacial tension as temperature increases. The range of interfacial tensions that we measured, ≃0.2−0.3 mJ m^{-2}, are lower than the values reported in literature.[9,20,48] This discrepancy is due to the fact that the previous calculations of the interfacial tension used eq 17 with $\rho_c = 3.33 \times 10^{19}$ cm^{-3}, which is the inverse volume of a single lysozyme molecule, rather than using the number density of lysozyme in a crystal, $\rho_c = 4.2 \times 10^{18}$ cm^{-3}, as we have done here. If we calculate the interfacial tension using $\rho_c = 3.33 \times 10^{19}$ cm^{-3}, then we obtain $\gamma \simeq 0.7$–1.1 mJ m^{-2}, which is consistent with the previous measurements.

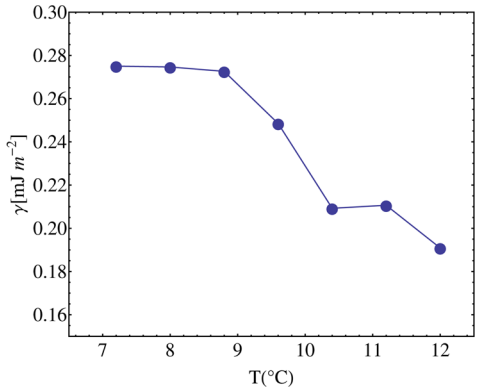


Figure 12. Interfacial tension vs temperature.

4.3.2. Fast Nucleation Barrier. Figure 13a 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 $\Delta G^* \propto (\partial \ln J/\partial\sigma)$,[47] we conclude that the nucleation barrier is unmeasurably small, i.e., $\Delta G^*/k_BT \ll 1$, for the fast nucleation process. The accuracy of the barrier height of the fast process was not good enough to discern between a spinodal process for which $\Delta G^* = 0$ and an activated process for which $\Delta G^*/k_BT \ll 1$. However, were the barrier of the fast process strictly zero, the crystallization phase transition would happen simultaneously at time $t = 0$ for all fast drops. This is not the case. Compare movie S2 for crystallization in emulsions with a finite activation barrier and movie S4 for spinodal liquid−liquid phase separation in emulsions with no activation barrier. The crystals appear at

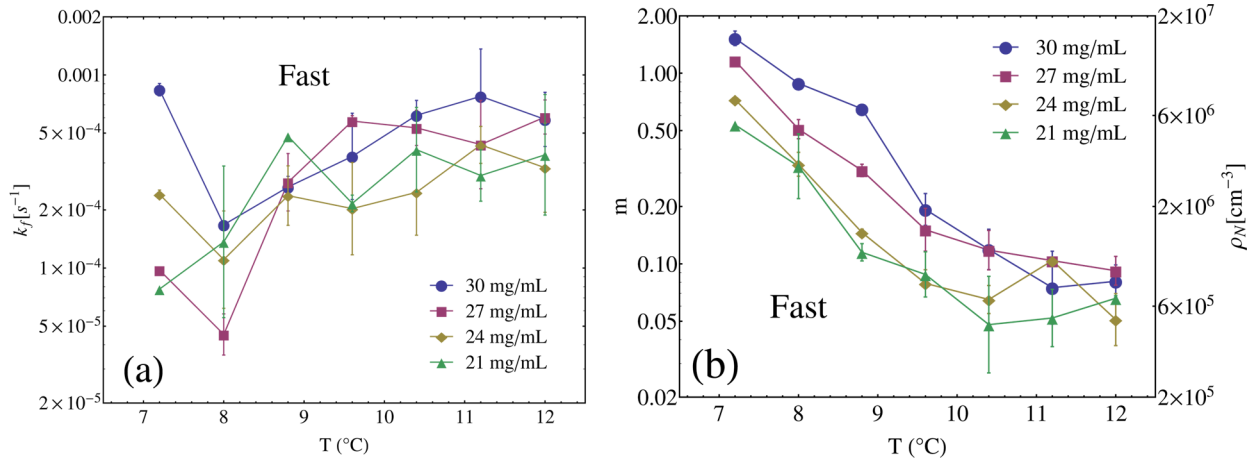


Figure 13. (a) Fast nucleation rates vs temperature. (b) Average number of nucleation sites, m, and the number density, ρ_N, of nucleation sites vs temperature. The crystallization conditions are lysozyme, 12.5% w/v PEG 8kD, and 5% w/v NaCl in 0.1 M NaAc buffer at pH 4.8.

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487−4509

random intervals, consistent with crossing a thermal activated barrier, whereas the liquid−liquid transition occurs simultaneously in all drops once the temperature crosses the spinodal line.

4.3.3. Fast Kinetic Prefactor. The number density of impurities in the fast process can be determined by measuring the average number of impurities per drop, m. As temperature is lowered, m increases, as shown in Figure 13b, which suggests creation of nucleation sites at lower temperatures. Figure 14

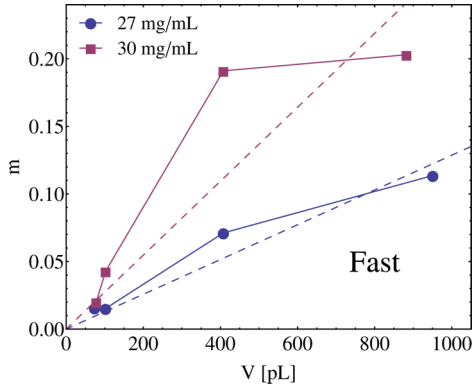


Figure 14. Average number of nucleation sites per drop, m, as a function of volume at 9 °C. The measured nucleation rates are J_s = 17.46 ± 3.41 cm^{-3} s^{-1} and k_f = (0.17 ± 0.02) × 10^{-3} s^{-1} for 30 mg mL^{-1}, and J_s = 3.41 ± 0.7 cm^{-3} s^{-1} and k_f = (0.77 ± 0.3) × 10^{-3} s^{-1} for 27 mg mL^{-1}.

shows m as a function of volume of the drop measured under two different crystallization conditions. The figure indicates that the number of impurities per drop is less than 1 and vanishes when the volume of the drop is sufficiently small. If the impurity concentration is a function of sample preparation and independent of drop volume, then the average number of impurities per drop will scale linearly in drop volume, as observed. The dotted lines represent the expected behavior in m as volume decreases. The kinetic prefactor for the fast nucleation process from eq 13b is $\rho DR^*Z\ e^{-\Delta F/kT}$. The measured value of prefactor taken from Figure 13a is ≈5 × 10^{-4} s^{-1} and is approximately independent of temperature. Using typical estimates of the monomer concentration ρ = 10^{18} cm^{-3}, D = 10^{-6} cm^2 s^{-1}, R^* = 3.5 × 10^{-7} cm, and Z = 0.2 for lysozyme crystallization trials, we calculate $\Delta F \simeq 19k_BT$ for the fast process. If the structures of the precritical nuclei of the fast and slow processes are the same, which are the assumptions of CNT as illustrated in Figure 2, then $\Delta F \simeq 19k_BT$ holds for both the slow and fast processes.

4.3.4. Slow Kinetic Prefactor. The data shown in Figure 15 are those terms in the slow process prefactor of eq 15 that are not independently determined, i.e., $\rho_N\ e^{-\Delta F/k_BT}$, as a function of temperature. In the slow process, the product $\rho_N\ e^{-\Delta F/k_BT}$ varies approximately 8 orders of magnitude from ∼10^{-2} to 10^6 cm^{-3} over 5 °C. The temperature variation could result from the temperature sensitivity of $\rho_N(T)$ or $\Delta F(T)/k_BT$ or both. In the previous paragraph, we argued that ΔF for the fast nuclei is temperature-independent. Now, we argue the same for slow nuclei. CNT assumes that the form of the pre- and postcritical nucleus is the same. If this is true, then as a corollary we also assume that the growth kinetics of a pre- and postcritical cluster are identical. This reasoning leads to the conclusion that if the temperature variation in the kinetic prefactor of a precritical

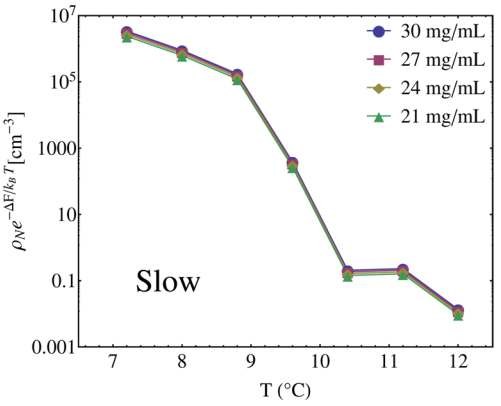


Figure 15. Prefactors obtained by fitting the measured nucleation rates to eqs (14) at every temperature. The crystallization conditions are lysozyme, 12.5% w/v PEG 8kD, and 5% w/v NaCl in 0.1 M NaAc buffer at pH 4.8.

nuclei is due to a temperature dependence of $\Delta F(T)/k_BT$, the activation barrier for the addition of a molecule to an existing cluster, then a similar temperature dependence of ΔF should be observed in the growth rates of the postcritical cluster (crystal) because ΔF is the rate-limiting step in determining the crystal growth rate. We measured growth rates, shown in Figure 16

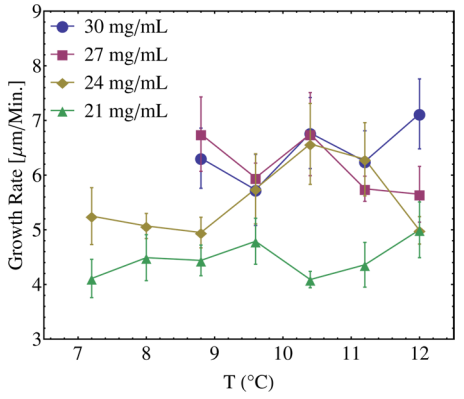


Figure 16. Measured growth rates vs temperature. The growth rates do not considerably change as a function of temperature suggesting that ΔF is constant. The rates approximately vary linearly with concentration.

and discussed in more detail in Appendix F, of crystals under the same crystallization conditions and in the same drops in which the nucleation rate experiments were performed, and the growth rates do not vary significantly with temperature. Therefore, ΔF of the postcritical crystal is independent of temperature. We can also infer from Figure 16 that the growth rates vary approximately linearly with lysozyme concentration. The data are consistent with a growth rate that is approximately proportional to supersaturation, which is expected for diffusion-dominated growth. If we accept the assumption that the pre- and postcritical crystal kinetics are identical, then the variation in the slow prefactor must arise from an increase in the number density of nucleation sites as temperature is lowered. 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 illustrated schematically in Figure 17.

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487−4509

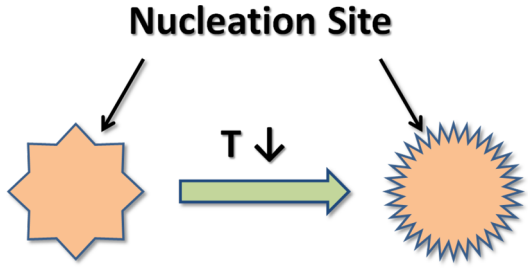


Figure 17. Creation/activation of nucleation sites. In the model, there are clusters with active sites represented by vertices. On lowering temperature, the number of clusters remains constant, but the number of active sites per cluster increases.

Another notable point is that although the slow and fast processes represent two nucleation pathways leading to crystallization, we measure only one rate of growth of the postcritical nuclei (crystals). Additionally, X-ray studies, which are described in a manuscript submitted elsewhere, on these same crystals reveal that the crystals formed by the two processes are identical. While this proves that the postcritical structures of the crystals formed from the fast and slow processes are the same, it does not prove whether the precritical structures are the same. One should keep in mind the possibility that the assumptions of the CNT, as illustrated in Figure 2, are wrong and that the structure and growth kinetics of the precritical nucleus and postcritical nucleus are different. In such a case, there could be a large ΔF for addition of a monomer to a precritical nucleus and a small ΔF for addition to a postcritical nucleus.

4.4. Impurities. *4.4.1. Impurities Causing Slow Process.* In all of our experiments, as discussed previously in section 3.1 on formulation, we prepared, off-chip, a 1:1 mixture of 80−120 mg mL^{-1} protein and 25% w/v PEG 8kD to create the protein stream that was subsequently mixed on-chip with a precipitant inside the microfluidic drop-generating device. The off-chip mixture of PEG and protein contains low salt and does not produce crystals. However, aggregates are formed at the moment of mixing, when the high concentration protein and high concentration PEG streams first come into contact. Once homogenized, which occurs in under 1 s, the mixture is stable, and no further aggregation occurs. The size of the aggregates range from 0.1−10 μm, as measured by dynamic light scattering (DLS) measurements and also using differential interference contrast (DIC) and fluorescence microscopy. See Appendix E for more discussion on characterizing protein aggregates. All experimental results reported in this research work are obtained with this unfiltered protein stream, which contains protein aggregates, except when otherwise noted.

In order to assess whether the aggregates in the low-salt PEG/lysozyme solution prepared off-chip as mentioned in the previous paragraph are in equilibrium, the mixture was filtered through 0.22 μm cellulose acetate filters to remove the aggregates. After filtration, additional aggregates did not form, even after 48 h, indicating that the aggregates are not in thermodynamic equilibrium with the lysozyme monomers but are in a dynamically arrested state. This observation is evidence that the aggregates were formed during mixing of the concentrated PEG and protein solutions. The aggregation is presumed to be caused by the depletion interaction, but depletion is not strong enough to form additional lysozyme aggregates once the equilibrium concentrations are attained. As the temperature was lowered, the aggregates do not undergo any measurable growth. Additionally, 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. We measured the concentration of lysozyme using Thermo Scientific nano UV−vis spectrophotometer with lysozyme extinction coefficient ϵ = 2.64 mL mg^{-1} cm^{-1} at 280 nm.

We produced the emulsion drops using the filtered protein stream and performed nucleation rate measurements on the filtered samples. Once again, we obtain two nucleation rates, slow and fast; however, the slow rates are at least 4 times slower than the slow rates obtained from unfiltered samples. This measurement independently confirms that slow nucleation is heterogeneous. If the nucleation were homogeneous, then

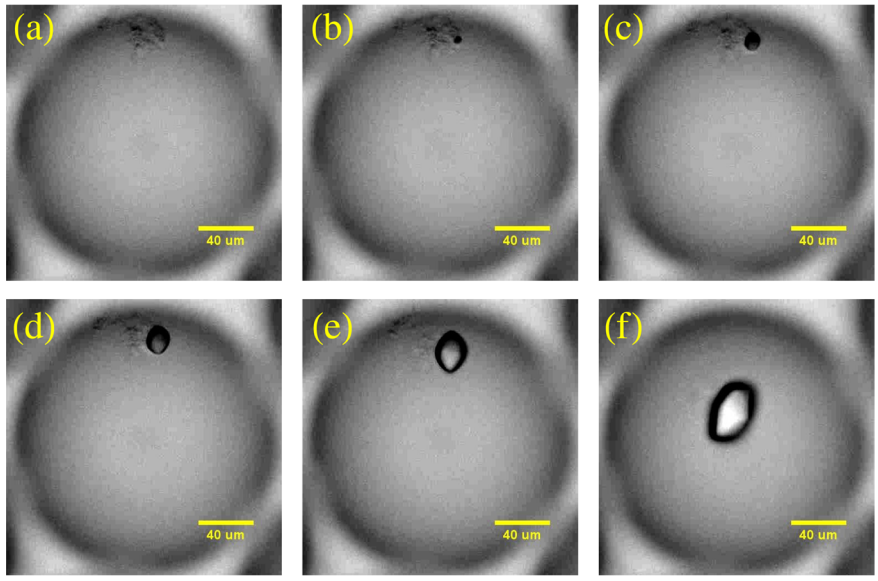


Figure 18. Nucleation and growth of a crystal from a protein dense aggregate/gel at (a) t = 0, (b) t = 15 s, (c) t = 45 s, (d) t = 75 s, (e) t = 120 s, and (f) t = 300 s. The aggregate dissolves as the crystal grows. See movie S3 in the Supporting Information.

filtering aggregates would not decrease the nucleation rate because homogeneous nucleation is only a function of monomer concentration, which is unaffected by filtration. Notably, the average number of impurities per sample, m, resulting in fast nucleation in the filtered samples does not change considerably from that of 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 do not know the mechanism for the induction of fast nucleation by unfilterable impurities. One possibility is that the surface energy required for wetting the unfilterable impurities might be smaller than the required for the lysozyme aggregates. Alternatively, it was found in simulations by Sear[49] that porous impurities enhance nucleation when the dimensions of the pore are comparable to the protein of interest. Therefore, a second possibility is that the mean pore size of unfilterable impurities is comparable to the molecular lysozyme.

Using optical microscopy (see movie S3 in the Supporting Information), we observed that nucleation always occurs on a protein aggregate and that as the crystal grows, the protein aggregate is consumed. Figure 18 shows the nucleation and growth of a crystal from a protein aggregate. We have observed the crystals that appear at later times (≥2 h) are almost always clearly associated with protein aggregates. Because after 2 h the probability that a nucleation event is due to a fast process is very low, we identify the aggregates as the source of the slow nucleation process observed in our experiments.

Figure 15 shows the measured prefactors as a function of temperature at different concentrations of lysozyme. Note the continuous increase in the prefactor as temperature lowered, which we suggest is a result of the activation of more nucleation sites and not an increase in the number of aggregates, as depicted in Figure 16. We also calculated the number density, ρ_N, of impurities causing the slow process as a function of temperature using an estimated $\Delta F \simeq 19k_BT$ (estimation explained in section 4.3.3). Figure 19 is the plot of ρ_N for the slow process as a function of temperature calculated for different supersaturations of lysozyme. The prediction for homogeneous nucleation is that ρ_N is independent of temperature at a concentration $\rho_N = \rho = 10^{18}$ cm^{-3}. At all temperatures, we observe $\rho_N \ll \rho$, which is consistent with the claim that the slow process is heterogeneous.

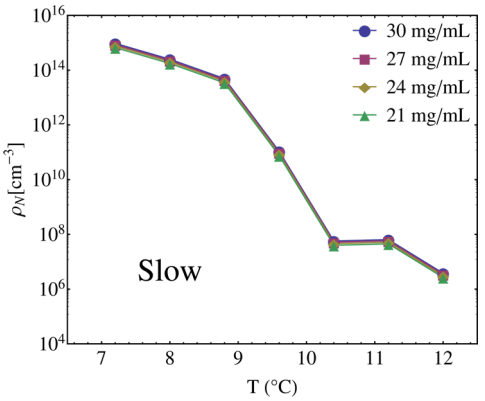


Figure 19. Number density, ρ_N, of nucleation sites as a function of temperature calculated at different supersaturations of lysozyme. Calculated assuming $\Delta F = 19k_BT$. If nucleation was homogeneous, then ρ_N would be 10^{18} (cm^{-3}) and independent of temperature.

4.4.2. Impurities Causing Fast Process. Our experiments, analyzed above, suggest that there are two heterogeneous processes in lysozyme crystallization. We attribute the slow process to nucleation on large protein aggregates and the fast process to nonfilterable aggregates. The number of fast nuclei increases rapidly with lowering temperature, as shown in Figure 13b.

4.5. Validation of Technique and Comparison with Previous Experiments. Galkin and Vekilov[9] developed a drop-based method to measure nucleation rates of protein crystals that involves nucleating the supersaturated protein solution for a quench period of Δt at temperature T_1 and subsequently growing the nucleated crystals at temperature $T_2 > T_1$. Under the assumption that the time scale for growth is much longer than the quench period, nucleation occurs only at temperature T_1, whereas at temperature T_2, no nucleation occurs, but growth of the nucleated crystals continues, i.e., nucleation and growth are decoupled. As long as the nucleated crystals remain small throughout interval T_1, the supersaturation and hence nucleation rate remains constant and further nucleation from bulk is not suppressed. The average number of crystals formed per drop is given by $\langle N \rangle = N_{hetero} + Jv\Delta t$, 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 a later time, $\Delta t > 0$, to be due to homogeneous nucleation. Galkin and Vekilov claim that heterogeneous nucleation primarily occurs at the droplet−oil interface, but we speculate disordered lysozyme aggregates serve as heterogeneous nucleation centers. See Appendix D for more discussion on heterogeneous nucleation at the droplet−oil interface. Galkin and Vekilov[17] also observed an increase in N_{hetero} with increasing supersaturation, consistent with our speculation that the heterogeneous nucleation sites are protein aggregates.

We have performed nucleation rate measurements under crystallization conditions similar to those of Galkin and Vekilov, i.e., with salt, but no PEG, using our drop-based method. We fitted the fraction of drops that do not have crystals to eq 10 and obtained two nucleation rates, namely, fast and slow. The slow nucleation rate was in agreement with the reported nucleation rates of Galkin and Vekilov. Figure 20 shows the nucleation rate vs supersaturation obtained by Galkin and Vekilov. The filled circles (●) are the measurements by Galkin and Vekilov, and the continuous line is the fit to nucleation rate described by CNT. The filled square(■) is our measurement of the slow nucleation rate performed using Turnbull's method of measuring $f_\phi(t)$. The fact that both data sets fall on the same line demonstrates that Vekilov's and our measurements are consistent and serves as a validation of the methods. Galkin and Vekilov[17] estimated the average number N_{hetero} of impurities per drop to be 0.2. The measured average number, m, of impurities per sample in our experiments was ≃0.12 per drop, which also is consistent with Galkin and Vekilov's experiments.[17] In Galkin and Vekilov's experiment, the fast rate was too rapid for them to observe, whereas our method was able to resolve the fast rate. We note that the results of the two methods are completely consistent. Using the data in Galkin and Vekilov,[17] we calculated the kinetic prefactor, $\rho_N e^{-\Delta F/k_BT} \approx 10^{-3}$ cm^{-3}, which is orders of magnitude less than $\rho \approx 10^{18}$ cm^{-3}, which is the monomer concentration, or, equivalently, the concentration of sites available for homogeneous nucleation

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487−4509

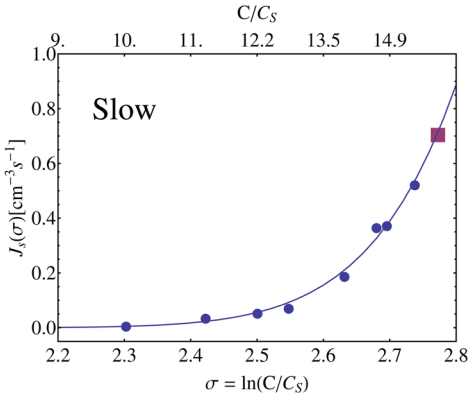


Figure 20. Measurement of nucleation rate under conditions mentioned by Galkin and Vekilov.[9] The crystallization conditions are lysozyme and 2.5% w/v NaCl in 0.05 M NaAc at pH 4.5, T = 12.6 °C. The filled square (■) is the nucleation rate obtained using the Turnbull method described in this article. Filled circles (●) are the measurements by Galkin and Vekilov,[9] and the continuous line is the fit to nucleation rate described by Classical Nucleation Theory.

assuming that $\Delta F = 0$. Vekilov attributed the small value of the kinetic prefactor as evidence for a two-step mechanism,[50] but we raise the possibility that heterogeneous nucleation accounts for the kinetic prefactor being orders of magnitude smaller than the diffusion-limited rate predicted for homogeneous nucleation. The distinction between two-step nucleation and heterogeneous nucleation is almost a matter of semantics. In two-step nucleation, dense disordered aggregates of protein form first and then crystals are nucleated by these aggregates. It is possible that the disordered aggregates are a metastable liquid–liquid phase, but they could also be disordered nonequilibrium aggregates or gel particles. In this latter case, the distinction between heterogeneous nucleation from precipitates versus two-step nucleation becomes blurred. The most direct evidence for heterogeneous nucleation in our experiments with PEG/salt/lysozyme is the observation of a permanent reduction in the slow nucleation rate with filtration and the direct observation of nucleation on protein aggregates. This rules out the possibility of a reversible two-step process in the PEG/salt/lysozyme case, but it does not rule out a liquid–liquid two-step process in the salt/lysozyme case studied by Vekilov.

5. CONCLUSIONS

We have developed an emulsion based technique for the measurement of protein crystal nucleation rates based on counting the number of drops without crystals as a function of time after a quench to deep supersaturation. The emulsion drops are produced using a flow-focusing microfluidic device. We applied the technique to measure the nucleation rates of lysozyme crystals. We found that on the order of 1000 drops are needed to obtain good statistics of a nucleation rate. We observe two nucleation rates within the drop population. The measured nucleation rates, denoted slow and fast, are analyzed according to CNT. The Pound and La Mer model[1] ascribes the fast nucleation rate to impurities present in a small subset of the drop population, but it makes no prediction on whether the slow nucleation rate is due to homogeneous or heterogeneous nucleation. A study of the kinetic prefactor of the slow rate suggests that it is also due to heterogeneous nucleation. From the CNT analysis, we extracted the barrier heights and size of critical nuclei associated with the nucleation processes. As the temperature was 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, counterintuitively, the CNT analysis reveals that the barrier heights increase as the temperature is lowered. As the temperature is lowered by 5 °C, the kinetic prefactor increases 8 orders of magnitude, whereas the barrier produces a decrease of 5 orders of magnitude in the nucleation rate. This was surprising, as the kinetic prefactor is often considered to be a weak function of temperature,[11] whereas in this case, the prefactor dominates the nucleation rates. The unexpected nature of this result calls for careful scrutiny. We have also successfully validated our technique by comparing it to the experiments of Galkin and Vekilov.[9] We believe this method can be applied to a wide variety of proteins and can be used to answer a key question facing the protein crystallization field: To what extent is nucleation heterogeneous?

■ APPENDIX

A. Chemical Homogeneity in the Drops

Drops are generated using a coflow microfluidic device with two aqueous inputs, as shown in Figure 21.[33] This device allowed us to mix protein and precipitant solutions at high supersaturation milliseconds prior to making the drops. At high supersaturation, the nucleation frequency per unit volume is large, but if the drops are small enough, then the nucleation rate per drop is low. Since the mixing is done on chip, we need to ensure that all drops are identical. We selected the cloud point, also known as the liquid–liquid phase transition, as our probe for drop homogeneity. Our tests, described below, indicate a compositional variation in our drops of less than 3%.

To test the compositional variation during in drop making, the chemical conditions are chosen such that the protein does not crystallize during off-chip mixing (in bulk), which allows us

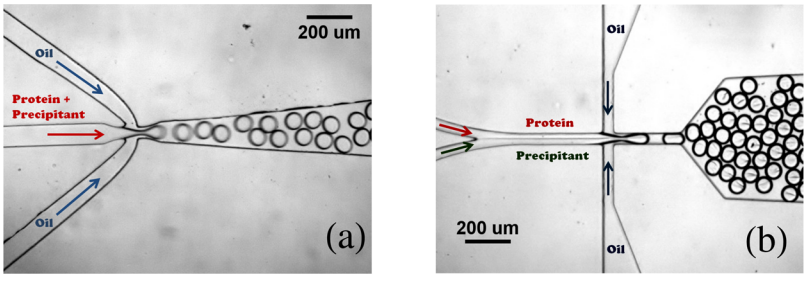


Figure 21. Generation of emulsion drops using (a) off-chip mixing and (b) on-chip mixing microfluidic devices. Note that both microfluidic chips are fabricated using PDMS.

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487–4509

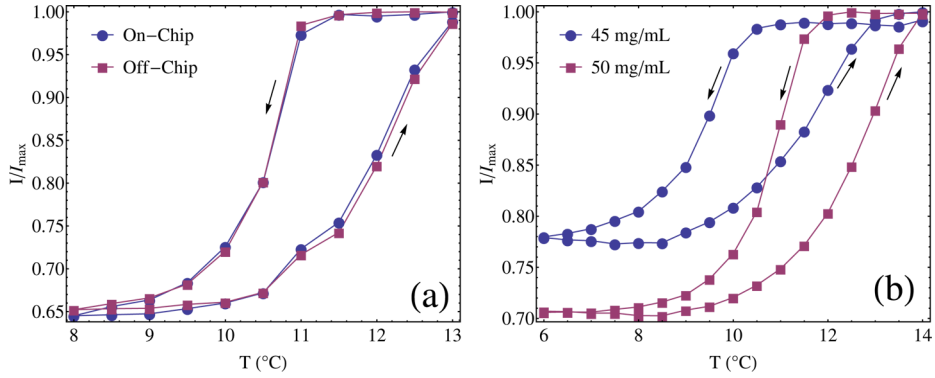


Figure 22. Average fractional intensity of drops while lowering and raising temperature. (a) Comparison between on-chip vs off-chip mixing of protein and precipitant. In both cases, the final composition in the drops is 50 mg mL^{-1} lysozyme, 5% w/v PEG 8kD, and 0.5 M NaCl in 0.2 M phosphate buffer at pH 6.2. (b) Sensitivity of the cloud point temperature to protein concentration.

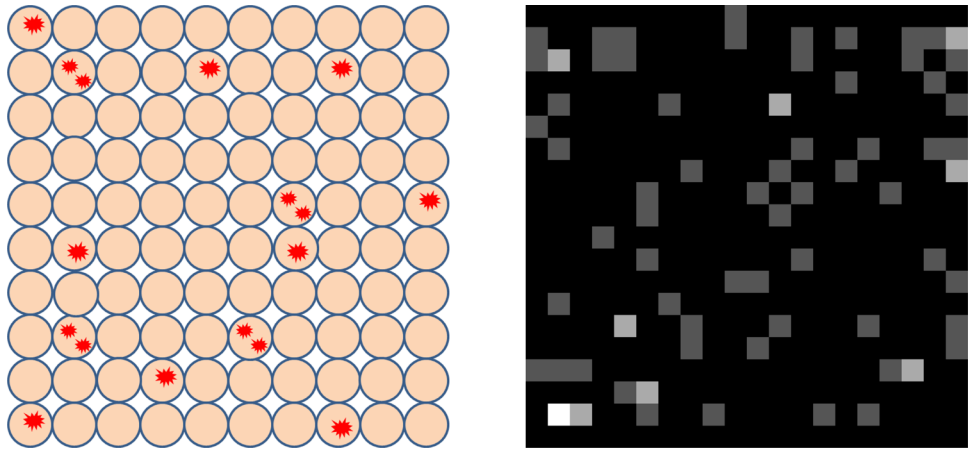


Figure 23. Initial conditions of simulation. Drops are independent and nucleate through two pathways. One pathway has the same probability for each drop, and a second pathway is proportional to the number of impurities in each drop. In both images, the average number of impurities is m = 0.2. (Left) Schematic of Pound and La Mer model. Drops contain a uniform, supersaturated solution, as well as a few randomly distributed nuclei. (Right) 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.

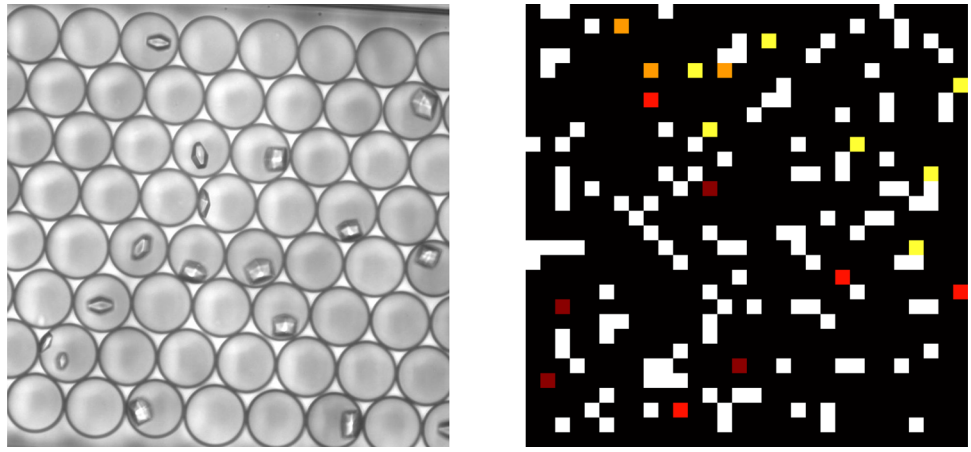


Figure 24. (Left) Nucleating drops in an experiment. (Right) Nucleating drops in simulation. Black and white pixels correspond to noncrystallized and crystallized drops, respectively, and any other colored pixel is a nucleated drop but the crystal has not yet grown to the detectable size.

to compare the mixing on-chip vs off-chip. In both cases, the intended final composition in the drops is 50 mg mL^{-1} lysozyme, 5% w/v PEG 8kD, and 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 concept behind this test is that all drops generated with off-chip mixing will have the same composition. If the microfluidic flows are unsteady, then the drops will be different sizes, but the composition will be identical. In contrast, for the drops formed with on-chip mixing, if the flows are unsteady, then drops will have different amounts

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487−4509

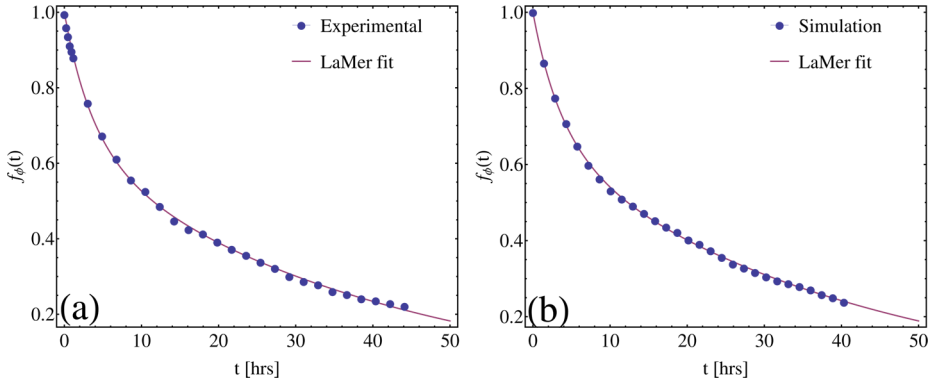


Figure 25. (a) Experimentally obtained fraction of noncrystallized drops and the fit to eq 20 from an experiment with k_0 = 0.02504 nL^{-1} h^{-1}, k = 0.2024 h^{-1}, and m = 0.4506. (b) Fraction of noncrystallized and the fit to eq 20 from the Monte Carlo simulation also with k_0 = 0.02504 nL^{-1} h^{-1}, k = 0.2024 h^{-1}, and m = 0.4506 and lag time t_g = 0.1 h. The parameters obtained from fit to eq 20 are k_0 = 0.02539 nL^{-1} h^{-1}, k = 0.1814 h^{-1}, and m = 0.4509. The simulation was not corrected for the lag time.

of protein and precipitant stream and thus the concentration will vary from drop to drop. After creating the drops in these two methods, drops were stored in sealed capillaries on a temperature stage. The temperature of the drops was lowered in steps of 0.5 °C until we reached the cloud point temperature, and it was then raised in steps of 0.5 °C until the drops were clear. At each temperature step, the drops were imaged. Figure 22a 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 $T_{\downarrow}^{C}$ and $T_{\uparrow}^{C}$, where $T_{\downarrow}^{C}$ (or $T_{\uparrow}^{C}$) is the temperature at which the drops are half as cloudy as they could become 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 produces no detectable compositional variation.

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 coflow microfluidic device. Figure 22b shows the cloud point hysteresis measured in two sets of drops with a 10% difference in protein concentration. The measured cloud points are ≈1°C apart, indicating that our measurement is sensitive to changes in composition of approximately 3%.

B. Monte Carlo Simulation of Nucleating Drops

We performed simple Monte Carlo simulations of nucleating drops to estimate the errors associated with our experiment. One error we are concerned with is the scan time, t_s, which is the time taken by the robotic stage to acquire images of all the drops. A second source of error we modeled is the lag time, t_g, which is the characteristic time for a nucleus to grow to a detectable size (≈ 5 mm). A third error we wish to model is the effect of finite sample size on accuracy. The simulation is based on a model proposed by Pound and La Mer.[1] Our model, shown in Figure 23a, consists of large number of drops containing an average number, m, of nucleation sites per drop, which are randomly distributed among the drops. Nucleation from the drops that do not have any nucleation sites occurs at a rate k_0, and nucleation from a drop containing p nucleation sites proceeds with rate $k_0 + pk$, i.e., nucleation can occur from

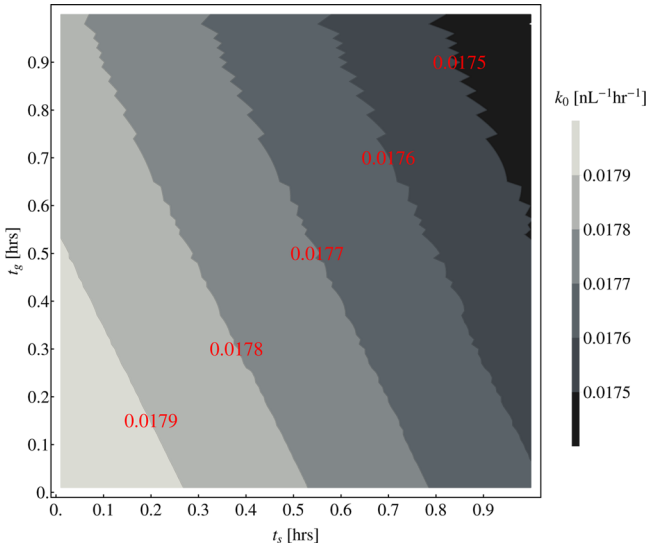


Figure 26. Contour plot illustrates the effect of scan time, t_s, and growth time, t_g, on the measurement of nucleation rate k_0. The nucleation rate k_0 is weakly affected by the scan and/or growth times because the characteristic nucleation time, $1/k_0 \sim 56$ h is much longer than t_s (~0.1 h) and t_g (~0.1 h). The simulation is performed with k_0 = 0.018 nL^{-1} h^{-1}, k = 0.72 h^{-1}, and m = 0.2.

the solution as well as from the sites, where k is the rate of nucleation from a single nucleation site. Our model is slightly different than Pound and La Mer's original model,[1] which considered that nucleation from a drop containing p nucleation sites proceeds with rate kp, i.e., nucleation occurs only from sites in drops that contain sites and not from solution. Assuming that the nucleation sites are distributed according to the Poisson distribution, the probability, s, that a drop contains p nucleation sites is

$$s(p) = \frac{e^{-m} m^{p}}{p!} \tag{19}$$

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

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487−4509

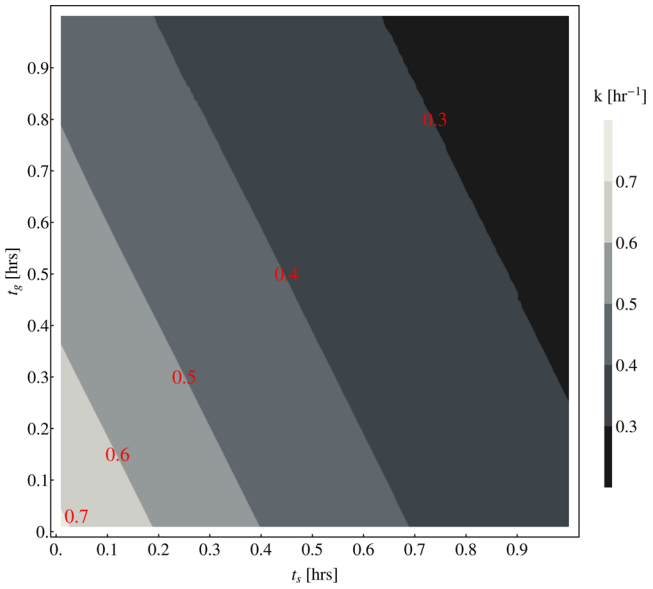


Figure 27. Contour plot shows the effect of scan time, t_s, and growth time, t_g, on the measurement of nucleation rate k. At smaller t_s and t_g, the error in the nucleation rate k is small ,but as $t_s \sim 1/k$ and/or $t_g \sim 1/k$, the error is large. Note that the error in k is about 10% for typical experimental values of t_s (∼0.1 h) and t_g (∼0.1 h). The simulation is performed with k_0 = 0.018 nL^{-1} h^{-1}, k = 0.72 h^{-1}, and m = 0.2.

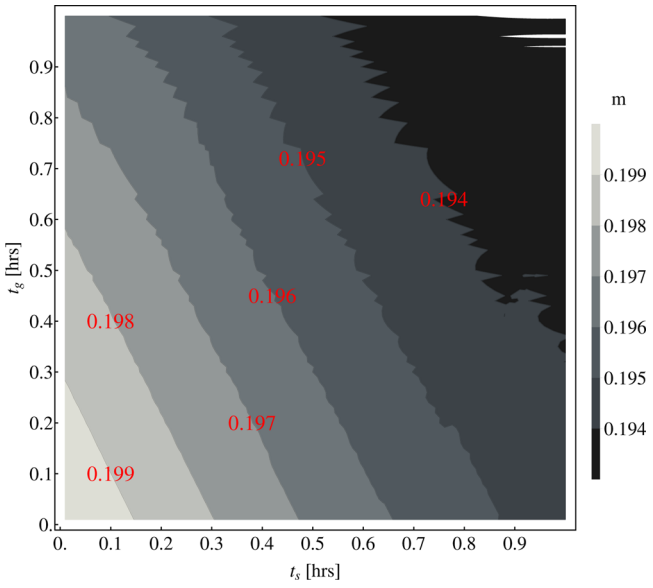


Figure 28. Contour plot shows the average number, m, of impurities per drop as a function of scan time, t_s, and growth time, t_g. The parameter m weakly depends on scan time and growth time. The simulation of $f_\phi(t)$ is performed with k_0 = 0.018 nL^{-1} h^{-1}, k = 0.72 h^{-1}, and m = 0.2.

$$f_\phi(t) = \sum_{p=0}^{\infty} s(p)\, e^{-(k_0+pk)t} = \sum_{p=0}^{\infty} \frac{e^{-m} m^p}{p!} e^{-(k_0+pk)t} = e^{-m}\, e^{-k_0 t}\, e^{m\, e^{-kt}} \quad (20)$$

Figure 23b represents the simulation model in which each pixel is 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 (noncrystallized drops) follows eq 20. The lag time, 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 24b shows the state of the simulation model at an intermediate time, 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 lag time, $t_g \sim 0.1$ h, corresponds to the experimental value for lysozyme under the conditions we studied. Figure 25 shows the fraction of noncrystallized drops as a function of time from an experiment and the corresponding simulation. In a typical experiment, thousands of drops are stored in a rectangular capillary and are scanned at regular intervals ranging from ∼0.25 h to 2 h using a home-built robotic stage. The time required to complete one scan consisting of 40 locations amounting to a total of 4000 drops is t_s = 0.1 h. To model the effect that a finite scan time has on our analysis of the nucleation experiments, we assign all drops that crystallized during the scan as having occurred at the beginning of the scan. While the simulation proceeds at discrete times, the intervals are very small compared to the scan time.

Our method to assess errors consists of simulating $f_\phi(t)$ with a fixed set of parameters; k_0 = 0.018 nL^{-1} h^{-1}, k = 0.72 h^{-1}, and m = 0.2, which are typical values from our experiments. These simulations are performed as we vary, one at a time, the following three quantities: (1) the numbers of drops in the sample, (2) t_g, and (3) t_s. We then fit the simulation results to eq 20 in order to obtain fitted values for k_0, k, and m, which we then compare with the values used in the simulation to determine how sensitive the fits are to finite sample sizes, lag times, and scan times, as shown in Figures 26, 27, and 28.

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 ∼1200−4000 drops.

Conclusion 2. When t_s is much smaller than the characteristic nucleation times, the errors are negligible, but when t_s is comparable to the characteristic nucleation times, the errors are large. This effect is observed in Figures 27 when $t_s > 1/k$, where we define the characteristic nucleation time as the inverse of the nucleation rate. From our experiments, the scan time, t_s = 0.1 h, is much smaller than the typical characteristic nucleation times ∼56 h and ∼1.4 h corresponding to k_0 = 0.018 nL^{-1} h^{-1} and k = 0.72 h^{-1} respectively. Therefore, we expect to have less than 10% error in our measurement of the fast nucleation rate.

Conclusion 3. For a finite lag time t_g, corresponding to slow crystal growth, the error can be minimized by correcting for lag time by substituting $f_\phi(t)$ for $f_\phi(t - t_g)$, as shown in Figure 29. In our experiments, the correction did not improve the fit because $t_s \sim t_g$ and is so small as to be negligible.

C. Corrections Due to Non-ideal Solution Behaviour

The definition of supersaturation $\Delta\mu/k_BT = \ln(C/C_s)$ is an approximation appropriate for an ideal protein solution. To estimate the correction due to non-ideality of the protein solution,[3] we consider the second virial coefficient for $\Delta\mu$

$$\sigma = \Delta\mu/k_BT = \ln(C/C_S) + 2B_2M(C - C_s) \quad (21)$$

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487−4509

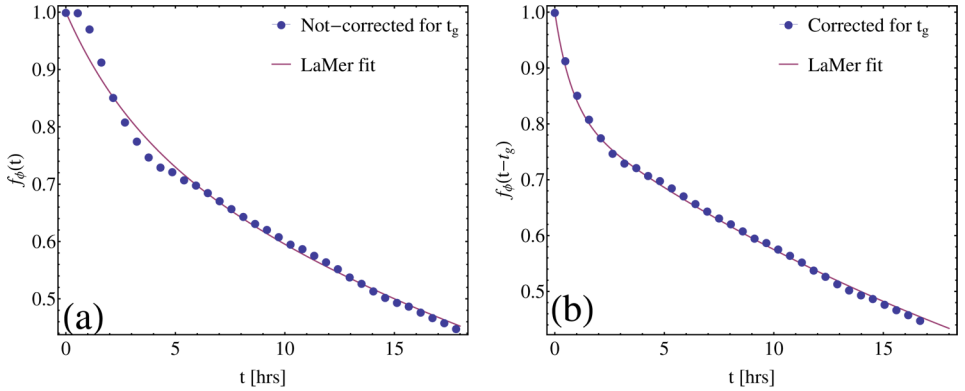


Figure 29. Correcting for finite lag time, t_g, by rescaling time. The nucleation simulation was performed with the number of drops N = 1600, k_0 = 0.036 nL^{-1} h^{-1}, k = 1.44 h^{-1}, m = 0.2, t_g = 1.2 h, and t_s = 0 h. (a) Fraction of drops with no crystals vs time, $f_\phi(t)$ and the La Mer fit, with fit parameters k_0 = 0.03268 nL^{-1} hr^{-1}, k = 0.265 h^{-1}, and m = 0.2065. (b) Fraction of drops with no crystals vs rescaled time, $f_\phi(t - t_g)$ and the La Mer fit, with fit parameters k_0 = 0.0352 ± 0.0002 nL^{-1} h^{-1}, k = 1.43 ± 0.0504 h^{-1}, and m = 0.2 ± 0.0036.

Table 1. Estimated Second Virial Coefficients for Crystallization Conditions[a]

T (°C)	B_2 (mL mol g^{-2})
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}

[a]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.[46].

where B_2 is the second virial coefficient, M is the molecular weight of lysozyme, and C_s is the lysozyme solubility. The virial coefficient has not been measured for the conditions used in our experiments. However, the virial coefficients were measured under similar conditions to our crystallization conditions, and we will use these values to approximate virial coefficients for our experiments.[46] Table 1 lists the extrapolated values for the second virial coefficient from the previous crystallization conditions of 21−30 mg mL^{-1} lysozyme, 12.5% w/v PEG 8kD, and 0.5 M NaCl in pH 6.2 phosphate buffer at temperatures of 7.2−12 °C.[46] However, the crystallization conditions in this study are at a higher salt concentration of 0.86 M NaCl and at a lower pH of 4.8. At pH 4.8, the net charge on lysozyme is greater than the net charge on lysozyme at pH 6.2. Under these conditions, the second virial coefficient, B_2, is expected to be more positive as a result of stronger Coulombic repulsive forces at pH 4.8 than at pH 6.2. 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, the decrease in pH and increase in NaCl concentration are expected to act in opposite ways on the virial coefficient. Hence, we expect the previously measured B_2 values to be a reasonable first guess for the virial coefficients in our study.

Using the estimated B_2's and eq 21, we corrected σ for the non-ideality of the crystallization conditions. The measured slow nucleation rates are fitted to eq 22 using the non-ideal σ.

$$J_s(\sigma) = A \cdot C \cdot e^{-B/\sigma^2} \tag{22}$$

We calculated the barrier height, ΔG^*, and the kinetic prefactor as follows.

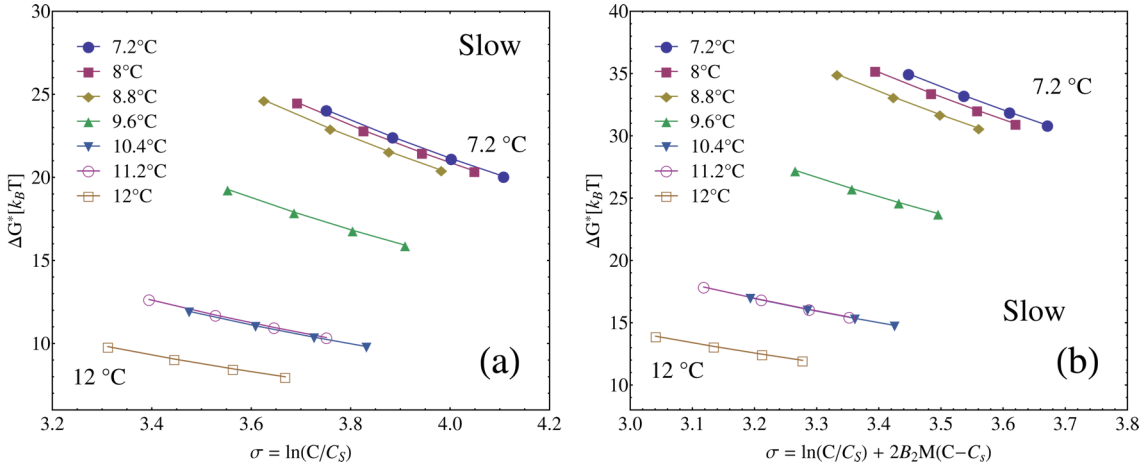


Figure 30. Barrier heights obtained using (a) σ for ideal solutions and (b) σ for nonideal solutions (eq 21). There is a ≃50% increase in the barrier heights, ΔG^*, obtained using the non-ideal σ.

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487−4509

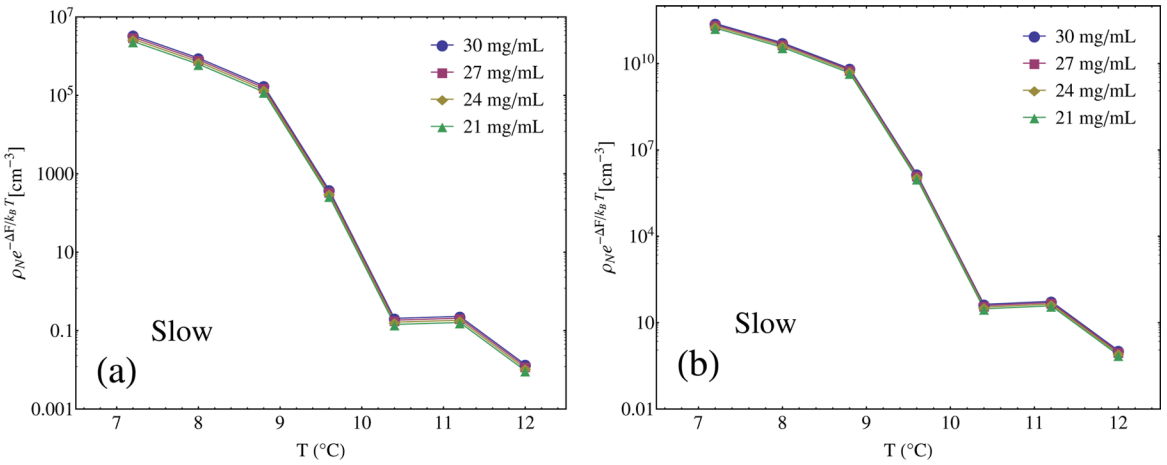


Figure 31. Kinetic prefactors obtained using (a) σ for ideal solutions and (b) σ for non-ideal solutions (eq 21). The kinetic prefactors that were obtained assuming ideal solution conditions are approximately 4 orders of magnitude smaller than the prefactors obtained assuming nonideal solution conditions (eq 21).

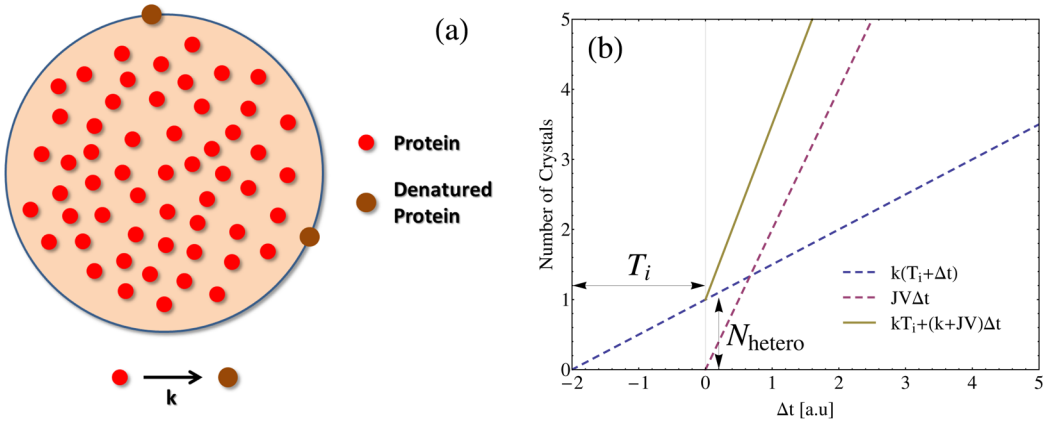


Figure 32. (a) Proposed model to understand the resulting offset in number of crystals per drop due to heterogeneous nucleation occurring at the protein droplet−oil interface, (b) The number of crystals per drop as a function of quench period Δt according to the proposed model.

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

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.

Conclusion 1. The barrier heights, ΔG^*, obtained using the non-ideal σ (eq 21) show qualitatively the same trend as the ideal σ, as shown in Figure 30. There is a ≃50% increase in the non-ideal case compared to the ideal case.

Conclusion 2. The kinetic prefactors that were obtained from assuming ideal solution conditions are approximately four orders of magnitude smaller than the prefactors obtained assuming non-ideal solution conditions, as shown in Figure 31. However, the qualitative trends are similar.

D. Understanding the Galkin and Vekilov Experiments[9]

Galkin and Vekilov performed lysozyme nucleation experiments in emulsions.[9] They attributed the number of crystals formed at $\Delta t = 0$ to heterogeneous nucleation occurring primarily at the protein droplet−oil interface. Such a model has several testable outcomes. The model shown in Figure 32a consists of 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 denature at rate k to form denatured protein molecules, which serve as heterogeneous nucleation centers at the interface. The number of denatured proteins at the interface is $k(T_i + \Delta t)$, where T_i is the incubation time, which is the elapsed time that the aqueous protein solution and oil have been in contact before the sample is quenched to start nucleation. Note that we define the start of experiment as the time at which target supersaturation is achieved and Δt is the elapsed time after quenching. Assuming that nucleation from a heterogeneous nucleation center occurs instantaneously, the rate at which crystals appear at the interface is also $k(T_i + k\Delta t)$, because in this model denaturation of the protein at the oil−water interface is the rate-limiting step in nucleating the crystals. Galkin and Vekilov observed 1 or 2 heterogeneous nucleation events at $\Delta t = 0$, which is the instant at which the sample is quenched. This 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 32b). Note that $N_{hetero} = kT_i$ is the number of crystals observed at $\Delta t = 0$. The model has the following predictions.

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

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487−4509

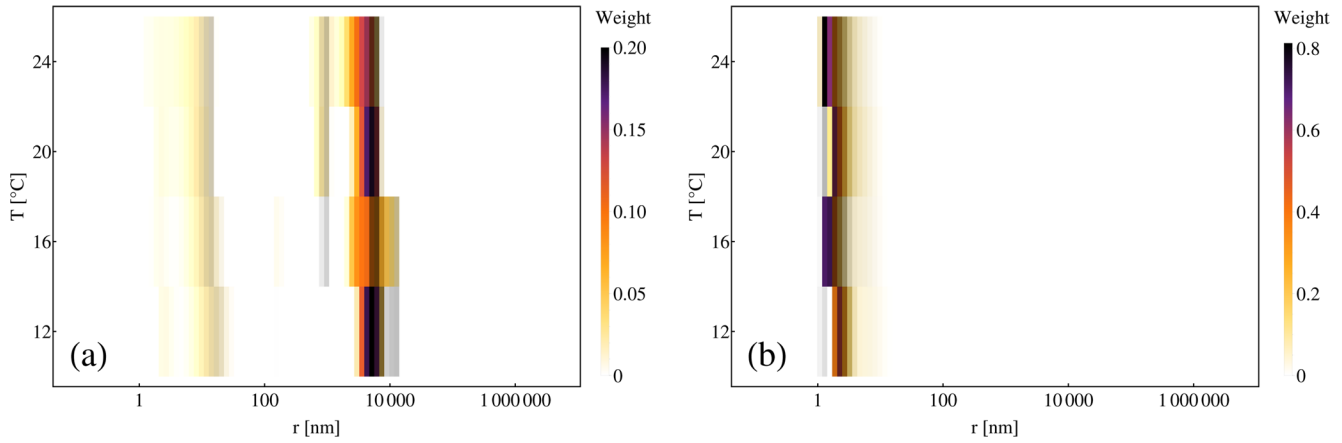


Figure 33. Dynamic light scattering measurement of aggregates in 50 mg mL^{-1} lysozyme and 12.5% w/v PEG 8kD in 0.1 M NaAc at pH 4.8. Data acquired on ALV-SLS/DLS 5000 system (ALV-GmbH, Langen (Hessen), Germany). (a) Intensity-weighted particle size distribution and (b) mass-weighted particle size distribution.

(2) The average number of crystals N_{hetero} per drop at $\Delta t = 0$ is proportional to drop area.

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 ≃0.1 to 10 μm. We have used 1-anilino-8-napthalene sulfonic acid ammonium salt (1,8-ANS) (Sigma, BioChemika 10417-25G-F), 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.

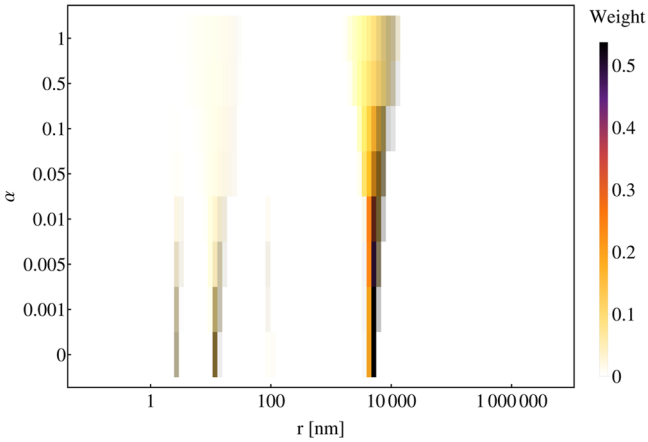


Figure 34. Intensity-weighted particle size distribution of a 50 mg mL^{-1} lysozyme and 12.5% w/v PEG 8kD in 0.1 M NaAc at pH 4.8, T = 12, solution. Smaller/weaker peaks disappear as the regularization parameter, α, is increased.

E.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 submicron range. The scattered intensity from the diffusing particles is used to calculate the intensity autocorrelation as a function of lag time. For a system with monodisperse particles, the intensity autocorrelation function decays exponentially with decay rate proportional to the diffusion constant. Using the Stokes−Einstein relation, the particle size is calculated once diffusion constant, viscosity, and temperature of the suspension are known. Figure 33a,b shows the intensity and mass-weighted particle size distributions of lysozyme/PEG mixtures as a function of temperature with the fraction of the particle size mapped on to an intensity colormap. 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 wavelength of light. The particle size at which the Rayleigh limit breaks down is $\rho \approx 2$, where ρ is a reduced variable defined as $\rho = 4\pi(n-1)R/\lambda$, n is the refractive index of the particle, R is the radius of the particle, and λ is the wavelength of probing light.[51] In our case, we have used λ = 633 nm; the corresponding particle size at which Rayleigh limit breaks down is R ≃ 250 nm. Therefore, particles of radius 250 nm and larger 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 ∼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 ∼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 autocorrelation 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 trustworthy. Figure 34 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.

E.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 by 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) (Sigma, BioChemika 10417-

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487−4509

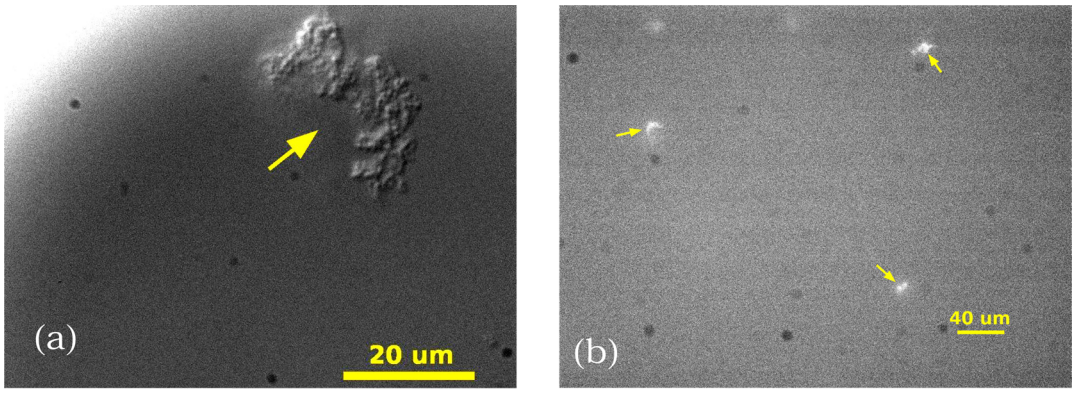


Figure 35. Presence of aggregates in two different trials of 50 mg mL^{-1} lysozyme and 12.5% w/v PEG 8kD in 0.1 M NaAc at pH 4.8 in (a) DIC and (b) fluorescence microscopy. 1-Anilinonaphthalene-8-sulfonic acid (100 mM) was used as the fluorescent dye, which fluoresces only when present in the hydrophobic regions of protein.

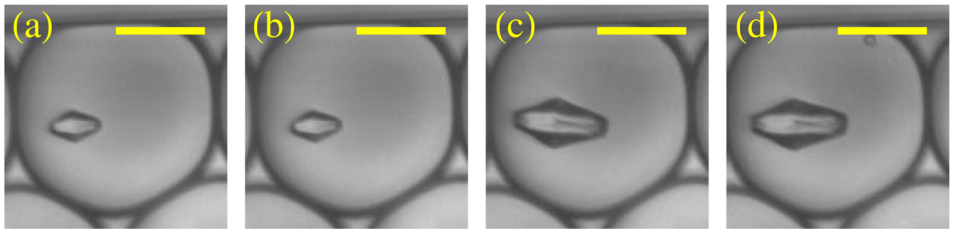


Figure 36. Measuring the growth of the longest dimension at (a) t = 0 s, (b) t = 14 min, (c) t = 28 min, and (d) t = 42 min. The scale bar is 50 mm.

25G-F) 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 35 shows the aggregates in DIC and fluorescence microscopy.

F. Growth Rate Measurements

We measured the length of the longest dimension in a crystal as a function of time. Figure 36 shows 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.

$$L(t) = a(1 - e^{-b(t-t_0)}) \tag{23}$$

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

$$\gamma = \left.\frac{dL}{dt}\right|_{t=t_0} = ab \tag{24}$$

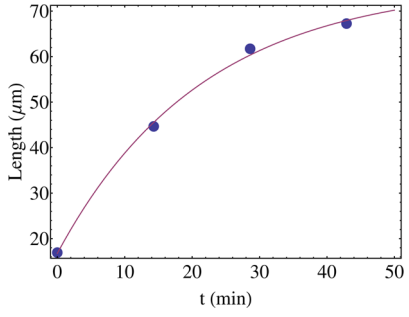


Figure 37. Length of the longest dimension as a function of time and the fit to eq 23.

Figure 37 shows the length of crystal as a function of time and the fit to eq 23.

G. Inverse Laplace Transformation

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

$$f_\phi(t) = \mathcal{L}g(s) = \int_0^\infty g(s)\, e^{-ts}\, ds \tag{25}$$

Since we measure $f_\phi(t)$ and wish to obtain the rate distribution function, we need to inverse Laplace transform (ILT) $f_\phi(t)$. The inverse Laplace transform is computed numerically using

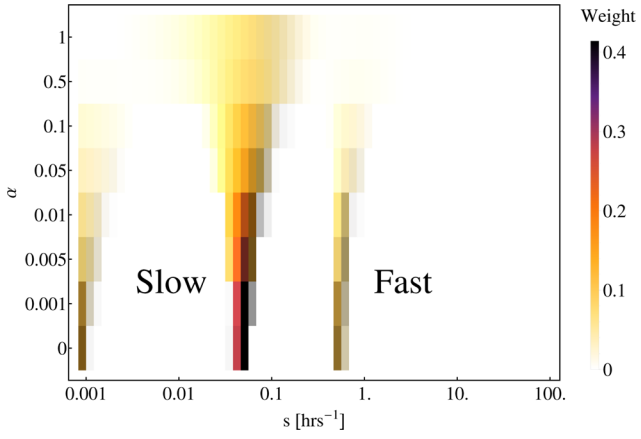


Figure 38. Decay rate distribution, $g(s)$, as a function of regularization parameter α and s. The presence of two peaks in $g(s)$ is not dependent on the choice of α.

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487−4509

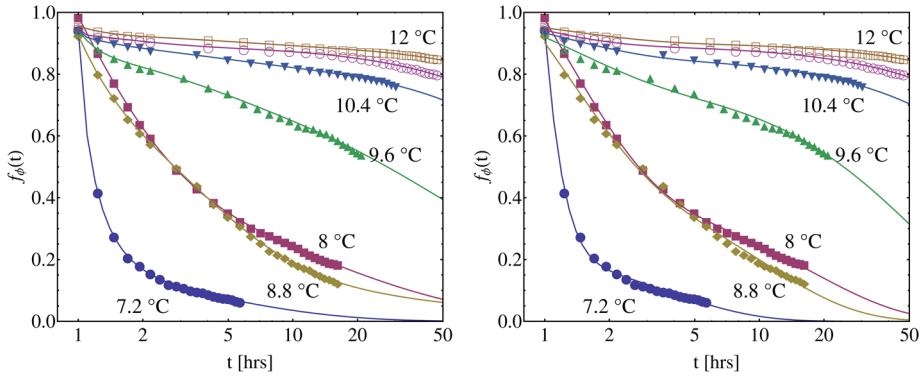


Figure 39. Measured (colored symbols) and fitted (colored lines) $f_\phi(t)$ as a function of time, t. The crystallization conditions are 30 mg mL^{-1} lysozyme, 12.5% w/v PEG 8kD, and 5% w/v NaCl in 0.1 M NaAc at pH 4.8 at different temperatures. (a) Fits using the inverse Laplace transform method. (b) Fits using a sum of two exponentials.

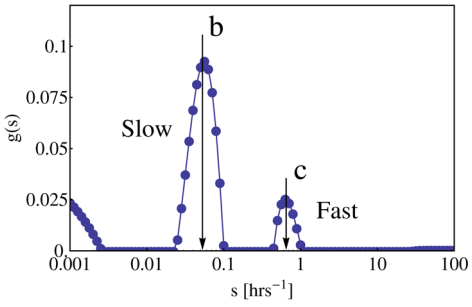


Figure 40. Decay rates b and c in eq 27 correspond to the two peaks observed in the decay rate distribution, $g(s)$.

Tikhonov regularization.[39] This method involves the minimization of the function

$$g(s, \alpha) = \min_{g(s)}\{\| f_\phi(t) - \int_0^\infty g(s)\, e^{-ts}\, ds \|^2 + \alpha \| g(s) \|^2\} \tag{26}$$

where α is the regularization parameter, which smooths the distribution $g(s)$. We used "fmincon," a MATLAB function, to perform the minimization of eq 26 while constraining $g(s)$ to be non-negative.

Figure 38 is the plot of decay rate distribution, $g(s)$, obtained as an ILT of the data shown in Figure 5, as a function of regularization parameter, α. Two peaks are present over a wide range of α, which suggests that their presence in the decay distribution, $g(s)$, is not an artifact due to the value of the regularization parameter. Note that the peak in $g(s)$ at small rates s, or long times, is due to the inaccuracy in establishing the baseline or long-time behavior of $f_\phi(t)$, as the experiments are stopped after 80 h.

Figure 39a, shows the fraction of drops with no crystals, $f_\phi(t)$, and the corresponding fits obtained using the ILT at various temperatures. Figure 39b is the plot of the fraction of drops with no crystals and the fits to a sum of two exponentials with the following functional form

$$f_\phi(t) = a\, e^{-bt} + (1 - a)\, e^{-ct} \tag{27}$$

where b and c are the decay rates with weights a and $1 - a$, respectively. The fitting parameters b and c correspond to the slow nucleation rate, k_s, and fast nucleation rate, k_f. Calculating $g(s)$ using an ILT is a regression analysis involving an infinite number of parameters since $g(s)$ is a continuous function; hence, it is expected that the ILT of Figure 39a fits the data better than does the two-exponential fit of Figure 39b. The decay rates b and c correspond to the two peaks observed in the decay rate distribution, $g(s)$, illustrated in Figure 40. Figure 41a,b shows the rates obtained using the ILT and two-exponential fits. The dotted lines represent the case in which both fits yield the same value. The ILT and two exponential fits yielded identical slow rates, b; however, there is a systematic

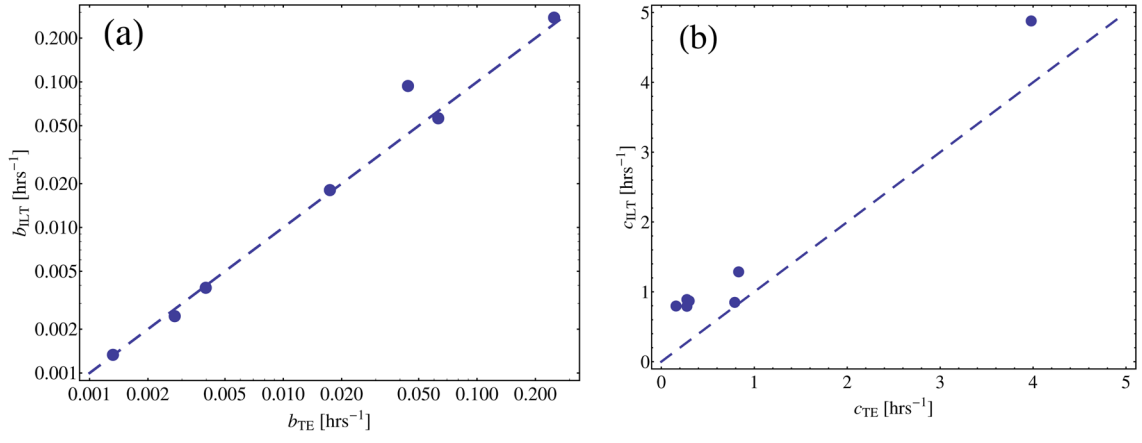


Figure 41. (a) Slow and (b) fast decay rates obtained from inverse Laplace transformation and sum of two exponential fit. The dotted lines represent a perfect scenario in which both fits yield the same fitting parameters.

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487−4509

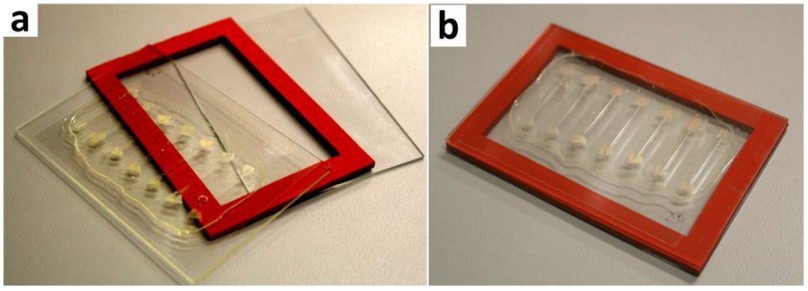


Figure 42. Creating an air chamber using a rectangular washer.

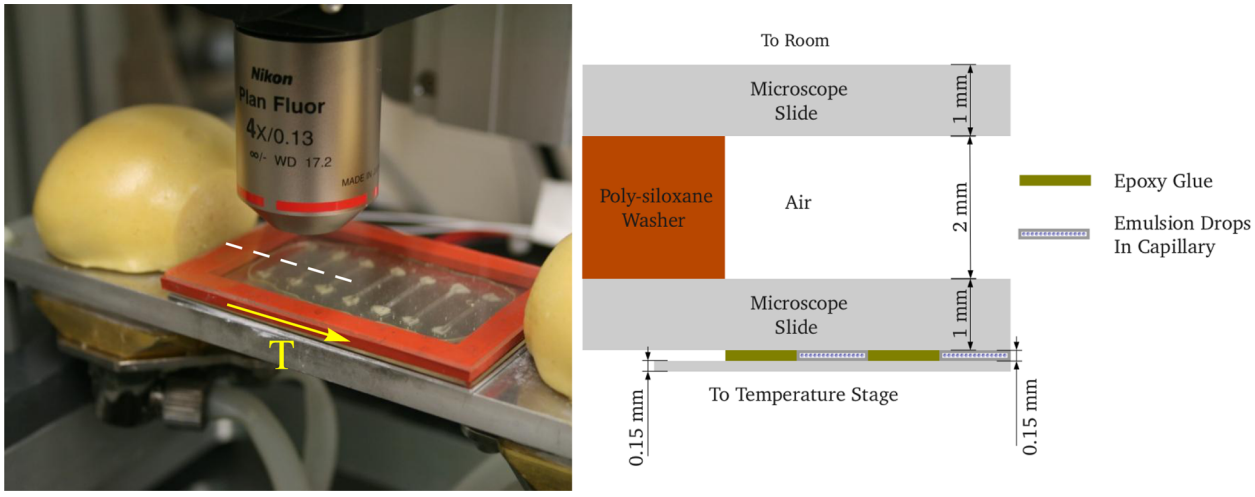


Figure 43. Sample mounted on the temperature/translation stage. The temperature stage consists of two thermoelectric coolers (TECs) that are independently controlled. The yellow colored blobs are the insulating foam covering the thermistors surface mounted to the temperature stage. To achieve a temperature gradient, each of the TECs is maintained at a different temperature. The arrow indicates the direction of temperature gradient. The inset figure is the cross-section along the dashed line, illustrating the relative dimensions of each of the components.

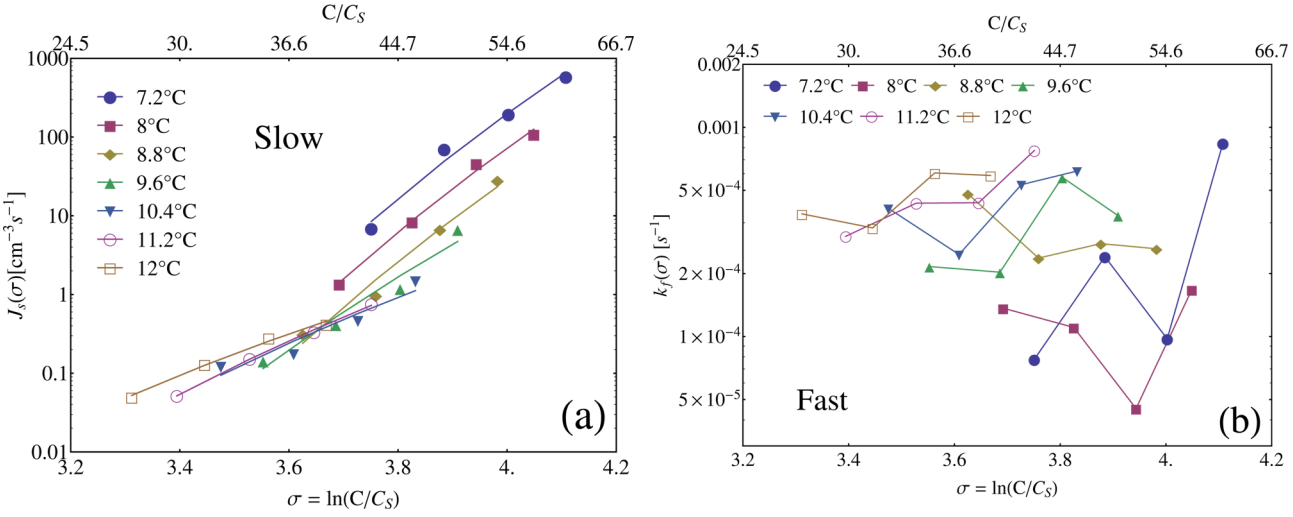


Figure 44. (a) Slow nucleation rates vs supersaturation. The colored symbols are measured data, and the colored lines are fit to eq 14a. (b) Fast nucleation rates as a function of supersaturation. The fast nucleation rates do not follow any systematic variation as a function of supersaturation. The crystallization conditions are lysozyme, 12.5% w/v PEG 8kD, and 5% w/v NaCl in 0.1 M NaAc buffer at pH 4.8.

deviation in the fast rate, c, perhaps due to inaccuracy in determining the baseline, which is manifested as a peak in $g(s)$ for small values of s in Figure 40.

H. Sample Preparation

The emulsion drops produced using the microfluidic chip are collected in a 0.5 mL eppendorf tube. The emulsion drops are stored in rectangular capillaries, purchased from VitroCom, with inner dimensions of width, 1 mm; height, 50/100 μm. The dimensions of capillaries are chosen such that droplets are packed in a hexagonal monolayer. The emulsion filled capillaries are sealed with VALAP,[36] a mix of equal parts of VAseline, LAnolin, and Paraffin wax with low melting

dx.doi.org/10.1021/cg500562r | Cryst. Growth Des. 2014, 14, 4487−4509

temperature. After the application of VALAP, the extra wax is scrapped off using a razor blade. The capillaries are then sandwiched between a microscope slide (75 × 50 × 1 mm) and a coverslip (48 × 65 × 0.15 mm), which are held together by epoxy (5 min) (purchased from Amazon Inc.) and are sealed air-tight.

To minimize the temperature difference between the metal plate connected to the TEC and the samples, we place the sandwiched samples with the coverslip side in contact with the metal plate. We insulate the sample from the surroundings using an air chamber as shown in Figures 42 and 43. The sample and a same sized microscope (75 × 50 × 1 mm) slide are pressed together with the rectangular poly-siloxane (product no. 3788T24, McMaster-Carr) washer, outer dimensions 75 × 50 × 2 mm, in between to create a sealed air chamber. To calibrate the system, we embedded 5 thermistors in epoxy in-between a glass coverslip and a glass slide. These thermistors were placed at the same height above the metal plate and the same lateral position as the samples. We neglected the temperature variations across the width of the capillary, i.e., perpendicular to the temperature gradient.

I. Nucleation Rates vs Supersaturation

The slow and fast nucleation rates vs supersaturation are shown in Figure 44.

■ ASSOCIATED CONTENT

Supporting Information

Movie S1: Droplet generation using a coflow microfluidic device; the protein and precipitant are mixed on-chip to avoid any nucleation before starting the experiment. The stream labeled "Protein" contains lysozyme, 12.5% w/v PEG 8kD, and 0.1 M NaAc at pH 4.8, and the stream labeled "Precipitant" contains 12.5% w/v PEG 8kD, 10% w/v NaCl, and 0.1 M NaAc at pH 4.8. Movie S2: Lysosyme crystallization in emulsion droplets of volume <1 nL produced using microfluidics. The crystallization conditions are 30 mg mL^{-1} lysozyme, and 12.5% w/v PEG 8kD, and 5% NaCl in 0.1 M NaAc buffer at pH 4.8 and at 6 °C. The total duration of the crystallization trial is approximately 36 h. The stochastic nature of crystal appearance is a characteristic of an activated process. Movie S3: Nucleation and growth of a crystal from a protein dense aggregate/gel. The aggregate dissolves as the crystal grows. Lysosyme crystallization in emulsion droplets of volume <1 nL produced using microfluidics. The crystallization conditions are 30 mg mL^{-1} lysozyme, 12.5% w/v PEG 8kD, and 5% NaCl in 0.1 M NaAc buffer at pH 4.8 and at 9 °C. The total duration of the movie is approximately 5 min. Movie S4: Liquid−liquid phase transition in lysozyme drops. Note the deterministic nature of the phase transition, i.e., all drops simultaneously undergo the phase transition, which is a characteristic of nucleation process with no activation barrier. The final composition in the drops is 50 mg mL^{-1} lysozyme, 5% w/v PEG 8kD, 0.5 M NaCl in 0.2 M phosphate buffer at pH 6.2. The temperature is varied from 8 to 13 °C. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: fraden@brandeis.edu; Phone: +1 (781) 736-2888.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Science Foundation MRSEC DMR-0820492 and NSF MWN 1209518. We thank Michael F. Hagan, Richard Sear, Bramie Lenhoff, Daan Frenkel, and Iskra Staneva for multiple discussions and suggestions.

■ REFERENCES

(1) Pound, G. M.; La Mer, V. K. *J. Am. Chem. Soc.* **1952**, *74*, 2323−2332.
(2) Gibbs, J. W. *Trans. Conn. Acad. Arts Sci.* **1876**, *III*, 108−248.
(3) Guo, B.; Kao, S.; McDonald, H.; Asanov, A.; Combs, L. L.; Wilson, W. W. *J. Cryst. Growth* **1999**, *196*, 424−433.
(4) McPherson, A. *Crystallization of Biological Macromolecules*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1999.
(5) Garcia-Ruiz, J. M. *J. Struct. Biol.* **2003**, *142*, 22−31.
(6) Chayen, N. E.; Saridakis, E. *Nat. Methods* **2008**, *5*, 147−153.
(7) Debenedetti, P. G *Metastable Liquids: Concepts and Principles*; Princeton University Press: Princeton, NJ, 1996.
(8) Gunton, J. D.; Shiryayev, A.; Pagan, D. L. *Protein Condensation: Kinetic Pathways to Crystallization and Disease*; Cambridge University Press: Cambridge, 2007.
(9) Galkin, O.; Vekilov, P. G *J. Phys. Chem. B* **1999**, *103*, 10965−10971.
(10) Sear, R. P. *J. Phys. Chem. B* **2006**, *110*, 21944−21949.
(11) Sear, R. P *J. Phys.: Condens. Matter* **2007**, *19*, 033101.
(12) Fisher, J. C.; Holloman, J. H.; Turnbull, D. *J. Appl. Phys.* **1948**, *19*, 775−784.
(13) Turnbull, D.; Fisher, J. C. *J. Chem. Phys.* **1949**, *17*, 71−73.
(14) Fisher, J. C.; Holloman, J. H.; Turnbull, D. *Science* **1949**, *109*, 168−169.
(15) Turnbull, D. *J. Appl. Phys.* **1949**, *20*, 817−818.
(16) Turnbull, D. *J. Appl. Phys.* **1950**, *21*, 1022−1028.
(17) Galkin, O.; Vekilov, P. G. *J. Cryst. Growth* **2001**, *232*, 63−76.
(18) Bhamidi, V.; Varanasi, S.; Schall, C. A. *Cryst. Growth Des.* **2002**, *109*, 395−400.
(19) Astier, J.-P.; Veesler, S. *Cryst. Growth Des.* **2008**, *8*, 4215−4219.
(20) Selimovic, S.; Jia, Y.; Fraden, S. *Cryst. Growth Des.* **2009**, *9*, 1806−1810.
(21) Ildefonso, M.; Candoni, N.; Veesler, S. *Cryst. Growth Des.* **2011**, *11*, 1527−1530.
(22) Selimovic, S.; Gobeaux, F.; Fraden, S. *Lab Chip* **2010**, *10*, 1696−1699.
(23) Dombrowski, R. D.; Litster, J. D.; Wagner, N. J.; He, Y. *Am. Inst. Chem. Eng.* **2010**, *56*, 79−91.
(24) Zheng, B.; Tice, J. D.; Roach, L. S.; Ismagilov, R. F. *Angew. Chem., Int. Ed.* **2004**, *43*, 2508−2511.
(25) Yadav, M. K.; Gerdts, C. J.; Sanishvili, R.; Smith, W. W.; Roach, L. S.; Ismagilov, R. F.; Kuhn, P.; Stevens, R. C. *J. Appl. Crystallogr.* **2005**, *38*, 900−905.
(26) Khvostichenko, D. S.; Kondrashkina, E.; Perry, S. L.; Pawate, A. S.; Brister, K.; Kenis, P. J. A. *Analyst* **2013**, *138*, 5384−5395.
(27) Heymann, M.; Opthalage, A.; Wierman, J. L.; Akella, S.; Szebenyi, D. M. E.; Gruner, S. M.; Fraden, S. *IUCrJ* **2014**, *1*, DOI: 10.1107/S2052252514016960.
(28) Malkin, A. J.; Thorne, R. E. *Macromol. Cryst.: Methods* **2004**, *34*, 273−299.
(29) Chapman, H. N.; Fromme, P. *Nature* **2011**, *470*, 73−77.
(30) Boutet, S.; Lomb, L. *Science* **2012**, *337*, 362−364.
(31) Stellato, F.; Oberthür, D.; Liang, M.; Bean, R.; Gati, C.; Yefanov, O.; Barty, A.; Burkhardt, A.; Fischer, P.; Galli, L.; Kirian, R. A.; Meyer, J.; Panneerselvam, S.; Yoon, C. H.; Chervinskii, F.; Speller, E.; White, T. A.; Betzel, C.; Meents, A.; Chapman, H. N. *IUCrJ* **2014**, *1*, 204−212.
(32) Sear, R. P. *Int. Mater. Rev.* **2012**, *57*, 328−356.
(33) Anna, S. L.; Bontoux, N.; Stone, H. A. *Appl. Phys. Lett.* **2003**, *82*, 364−366.
(34) Holtze, C.; Weitz, D. A. *Lab Chip* **2008**, *8*, 1632−1639.

(35) Roach, L. S.; Song, H.; Ismagilov, R. F. *Anal. Chem.* **2005**, *77*, 785−796.
(36) Inoue, S.; Spring, K. R. *Video Microscopy: The Fundamentals*; Plenum Press: New York, 1997.
(37) Provencher, S. W. *Comput. Phys. Commun.* **1982**, *27*, 213−227.
(38) Provencher, S. W. *Comput. Phys. Commun.* **1982**, *27*, 229−242.
(39) Tikhonov, A. N.; Arsenin, V. Y. *Solutions of Ill-Posed Problems*; Halsted Press: New York, 1977.
(40) Laval, P.; Salmon, J.-B.; Joanicot, M. *J. Cryst. Growth* **2007**, *207*, 622−628.
(41) Laval, P.; Crombez, A.; Salmon, J.-B. *Langmuir* **2009**, *25*, 1836−1841.
(42) Rosenbaum, D. F.; Kulkarni, A.; Ramakrishnan, S.; Zukoski, C. F. *J. Chem. Phys.* **1999**, *111*, 9882−9889.
(43) Kulkarni, A.; Zukoski, C. *J. Cryst. Growth* **2001**, *232*, 156−164.
(44) Forsythe, E. L.; Judge, R. A.; Pusey, M. L. *J. Chem. Eng. Data* **1999**, *44*, 637−640.
(45) Diamond, R. *J. Mol. Biol.* **1974**, *82*, 371−391.
(46) Bloustine, J.; Virmani, T.; Thurston, G. M.; Fraden, S. *Phys. Rev. Lett.* **2006**, *96*, 087803.
(47) Kashchiev, D. *J. Chem. Phys.* **1982**, *10*, 5098−5102.
(48) Vekilov, P. G.; Monaco, L. A.; Thomas, B. R.; Stojanoff, V.; Rosenberger, F. *Acta Crystallogr., Sect. D* **1996**, *52*, 785−798.
(49) Page, A. J.; Sear, R. P. *Phys. Rev. Lett.* **2006**, *97*, 065701.
(50) Vekilov, P. G. *J. Phys.: Condens. Matter* **2012**, *24*, 193101.
(51) van de Hulst, H. C. *Light Scattering by Small Particles*; Dover Publications, Inc.: New York, 1981.