Comparison of experimental estimates and model predictions of protein crystal nucleation rates

Narendra M. Dixit, Amit M. Kulkarni, Charles F. Zukoski *

114, Roger Adams Laboratory, Department of Chemical Engineering, University of Illinois at Urbana-Champaign, 600, S. Mathews Avenue, Urbana, IL 61801, USA

Abstract

Estimates of protein crystal nucleation rates determined from three independent sets of experimental observations are compared for internal consistency. Large discrepancies emerge. A qualitative understanding of these discrepancies is gained from an analysis of the experimental techniques used within the framework of extant models for nucleation rates. Quantitatively, however, these models fail to predict nucleation rates of protein crystals. Possible origins of this failure lie in the approximate descriptions of the interaction potentials between protein molecules and the processes of aggregation and dissociation of single molecules on cluster surfaces that lead to crystal nucleation. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Protein crystallization; Nucleation rate; Induction time

1. Introduction

Driven by the demand for high quality protein crystals, essential for the determination of protein structures, fundamental studies of protein phase behavior have endeavored to identify optimal solution conditions for crystallizing proteins. These studies demonstrate that the broad features of the equilibrium thermodynamics of protein solutions can be understood using models based on simple fluid theory [1–5]. However, the dynamics of phase transitions in protein solutions remain poorly understood. Indeed, when compared to small molecule systems, nucleation rates of protein crystals are inexplicably low. Developing an understanding of the mechanistic origin of this discrepancy has been the focus of several recent investigations of protein crystal nucleation and growth [6–16].

Recently, Galkin and Vekilov have proposed a temperature jump technique to decouple nucleation and growth in protein crystallization experiments [10,11]. Using this technique they have reported measurements of nucleation rates for lysozyme under several solution conditions. Kulkarni and Zukoski [13] have determined induction times for lysozyme crystal nucleation over a wide range of solution conditions using light scattering. Under similar solution conditions Darcy and Wieneck [9] have determined the enthalpy of crystallization for lysozyme using mi-
crocalorimetry. From these latter two sets of data, indirect estimates of crystal nucleation rates for lysozyme may be obtained. This makes available three independent sets of experimental observations all leading to estimates of protein crystal nucleation rates. These observations have been made on the crystallization of lysozyme under similar solution conditions but have not been compared for internal consistency. In this paper, we test the internal consistency of these measurements and compare the reported rates with models for crystal nucleation.

Predictions of nucleation rates can be made using the well-established classical nucleation theory [17]. Here, the nucleation rate is given by an expression of the form $A \exp(\Delta E^*)$, where $\Delta E^*$ is the free energy change associated with the formation of a cluster of a critical size, and $A$, the prefactor characterizing the diffusion limited aggregation of particles onto a cluster surface. The critical cluster size is defined as that size at which the free energy of cluster formation, determined from a balance between opposing volumetric and surface energy contributions, goes through a maximum. While the classical nucleation theory is known to capture $\Delta E^*$ well, the prefactor, $A$, remains much less understood [6,10,11,17].

Recently, using simulation techniques, ten Wolde and Frenkel [6] determined $\Delta E^*$ in the vicinity of the metastable critical point for protein solutions. They observed that the proximity to the critical point produces tremendous density fluctuations resulting in pockets of high densities within which the barrier to nucleation is reduced. A minimum in $\Delta E^*$ resulted at the critical point, suggesting that crystallizing proteins close to the critical point enhanced nucleation rates significantly. Further, they argued that while nucleation within these density pockets is rapid, the ensuing crystal growth, which occurs over timescales much larger than the timescale of the density fluctuations and is therefore governed by the average supersaturation of the system, can be made arbitrarily slow by shifting the location of the critical point close to the solubility boundary. This work suggests that by crystallizing proteins near the critical point, high quality crystals can be produced at enhanced rates.

Following these ideas, Dixit and Zukoski [18] have developed an analytical model to calculate nucleation rates in the presence of density fluctuations. They build on an alternative kinetic nucleation model of Narasimhan and Ruckenstein [19,20], which allows the calculation of nucleation rates in homogeneous suspensions without requiring estimates of the surface energy. Here, cluster growth or shrinkage is determined as a competition between the rates of particle aggregation onto and evaporation from a cluster surface. The rate of formation of clusters of the critical size, defined as that size at which these competing rates balance, gives the nucleation rate. In agreement with the predictions of ten Wolde and Frenkel [6], Dixit and Zukoski [18] find that the nucleation rate in the presence of density fluctuations goes through a maximum at the critical point. Enhancements of nearly 10 orders of magnitude in the nucleation rate in the presence of fluctuations over that in the absence of fluctuations are predicted, reiterating the advantage of crystallizing proteins near the critical point. Experimental verification of these predictions is difficult because the protein concentration near the critical point is large ($\sim 400 \text{ mg ml}^{-1}$) making experiments in this region prohibitive in terms of the amount of protein required. Protein crystallization experiments, such as the ones mentioned earlier [8–13], are usually conducted at significantly lower concentrations ($\sim 10–50 \text{ mg ml}^{-1}$) where the influence of fluctuations is difficult to observe.

Here, we aim to test the ability of standard models of nucleation rates, which are known to describe nucleation in small molecule systems well, to capture nucleation in protein systems. For this, we apply our kinetic nucleation model under the same conditions as the experiments described earlier. First, in Section 2 we compare the independent sets of experimental observations of protein crystal nucleation rates for internal consistency. We find large discrepancies in the rates reported for essentially identical systems measured by different methods. To reconcile these discrepancies, we examine the experimental techniques used and the assumptions made in interpreting the data. Possibilities of large uncertainties in some of the experimental tech-
niques become evident. With knowledge of these uncertainties, in Section 3 we compare model predictions of nucleation rates and experimental observations. Qualitative agreement is seen, but the comparison is marked by significant quantitative discrepancies. The model grossly overpredicts nucleation rates, exceeding even the upper bounds on the nucleation rates set by these experiments by several orders of magnitude. In Section 4, we present our analysis of the experimental techniques used, and in Section 5 we draw conclusions.

2. Experimental estimates of nucleation rates

Experimental investigation of nucleation kinetics is hindered by the inherent coupling between nucleation and ensuing crystal growth: While new crystals nucleate, existing crystals grow. Nucleation and growth deplete a crystallizing sample of the dissolved solute. This lowers the supersaturation and in turn reduces the nucleation and growth rates. This coupling is captured by the Johnson–Mehl–Avrami–Kolmogrov (JMAK) integral [21], which determines the total amount of solute converted to crystal form due to simultaneous nucleation and growth

\[ V(t) = \int_0^t I(\phi(\tau)) \, d\tau \frac{4\pi}{3} \times \left[ R^*(\phi(\tau)) + \int_0^{\phi(t)} \frac{dR(\phi(\theta))}{d\theta} \, d\theta \right]^3 \]  (1)

where \( V \) is the volume of solute in crystal form per unit volume of the solution after a time \( t \) from the onset of nucleation, \( I(\phi) \) is the nucleation rate of crystals of size \( R^*(\phi) \) and \( dR/dt \), their growth rate. The time dependence of the volume fraction of the solute, \( \phi \), in the solution is determined by the mass balance

\[ \phi(0) = V(t) \phi_{cp} + (1 - V(t)) \phi(t) \]  (2)

where \( \phi_{cp} \) is the packing fraction of the solute molecules in the crystals. While several models provide independent descriptions of nucleation and growth, for realistic forms of \( I(\phi) \) and \( dR/dt \), solving the above-mentioned simultaneous equations is extremely difficult. Extracting kinetic information from experiments involves solving the exact inverse of this problem, viz. to determine \( I(\phi) \) and \( dR/dt \) given \( \phi(t) \), for example, and is almost impossible. The equations, however, become extremely simplified if one of the processes — nucleation or growth — is suppressed.

Galkin and Vekilov [10,11] devised an experimental protocol for determining the rate of protein crystal nucleation that aims to achieve this decoupling. At a fixed concentration, a protein solution is quenched to a temperature \( T_1 \) well below the solubility boundary. The supersaturated solution is held at \( T_1 \) for an interval of time \( t_1 \), and then the temperature is raised to \( T_2 \) just below the solubility boundary. Finally, the solution is held at \( T_2 \) for an extended period of time \( t_2 \) after which the number of crystals in the solution is counted. Galkin and Vekilov analyze their data by arguing that crystals nucleate at \( T_1 \) and grow at \( T_2 \). The analysis proceeds by assuming that in the time interval \( t_1 \), crystals are nucleated but undergo negligible growth. Thus nucleation is assumed to occur at constant supersaturation. The temperature \( T_2 \) is chosen to lower the supersaturation to a point arresting nucleation yet allowing existing nuclei to grow. As a result, the crystals nucleated during \( t_1 \) grow to an observable size at \( T_2 \) and the data are analyzed, assuming no new crystals are nucleated after the temperature is raised to \( T_2 \). Carrying out a large number of experiments, Galkin and Vekilov report the average number of crystals formed, \( N \), in a sample volume \( V_S \). These experiments are repeated for different values of \( t_1 \). The average number of crystals is found to be a linear function of \( t_1 \) with a form \( N/V_S = N_{\text{hetero}}/V_S + t_1 I_{\text{measured}} \), where the intercept, \( N_{\text{hetero}} \), is assumed to be the number of crystals formed by heterogeneous nucleation and \( I_{\text{measured}} \) is the reported rate of homogeneous nucleation. Nucleation rates reported are presented in Fig. 1.

Darcy and Wiencek [9] determined the enthalpy of lysozyme crystallization by microcalorimetry under conditions similar to those used by Galkin and Vekilov. In these experiments, the heat released by a crystallizing sample was recorded in terms of power, \( P \), as a function of time \( t \) over a time interval \( \tau \). The difference between the solu-
tion concentrations at the beginning and the end of the experiment gave the mass of protein, $M$, converted to crystalline form. The enthalpy of crystallization was obtained as

$$\Delta H = \lim_{t \to \infty} \frac{1}{M} \int_0^t P(t) \, dt$$

The heat released during crystallization is directly related to the rate of crystal nucleation and growth.

Assuming that the characteristic lengthscale of the crystals was proportional to the total mass in crystal form, Darcy and Wiencek determined crystal face growth rates from the cumulative power at any time $t$. Fits to model parameters yielded good agreement with lysozyme crystal growth rates determined by Nadarajah et al. [22] providing a consistency check with other data in the literature. Using this heat release data, an upper bound on the nucleation rate may be obtained by assuming that only nucleation and no growth occurs so that all of the heat released comes from the formation of the critical nuclei. We note that this assumption is extreme in the light of the observations of Darcy and Wiencek but consistent with the assumptions of Galkin and Vekilov who studied lysozyme crystallization under very similar conditions.

Assuming no growth occurs during the time interval $t_1$, the rate at which protein mass is converted into crystalline form is given by

$$\frac{dM}{dt} = IV_s n^* M_w$$

where $n^*$ is the number of protein molecules in a cluster of the critical size (described later), $M_w$ is the molecular weight of lysozyme, and $I$, the nucleation rate. We choose, $n^* = 4$, consistent with the estimates of Galkin and Vekilov, and $M_w = 14\,600 \, \text{g mol}^{-1}$. With $\Delta H$ as the enthalpy of crystallization, the power generated by the nucleation process can be written as

$$P(t) = \Delta H \times \frac{dM}{dt}$$

Combining Eqs. (3) and (4), the nucleation rate can be written as:

$$I = \frac{P(t)}{V_s n^* M_w \Delta H}$$

Knowing $\Delta H = 14.3 \, \text{kcal mol}^{-1}$ as reported by Darcy and Wiencek, the denominator in the above equation is a known constant, and the nucleation rate tracks the power released with time. In the Darcy and Wiencek experiments, $P(t)$ begins at zero, rises to a maximum in a short time, and then decays slowly to zero. The time to reach the maximum characterizes an induction time required for nucleation to reach steady state. Had the nucleation remained steady for a period of time, $P(t)$ would have been a constant. However, even before the power signal can reach steady state, the number of crystals formed becomes significant enough to reduce the protein concentration and thus the supersaturation. This in turn causes the nucleation rate and thus $P(t)$ to fall until they become zero at equilibrium. Assuming that the protein concentration does not change much until $P(t)$ reaches a maximum, we calculate $I$ corresponding to the maximum $P(t)$ as

![Fig. 1. Comparison of estimates of nucleation rates of lysozyme crystals in 3% NaCl, 0.05 M NaAc buffer at pH 4.5, from different experimental methods. Large discrepancies exist between the different experimental estimates. Calculations using the kinetic nucleation model under similar conditions overestimate nucleation rates by several orders of magnitude.](image)

- **Kinetic model**
- **Darcy and Wiencek**
- **Kulkami and Zukoski**
- **Galkin and Vekilov**

an estimate of the nucleation rate that corresponds to the supersaturation at the beginning of the experiment.

Knowing \( P(t) \) at one volume fraction, we can estimate \( P(t) \) at other volume fractions as follows. Darcy and Wiencek define a characteristic length-scale of the crystals, \( L \), such that \( M(t) \) is related to \( L(t) \) as

\[
\frac{M(t)}{M_i} = \left( \frac{L(t)}{L_i} \right)^3
\]

where \( M_i \) and \( L_i \) correspond to the equilibrium values of \( M \) and \( L \). Then, from Eqs. (4) and (6), it follows that

\[
\frac{dL}{dt} = \frac{L_i^3}{2} \left( \frac{P(t)}{\Delta H} \right)^{2/3} \left[ \int_0^t P(t') dt' \right]^{1/3}
\]

Since \( P(t) \) rises sharply to a maximum, we can write, at \( t = t_{\text{max}} \),

\[
\int_0^t P(t') dt' \approx P(t_{\text{max}}) t_{\text{max}}/2 = P_{\text{max}} t_{\text{max}}/2
\]

say, so that, \( dL/dt \) at \( t = t_{\text{max}} \) varies as \( (P_{\text{max}})^{1/3} \). Here, we assume that for the range of concentrations considered, \( t_{\text{max}} \) does not vary much. Then, following the assumption of Darcy and Wiencek that \( dL/dt = k[c(t) - c_{\text{sat}}]^n \), where \( c \) stands for the concentration and \( c_{\text{sat}} \) the solubility, we obtain

\[
\frac{P_{\text{max}}(c_1)}{P_{\text{max}}(c_2)} = \left( \frac{c_1 - c_{\text{sat}}}{c_2 - c_{\text{sat}}} \right)^{3n}
\]

Under the conditions considered, Darcy and Wiencek report \( c_{\text{sat}} = 6 \text{ mg ml}^{-1} \) and determine \( n = 2.72 \) for \( c = 64 \text{ mg ml}^{-1} \). Assuming \( n \) does not vary much with protein concentration and knowing from measurements that \( P_{\text{max}} = 2.5 \mu\text{W} \) for \( c = 64 \text{ mg ml}^{-1} \), we can estimate nucleation rates at various protein concentrations. Comparisons with the work of Galkin and Vekilov are possible because similar values of supersaturation and temperature were studied. Of interest is the observation that \( t_{\text{max}} \) lies in the range of values of \( t_1 \) studied by Galkin and Vekilov. Nucleation rates estimated in this manner are given in Fig. 1.

Kulkarni and Zukoski [13] invoke classical nucleation theory to argue that induction times and nucleation rates have inverse, yet exponential dependencies on the work of formation of a cluster of the critical size, but the two differ in the pre-exponential terms. Thus, induction time may be written as \( t_{\text{ind}} = B \exp(-\Delta E^\ast) \), where \( \Delta E^\ast \) is the same as in the expression for nucleation rates. Kulkarni and Zukoski calculate \( B \) from their experiments. Using estimates of protein diffusivity to link \( B \) to the pre-exponential factor \( A \), they are able to calculate nucleation rates from their measurements of induction times. Rates estimated in this fashion are shown in Fig. 1. In agreement with classical nucleation theory, the induction times — and thus, the nucleation rates — have an explicit dependence only on the supersaturation and not on the protein concentration or the temperature. The estimates of surface tension obtained from \( \Delta E^\ast \) agree with other estimates in the literature based on strengths of attraction, enthalpies of crystallization and step growth rates. This reinforces the general notion that even in protein suspensions, classical nucleation theory captures \( \Delta E^\ast \) well. The pre-exponentials \( A \) and \( B \) are related to the particle diffusivity in classical nucleation theory and can be estimated in an a priori manner. Kulkarni and Zukoski find that values of \( A \) and \( B \) determined via experiments are substantially different from \( (A \text{ being much smaller than}) \) predictions from classical theory.

To compare the results of the above three experimental estimates of nucleation rates, we focus attention on one set of buffer conditions: 3% NaCl salt, 0.05 M NaAc buffer, pH 4.5. \( T = 286 \text{ K} \), where data from all the three sets of experiments exists. In Fig. 1 we show the nucleation rates measured by Galkin and Vekilov and the rates estimated from the data of Darcy and Wiencek and Kulkarni and Zukoski as described earlier. The comparison is poor. Rates estimated from the Darcy and Wiencek and the Kulkarni and Zukoski data are more than 10 orders of magnitude above the measurements of Galkin and Vekilov. That the data of Darcy and Wiencek is also above that of Kulkarni and Zukoski can be accounted for by the fact that we have assumed that only nucleation and no growth occurs up to \( t_{\text{max}} \) and have neglected heterogeneous nucleation. However, the glaring discrepancy between the other sets of data needs to be understood. For
this purpose, we develop here an analysis based on models of nucleation rates.

3. Predictions of nucleation rate models

The classical approach to calculating nucleation rates treats nucleation as an activated process [17]. Here the free energy change associated with cluster formation, $\Delta E^*$, has two components: a negative component from the formation of the thermodynamically favorable crystal nucleus and a positive component from the formation of an interface between the nucleus and the bulk solution. The individual components have dependencies on cluster size such that small clusters shrink and large ones grow. This defines a critical cluster size as that size at which a cluster is just stable with no tendency to grow or shrink. $\Delta E^*$ goes through a maximum at this size, giving rise to a barrier to nucleation which growing clusters must cross to become stable. The nucleation rate, $I$, is then given as the rate at which clusters cross this barrier. Drawing from reaction kinetics, this rate is written as

$$I = A \exp(\Delta E^*)$$

where $A$ is the prefactor.

To calculate induction times using classical nucleation theory one argues that the induction time is a measure of the time required for a cluster to grow to a critical size and thus must be inversely related to the nucleation rate. Thus the induction time is written as

$$t_{\text{ind}} = B \exp(-\Delta E^*)$$

Kulkarni and Zukoski [13] use this link of $\Delta E^*$ between $I$ and $t_{\text{ind}}$ in their estimates of nucleation rates from induction times. Sohnel and Mullin [23] argue that the induction time is the sum of the time required for the nucleation of a stable cluster, $t_\text{n}$, and the time for its growth to a detectable critical size, $t_\text{g}$. On an average, $t_\text{n} = 1/I$, where $I$ is the nucleation rate. Thus, the assumption in Eq. (10) that $t_{\text{ind}}$ varies as $1/I$ ceases to hold when $t_\text{g}$ is large compared to $t_\text{n}$.

An alternative completely kinetic approach to calculating nucleation rates has been proposed by Narasimhan and Ruckenstein [19,20]. Here, cluster growth or shrinkage is determined as a competition between two processes: aggregation of monomers onto and evaporation of monomers from a cluster surface. The rate of aggregation, $\beta$, is determined by solving the diffusion equation for particles diffusing down a concentration gradient, assuming the concentration to equal the bulk concentration far away from the cluster surface and zero at the cluster surface. Molecules on the surface are assumed to be in a potential well because of their bonds with neighboring molecules. The rate at which surface molecules escape out of this potential well is calculated by solving the Smoluchowski equation for the motion of a particle in a potential well. This gives the rate of evaporation, $\alpha$. A cluster packing profile is chosen such that molecules on the surface of smaller clusters are bound weaker than on larger clusters. Thus, $\alpha$ is large for small cluster sizes and decreases as the cluster size increases. Being proportional to the area of the cluster surface, $\beta$ is small for small clusters and increases with cluster size. Small clusters thus shrink and large clusters grow. A critical cluster size is defined as that size at which $\alpha = \beta$. The rate of formation of clusters of this size gives the nucleation rate. Dixit and Zukoski [18] have modified this model to establish thermodynamic consistency (i.e. the variations in solubility with well depth and temperature are accounted for in the rate equation in a consistent manner) and incorporate the dependence of the gradient diffusivity of monomers on the volume fraction. Here, we use this modified kinetic model to calculate nucleation rates under the conditions of the experiments described earlier.

Protein molecules are treated as spherical particles interacting with a centrosymmetric square well potential, with $e/kT$, the magnitude of the attractive well depth, and $\lambda > 1$, the extent of the square well. Ramakrishnan and Zukoski [24] have built on the work of Heyes and Aston [25] to develop an equation of state for square well fluids. Varying the parameter $\lambda$ alters the relative stability of the critical point with respect to the solubility boundary. By matching the second virial coefficient, $B_2$, for square well systems with experimental observations of $B_2$ at the critical point for
lysozyme, Ramakrishnan and Zukoski have shown that $\lambda = 1.1$ corresponds to the phase diagram of lysozyme in 3% NaCl, 0.05 M NaAc buffer, pH 4.5, as used in the nucleation experiments discussed here. This phase diagram is shown in Fig. 2. Under these conditions Rosenbaum et al. [26] have recently determined $B_2$ as a function of temperature, $T$. Matching the values of $B_2$ to experimental values yields $\langle \varepsilon/kT \rangle_1 = 2.22$ and $\langle \varepsilon/kT \rangle_2 = 2.05$, corresponding to $T_1 = 12.5^\circ C$ and $T_2 = 20^\circ C$ used by Galkin and Vekilov. At these temperatures, the lysozyme solubilities are $\phi_s = 0.0017$ and 0.005, corresponding to values obtained by Muschol and Rosenberger [3]. Here, the protein volume fraction in solution at the solubility boundary is $\phi_s = \rho (4\pi a^3/3)$, where $\rho$ is the number density of lysozyme molecules and $a$ is the radius of a lysozyme molecule set at 1.7 nm. Using these parameter values, we use the kinetic model of Dixit and Zukoski [18] to calculate nucleation rates at $T_1$ for various supersaturations. The resulting rates are shown in Fig. 1.

Given the continuous phase viscosity, solubility, $\varepsilon/kT$ and $\lambda$, the current kinetic model has no adjustable parameters. Model predictions, shown in Fig. 1, exceed the experimental estimates of nucleation rates by several orders of magnitude. The kinetic model for nucleation has been shown to predict nucleation rates for small molecule systems well [17,19,20]. Assuming the mechanisms of nucleation to be similar in proteins and small molecule systems, one would expect this model to capture nucleation rates in protein systems equally well. Surprisingly, however, model predictions exceed even the upper bound on nucleation rates set by the Darcy and Weincek data by several orders of magnitude. We therefore conclude that the model fails to capture the nucleation rates in protein systems. As mentioned earlier, using predictions of prefactor $A$ based on molecular diffusivities, classical nucleation theory also overpredicts reported nucleation rates. However, classical nucleation theory captures the supersaturation dependence of the nucleation rate. Thus there are consistent and large differences between model predictions and experimental estimates of nucleation rates reported for protein crystals and these differences appear to reside in the poor characterization of the aggregation and dissociation rates of monomers (e.g. the pre-exponential terms in classical nucleation theory).

In both the kinetic and classical nucleation rate models, protein molecules are assumed to be spherical particles interacting with centrosymmetric pair potentials. However, proteins are complex, non-spherical molecules known to experience strong, anisotropic interactions, which might account for the differences between model predictions and experimental estimates of nucleation rates. Models that allow highly directional ‘patchy’ interactions between molecules have been proposed [27,28]. Such interactions make bond formation between two colliding particles less probable because of the additional constraint imposed on their relative orientations. This approach raises the possibility that the low nucleation rates observed in protein systems as compared to small molecule systems might be explained using non-centrosymmetric potentials. However, calculations of nucleation rates incor-
porating patchy interaction potentials have not yet been performed. Extant nucleation models, nevertheless, capture the effects of supersaturation on induction times and are known to describe small molecule crystal nucleation rates well. Therefore, we build on these models to understand the discrepancies between the different experimental estimates of nucleation rates.

4. Analysis of nucleation rate experiments

To understand the discrepancy between the different experimental estimates of nucleation rates, we examine the techniques used and the assumptions made in interpreting the data. We note that the Galkin and Vekilov data yields nucleation rates that are lower than the other estimates by tens of orders of magnitude. Examining their temperature jump technique brings forth several questions that could lead to possible uncertainties in their estimates of nucleation rates: (1) How does the temperature jump from $T_1$ to $T_2$ affect the number of crystals seen at the end of period $t_2$ as compared to the number of crystals nucleated during the period $t_1$? (2) Does nucleation during the entire time period $t_1$ remain steady at constant supersaturation? (3) How significant is crystal growth during the period $t_1$? Within the framework of the nucleation models described earlier, we develop an analysis that answers these questions.

We note at the outset that both, the classical nucleation theory and the kinetic nucleation model, define a critical cluster size $R^*$ and predict that clusters smaller than $R^*$ shrink while clusters bigger than $R^*$ grow. The critical cluster size is a decreasing function of supersaturation, $s = \phi / \phi_s$, being infinite at the solubility boundary, $s = 1$, and approaching the monomer size, $a$, for large $s$. The rate of nucleation is an increasing function of $s$ and thus increases rapidly as $R^*$ decreases. The conclusion of the analysis given later is thus independent of the nucleation rate model used.

In their experiments, Galkin and Vekilov attempt to separate crystal nucleation and growth by holding protein solutions at $T_1$, where $s$ is significant, for a time $t_1$ and then raising the temperature to $T_2$, where $s$ is small. The argument is made that at $T_1$, clusters nucleate but do not grow sufficiently to appreciably reduce $s$. At $T_2$, $s$ is suppressed to a point where Galkin and Vekilov argue that no nucleation occurs but the existing clusters grow. These arguments are used to justify the assumption that all crystals observed at $T_2$ were nucleated at $T_1$ at a constant supersaturation. However, if the rate of nucleation or crystal growth were high, the supersaturation at $T_1$ would decrease and the rate of nucleation at $t_1$ would not be constant. Also, if the supersaturation is not suppressed sufficiently, nucleation could occur at $T_2$. We note that similar experiments are used to determine rates of liquid nucleation from the vapor phase [29]. These experiments are analyzed in a very different manner than proposed by Galkin and Vekilov because of the observed effects of reduced supersaturation during $t_1$, continued nucleation at $T_2$ and coalescence of liquid droplets at $T_1$ and $T_2$.

Following both classical nucleation models and kinetic nucleation models, at constant $\phi$, when $T_1 < T_2$, $R^*(T_1) < R^*(T_2)$. While crystals of size $R_1 = R^*(T_1)$ nucleate at $T_1$, only crystals of size $R_2 = R^*(T_2)$ or greater will survive at $T_2$. Thus, if crystals nucleate and do not grow at $T_1$, when the solution temperature is raised to $T_2$, these crystals will shrink and eventually dissolve. Galkin and Vekilov choose $T_2$ to lie extremely close to the solubility boundary in order to suppress the nucleation rate at this temperature. As a consequence, under these conditions it must be true that $R_2 \gg R_1$. Thus for crystals to survive at $T_2$, a significant amount of crystal growth must occur at $T_1$. A question of importance to the analysis of Galkin and Vekilov then is whether this growth, along with nucleation, reduces $s$ sufficiently to also change the nucleation rate during the time interval $t_1$.

If we let $I(T_1)$ be the steady state nucleation rate at $T_1$ and a volume fraction $\phi$, we can define the nucleation time $t_n = 1/I(T_1)$ as the time characterizing the formation of a crystal nucleus of size $R_1$ in a unit volume. If we assume that within this volume crystals are generated at regular intervals of time, $t_n$, then the number of crystals larger than a size $R > R_1$ at any time $t$ is given by
\[ N(R,t) = \frac{t - t_g(R)}{t_n}, \quad t > t_g(R) + t_n \]  

where \( t_g(R) \) is the time required for a crystal to grow from \( R_i \) to \( R \) at constant \( \phi \).

Galkin and Vekilov vary the nucleation period \( t_i \) and count the number of crystals at the end of the growth period \( t_f \) for each \( t_i \). However, both the kinetic and classical models indicate that after the jump in temperature to \( T_2 \), only crystals of size \( R_2 \) or bigger survive. Thus, if one assumes that no nucleation occurs at \( T_2 \), the number of crystals counted at the end of \( t_f \) must equal \( N(R_2,t_f) \). The nucleation rate calculated by Galkin and Vekilov can then be written as

\[ I_{\text{measured}} = \frac{\Delta N(R_2,t_f)}{\Delta t_f} \]

Since the total number of crystals nucleated during the period \( t_i \) is given by \( N(R_i,t_i) \), the number of crystals that melt at \( T_2 \) is given by

\[ N_{\text{melt}}(t_i) = N(R_1,t_i) - N(R_2,t_i) = \frac{t_g(R_2)}{t_n}, \]

\[ t_1 > t_g(R_2) + t_n \]  

where we use the fact that \( t_g(R_i) = 0 \). \( N_{\text{melt}} \) thus equals the number of crystals nucleated in the time required for the first nucleated crystal to grow from \( R_1 \) to \( R_2 \).

If the time required for a crystal to grow from \( R_1 \) to \( R_2 \) is much smaller than the time to nucleate crystals, i.e. \( t_g(R_2)/t_n < 1 \), then \( N_{\text{melt}} \approx 0 \) suggesting that almost all the crystals nucleated at \( T_1 \) survive at \( T_2 \). In this case, Eq. (12) reduces to

\[ I_{\text{measured}} = \frac{\Delta N(R_i,t_i)}{\Delta t_i}. \]

For small \( t_i > t_n \), where changes in \( \phi \) are negligible, this simplifies to

\[ I_{\text{measured}} = \frac{1}{t_n}, \quad I_{\text{measured}} = I(T_i). \]

Thus, if \( t_g(R_2)/t_n < 1 \) and \( t_1 \) were sufficiently small that nucleation did not alter \( \phi \) at \( T_1 \) over a time interval \( t_i \), the experiments of Galkin and Vekilov would yield accurate nucleation rates.

Wagner and Strey [29] have used similar pressure jump experiments to calculate nucleation rates of water vapor. In agreement with this analysis, they argue that accurate nucleation rate data can be obtained only in the limits of \( t_i \to 0 \) and \( t_g(R_2) \to 0 \) and attempt to tune their experiments accordingly.

In the experiments of Galkin and Vekilov, these limits are difficult to satisfy. For example, \( T_2 \) is chosen to lie close to the solubility boundary so that \( R_2 \to \infty \). This implies that \( R_2 >> R_1 \) and, therefore, that the time required to grow crystals from \( R_1 \) to an arbitrarily large \( R_2 \) would also be arbitrarily large implying that \( t_g(R_2)/t_n \gg 1 \). If \( t_g(R_2)/t_n \gg 1 \), \( N_{\text{melt}} \) is large implying that a significant fraction of the crystals nucleated at \( T_1 \) melt at \( T_2 \).

In addition, in order to see crystals at \( T_2 \) the nucleation time \( t_i \) must be larger than the time for crystals to grow from \( R_1 \) to \( R_2 \) (i.e. \( t_i > t_g(R_2) \)). As a consequence, if we are operating under conditions where the growth time is large in comparison to the time required for a crystal to nucleate, we can also conclude that \( t_i >> t_n \). This implies that a huge number of crystals will nucleate during the period \( t_i \), which, along with the growth of these crystals, will produce significant changes in \( \phi/\phi_sef \). Thus if \( T_2 \) is chosen to lie close to the solubility boundary such that no nucleation occurs during \( t_2 \), the hypothesis of Galkin and Vekilov that nucleation occurs at constant supersaturation throughout the period \( t_i \) cannot be true. Under these conditions, with \( t_i \) being large, the number of crystals formed and capable of growing to \( R_2 \) is governed by a much smaller value of \( s \) than the initial value, say \( s_0 \). At this smaller supersaturation, the growth rate of crystals is greatly reduced. On warming to \( T_2 \), a very limited number of crystals will have grown from \( R_1 \) to \( R_2 \) and thus most of the crystals nucleated at \( T_1 \) will melt. If most of the crystals melt, the apparent nucleation rate in Eq. (12) will dramatically underpredict the actual nucleation rate.

To illustrate these ideas further, we consider the same set of conditions of the Galkin and Vekilov experiments shown in Fig. 2. Fig. 3(a) shows the nucleation rates and Fig. 3(b), the values of \( R^* \), under these conditions, at the temperatures \( T_1 \) and \( T_2 \), calculated using the model of Dixit and Zukoski. At \( \phi = 0.02 \), the nucleation rates at \( T_1 \) and \( T_2 \) are \( 5.0 \times 10^{17} \) and \( 4.0 \times 10^{15} \) cm\(^{-3}\) s\(^{-1}\), respectively, indicating that the nucleation rate is suppressed by more than 2 orders of magnitude at \( T_2 \) in accordance with the expectation of Galkin and Vekilov. At \( \phi = 0.05 \), corresponding to a lysozyme concentration of 58.6 mg ml\(^{-1}\), \( I(T_1) = \ldots \)
5.0 × 10^{20} and I(T_2) = 1.1 \times 10^{20} \text{ cm}^{-3} \text{s}^{-1}, indicating a much lesser suppression at higher supersaturations. At these temperatures, for \( \phi = 0.05 \), the critical sizes are: \( R_1 = R^*(T_1) = 3.88 \times 10^{-7} \) cm and \( R_2 = R^*(T_2) = 4.20 \times 10^{-7} \) cm.

Assuming that crystals grow at constant supersaturation, their growth rate can be calculated as

\[
\frac{dR}{dt} = \frac{a^3}{3\phi_{cp}R^2} (\beta - x)
\]

where \( \beta = \beta(R, \phi) \) and \( x = x(R) \) are the rates of aggregation and dissociation defined earlier, \( a \) is the radius of a monomer and \( \phi_{cp} \) is the packing fraction of the cluster. The model of Dixit and Zukoski [18] provides expressions for \( \beta \) and \( \alpha \) for square well fluids used to predict the phase diagram in Fig. 2. Holding \( \phi \) constant at 0.05 and integrating the above equation gives the crystal size \( R \) as a function of time \( t \). Fig. 4 presents \( R(t) \) with \( t \) in units of \( t_n = 1/I(T_1) \). We note that under the conditions considered, \( t_g(R_2) = 8.0 \times 10^{14} t_n \) in agreement with our analysis earlier that \( t_g(R_2)/t_n \gg 1 \). Thus, by the time the first crystal grows to the size \( R_2 \), nearly \( 10^{15} \) more crystals are formed in the solution, lowering \( \phi \) tremendously. Since all the crystals formed must be of size \( R_1 \) or greater, the above nucleation rate results in a 50% reduction in \( \phi \) within \( 10^{17} t_n \) or \( 2 \times 10^{-4} \text{s} \).

Darcy and Wiencek record a nearly 30% reduction in protein concentration at the end of \( t_1 = 8 \) h, showing that the above calculations, although quantitatively inaccurate, capture the qualitative picture correctly. Galkin and Vekilov used the nucleation periods 10 min < \( t_1 < 8 \) h. Thus, the above model calculations and the data of Darcy and Wiencek agree that in the Galkin and Vekilov experiments as well, nucleation is unlikely to occur at steady state at constant \( \phi \) over the range of times studied by Galkin and Vekilov. As a result, \( I_{\text{measured}} \) is unlikely to equal \( I(T_1) \) and is in all likelihood only a small fraction of \( I(T_1) \) corresponding to a much lower value of \( \phi \). Identifying this fraction to back-calculate \( I(T_1) \) goes back to solving the JMAK integral (Eqs. (1) and (2)) which defeats the basic purpose of the Galkin and Vekilov experiments.

The recognition of the reducing supersaturation during the period \( t_1 \) can also be seen in the way Wagner and Strey [29] analyze data from their pressure jump experiments for nucleation rates. From measured light scattering signals at two angles, Wagner and Strey use the Mei theory to

---

**Fig. 3.** (a) Shows the nucleation rates and (b) the critical cluster sizes under the conditions used by Galkin and Vekilov [10], shown in Fig. 2, calculated using the kinetic nucleation model.
Fig. 4. Growth of a cluster from size $R_i(T_1)$ to $R_i(T_2)$ at a fixed protein volume fraction in the solution. $R_i(T)$ is the critical cluster size at a temperature $T$ at the same volume fraction. The time required for this growth is normalized by the characteristic nucleation time $t_n$.

deduce the number and size of the liquid droplets formed during the period $t_1$. To avoid droplet coalescence and secondary nucleation effects, they use the signal scattered just after the pressure jump. An experimental nucleation rate is then calculated as a ratio of the number of droplets and the time of the pressure pulse $t_1$. In their experiments, $t_1$ was of the order of 1 ms. Yet, to compare their experimental nucleation rate with theory, they recognize that the supersaturation varies with time and allow for this variation in their calculations. Using classical nucleation theory, they integrate the nucleation rate over the period $t_1$ assuming an appropriate time variant supersaturation profile to obtain the theoretical number of droplets that would be formed at the end of $t_1$. Dividing this number by $t_1$ gives them an average nucleation rate predicted by theory, which they found to compare well with their experimental rates.

The qualitative calculations above suggest that when $t_2 \gg t_n$ (since $R_2$ is arbitrarily large) the number of crystals observed at the end of time $t_2$ in the experiments of Galkin and Vekilov is much smaller than that which existed at the end of $t_1$. This can be tested experimentally by looking for the melting of crystals when the temperature is raised from $T_1$ to $T_2$. Note that if $R_2$ is not chosen arbitrarily large, it must be large enough that no crystals nucleate within the time interval $t_2$ (several months) when the protein solution at $\phi$ is held at $T_2$. Thus, while the above arguments are qualitative at this point, they are subject to experimental test. If no crystals nucleate in a solution at $T_2$ in a time much greater than $t_2$, $R_2$ must be very large. However, if crystals are seen in a time much smaller than $t_2$, $R_2$ must be at least comparable to $R_1$. Under these conditions, one must be concerned about nucleation at $T_2$ as contributing to the number of observed crystals. Experiments to illustrate these points were conducted in the following manner.

Lysozyme solution (3% NaCl salt, 0.05 M NaAc buffer, pH 4.5) containing 55 mg ml$^{-1}$ of lysozyme was made up and placed in a 100 µl scattering cell [13]. The light scattered at 90° from the laser was detected with a photodiode. While the cell was loaded, samples were held at 40°C, which lies above their solubility temperature. At $t=0$, the cell was quenched to $T_1$. The intensity of the scattered light was observed as a function of time. Nucleation was said to occur when there was a sudden increase in the intensity of the scattered light. As shown in Fig. 5, when $T_1 = 13°C$, nucleation was observed after 7 min while at 22°C, nucleation occurred after 2 h. Note that this suggests that the temperature $T_2$ used by Galkin and Vekilov was not sufficiently close to the solubility boundary to suppress nucleation if the concentration of protein is held constant. As a consequence, $R_2$ is not arbitrarily large and even if no growth occurred at $T_1$, nucleation at $T_2$ is a distinct possibility.

The intensity of scattered light depends on the size and concentration of scattering units. If all crystals can grow from $R_1$ to $R_2$ within the time $t_1$, then on increasing the temperature from $T_1$ to $T_2$, the intensity of scattered light should continue to increase monotonically. However, if some of the crystals have not reached a size $R_2$, stable at $T_2$, they will melt and the intensity of scattered
light will decrease. As shown in Fig. 6, if the temperature in the cell is held at 13°C for a period of 10 min and then raised to 22°C the scattered light intensity passes through a maximum. After a period of time, the scattering intensity increases rapidly a second time suggesting that a secondary nucleation event has taken place. The maximum in scattered intensity demonstrates that on raising the temperature to \( T_2 \), there is a decrease in the number and/or size of the scattering units and provides strong evidence that not all the clusters have grown to a size to be stable at \( T_2 \), and thus dissolve once the temperature is raised to \( T_2 \). The secondary nucleation event suggests that nucleation does occur during the period \( t_2 \).

Galkin and Vekilov carried out extensive experiments to eliminate the effects of heterogeneous nucleation. In our experiments, the induction times, shown in Fig. 5, were very reproducible for independent fillings of the scattering cell. This gives some confidence that heterogeneous nucleation does not dominate our signal. However, even if heterogeneous nucleation occurred in our system, the dissolving of some of the crystals on increasing the temperature from \( T_1 \) to \( T_2 \) indicates that not all the particles nucleated at \( T_1 \) will be seen after a long period of time at \( T_2 \). From our analysis and experiments, we conclude that the Galkin and Vekilov data is subject to large uncertainties and its comparisons with model predictions therefore remain inconclusive.

Examining the techniques of Kulkarni and Zukoski and Darcy and Wiencek also reveals possible uncertainties. In interpreting the heat release data of Darcy and Wiencek as nucleation rates, we have assumed that during the period \( t_{\text{max}} \), where the power released reaches a maximum, only nucleation and no growth of crystals occurs. Also, we have neglected possible heterogeneous nucleation. As a result, nucleation rates given in Fig. 1 derived from the data of Darcy and Wiencek provide, at best, an upper bound on the homogeneous nucleation rate of lysozyme crystals.

The estimates of nucleation rates obtained from the induction time data of Kulkarni and Zukoski may also have several uncertainties. The assump-
tion of the classical model that the dependence of induction time on $\Delta E^*$ is identical to that of the nucleation rate, holds only when the time for nucleation is at most on the order of the growth time for the nucleated crystal to become detectable in the nomenclature of Sohnel and Mullin described earlier and this happens when $t_g \ll t_n$ or $t_g \sim t_n$. That the assumption of the classical model holds under the experimental conditions of Kulkarni and Zukoski is proven by the remarkable agreement the induction time data shows with this model. Estimates of surface tension determined by assuming the classical form for $\Delta E^*$ are in agreement with other experimental and theoretical estimates. However, substantial uncertainties arise from a limited knowledge of the prefactors $A$ and $B$.

5. Conclusions

Large discrepancies exist in independent experimental estimates of nucleation rates. The assumptions used in interpreting the data in each experiment as nucleation rates are also subject to uncertainties. The calorimetric data of Darcy and Wiencek provides an upper bound at best of the rate of homogeneous nucleation. The Kulkarni and Zukoski data is interpreted using classical nucleation theory where the prefactors are poorly understood. The Galkin and Vekilov data has large uncertainties arising from the way the temperature jump technique is implemented and analyzed. On the other hand, we also conclude that models for nucleation rates consistently overpredict experimental estimates of protein crystal nucleation rates. One possible explanation lies in the simplicity of the pair potentials used for describing protein interactions. Centrosymmetric interaction potentials are certainly an oversimplification of protein interactions. More sophisticated patchy interactions that better approximate protein interactions with resulting modifications in the nucleation mechanism might provide an answer. Another explanation may lie in the poor descriptions of the rate processes of protein attachment to and dissociation from a crystal nucleus. In classical models, these uncertainties are concentrated in the prefactor and are thus purely kinetic in nature. However, in the kinetic nucleation model, these are tied to the equilibrium thermodynamic phase behavior thus greatly limiting the parameters that can be adjusted to ensure model agreement with both equilibrium phase behavior and the kinetics of phase transformations.

Acknowledgements

The support of NASA Grant NA G8-1376 for this work is gratefully acknowledged. We also thank Richard McClurg for helpful discussions.

References