# Nucleation of ordered solid phases of proteins via a disordered high-density state: Phenomenological approach

Cite as: J. Chem. Phys. **122**, 174905 (2005); https://doi.org/10.1063/1.1887168
Submitted: 28 September 2004 • Accepted: 15 February 2005 • Published Online: 04 May 2005

**Weichun Pan, Anatoly B. Kolomeisky and Peter G. Vekilov**

# Nucleation of ordered solid phases of proteins via a disordered high-density state: Phenomenological approach

Weichun Pan
*Department of Chemical Engineering, University of Houston, Houston, Texas 77204-4004*

Anatoly B. Kolomeisky
*Department of Chemistry, Rice University, Houston, Texas 77005-1892*

Peter G. Vekilov
*Department of Chemical Engineering, University of Houston, Houston, Texas 77204-4004*

(Received 28 September 2004; accepted 15 February 2005; published online 4 May 2005)

Nucleation of ordered solid phases of proteins triggers numerous phenomena in laboratory, industry, and in healthy and sick organisms. Recent simulations and experiments with protein crystals suggest that the formation of an ordered crystalline nucleus is preceded by a disordered high-density cluster, akin to a droplet of high-density liquid that has been observed with some proteins; this mechanism allowed a qualitative explanation of recorded complex nucleation kinetics curves. Here, we present a simple phenomenological theory that takes into account intermediate high-density metastable states in the nucleation process. Nucleation rate data at varying temperature and protein concentration are reproduced with high fidelity using literature values of the thermodynamic and kinetic parameters of the system. Our calculations show that the growth rate of the near-critical and supercritical ordered clusters within the dense intermediate is a major factor for the overall nucleation rate. This highlights the role of viscosity within the dense intermediate for the formation of the ordered nucleus. The model provides an understanding of the action of additives that delay or accelerate nucleation and presents a framework within which the nucleation of other ordered protein solid phases, e.g., the sickle cell hemoglobin polymers, can be analyzed. © *2005 American Institute of Physics*. [DOI: 10.1063/1.1887168]

## I. INTRODUCTION

The formation of protein ordered phases is of interest to many scientific and technological areas: structural biology,[1–3] patho-physiology of protein condensation diseases,[4–6] production of protein pharmaceutical preparations,[7,8] etc., and a deeper understanding of the underlying mechanisms is expected to facilitate progress in these areas.

The phase diagrams of protein solutions differ from those of simple substances, such as argon. The significant factor underlying this difference is the range of attraction between molecules:[9–11] this range is determined by the size of the solvent molecules[12] (at the typically used ionic strengths of $\geqslant 0.1$ M the electrostatics is screened and the Deriaguine–Landau–Veervey–Overback theory does not apply[13,14]), smaller by at least an order of magnitude than the protein molecules' sizes. As a result, in protein solutions the triple point disappears and the liquid-liquid (*L*-*L*) phase separation is submerged below the solution-crystal equilibrium.[15,16] Phase diagrams with a *L*-*L* separation line with a critical point lying below a smooth liquidus and a solidus lines have been found for several proteins.[17–21]

An investigation of the crystal nucleation mechanisms using the protein lysozyme revealed an unusual dependence of the rate of homogeneous nucleation on temperature: as the temperature is lowered from the equilibrium, the dependence passes through a maximum in the vicinity of liquid-liquid phase boundary,[22] Fig. 1. Another unusual result with the same system was that as supersaturation is increased, the nucleus size decreases and reaches one molecule.[23] This transition occurs at conditions that differ by 1.0 °C–1.5 °C (or $\sim$10–15 mg ml$^{-1}$) from the *L*-*L* coexistence line in the (temperature, protein concentration) plane and indicates that the nucleation free-energy barrier becomes less than the thermal energy and the rate of nucleation is solely limited by the kinetics of growth of the near-critical clusters.[24–26] These two observations contradict the assumptions and predictions of the classical nucleation theory.[27]

Prompted by computer simulations[28] and density functional calculations[29] results in the vicinity of the *L*-*L* critical point, the above unusual behavior of the nucleation rate was explained by the presence of a dense liquid intermediate in the nucleation reaction pathway, Fig. 2.[22,30] Since the nucleation rate maximum is consistently reached above the *L*-*L* coexistence line, it was concluded that a dense-liquid-like droplet, unstable or metastable with respect to the low density solution, serves as an intermediate in that region, Fig. 2.[30] Similar interpretations of these results were offered in several theoretical and computational investigations of protein crystallization.[31–33] Note also that the protein nucleation is a complex phenomenon that cannot be described by a single order parameter. Thus Fig. 2 is a simplified formal presentation of a free-energy landscape of the system, and the nucleation reaction coordinate might include several order parameters.

In this paper we develop a simple phenomenological

 

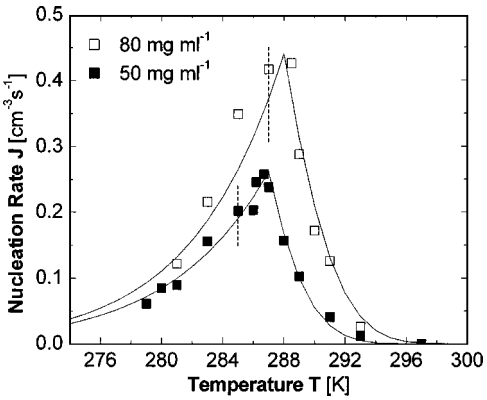


FIG. 1. The dependence of the rate of homogeneous nucleation $J$ of lysozyme crystals on temperature $T$ at two fixed lysozyme concentration indicated in the plot. The temperatures of equilibrium between crystals and solution are 315 K at $C_{\mathrm{lys}}=50$ mg ml$^{-1}$ and 319 K at $C_{\mathrm{lys}}=80$ mg ml$^{-1}$. The temperatures of $L$-$L$ separation are 285 K at $C_{\mathrm{lys}}=50$ mg ml$^{-1}$ and 287 K at $C_{\mathrm{lys}}=80$ mg ml$^{-1}$ (Ref. 37) and are marked with vertical dashed lines. Symbols represent experimental results from Ref. 22. Lines are results of the model discussed in the text.

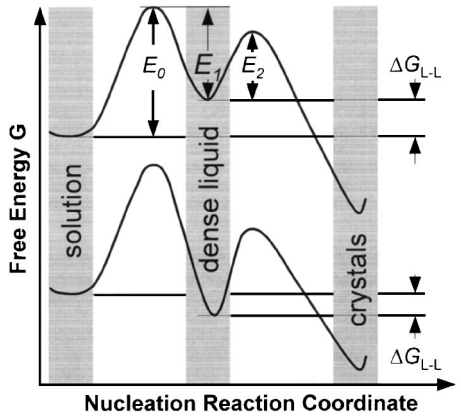


FIG. 2. The free energy $G$ along two possible pathways for nucleation of crystals from solution. $E_0$ is the barrier for formation of a quasidroplet of dense liquid as a result of a density fluctuation, $E_0$ is the barrier for decay of these droplets, $E_2$ the barrier for formation of an ordered cluster as a result of a structure fluctuation within the dense quasidroplet. $\Delta G_{L\text{-}L}$ is the free energy of formation of the dense liquid phase, $\Delta G_{L\text{-}L}>0$ above the $L$-$L$ coexistence line and upper curve applies; $\Delta G_{L\text{-}L}<0$ below the $L$-$L$ coexistence line, reflected by lower curve. See text for discussion of $E_1$ above the $L$-$L$ coexistence line.

model of protein crystallization via an intermediate liquid state. Our goal is a quantitative understanding of the dependence of nucleation on temperature and concentration using parameters that can be measured experimentally.

## II. THE MODEL

### A. The nucleation rate law

Our main assumption is that in the supersaturated dilute solution a liquidlike cluster is formed with a temperature-dependent and protein-concentration-dependent rate $u_0(C,T)$. This cluster can dissociate back into the solution with rate $u_1(T)$, or it can transform into an ordered crystal nucleus with rate $u_2(T)$. The crystal nucleus irreversibly grows to a macroscopic ordered phase. These processes can be formally described by the following rate scheme:

$$0 \leftrightarrows 1 \rightarrow 2, \tag{1}$$

where state 0 corresponds to the dilute solution, state 1 is the intermediate dense liquid cluster, and state 2 is the final crystalline phase. These three states correspond to the minima in the energy landscape picture in Fig. 2.

The thermodynamic and kinetic parameters of the intermediate and the depth $E_1$ of the second minimum determine the nucleation rate resulting form this model. For all temperatures above the critical $T_c$ for $L$-$L$ separation, we assume that the system, via concentration fluctuations, selects the intermediate state leading to fastest nucleation of crystals; $E_1<E_0$ so that the intermediate is metastable or unstable with respect to the dilute solution. At $T<T_c$, we test two possibilities: (i) state 1 is the same as the dense phase at this temperature; (ii) state 1 is selected according to the same criteria as at $T>T_c$. $E_1$ becomes greater than $E_0$ not at $T_c$ but at $T_{L\text{-}L}$ at which the chosen dilute solution is in equilibrium with a dense liquid. In all cases, the intermediate has a higher free energy than the final crystalline state.

If we define $P_i(t)$ as a probability to find the system in state $i=0,1,$ or 2 at time $t$, then a parameter $\tau = \int_0^\infty t[dP_2(t)/dt]dt$ determines the mean time to create one crystalline nucleus in a steady state process. Thus, the parameter $\tau$ represents a mean *first-passage time* for the transition from state 0 to state 2 and is given by[34]

$$\tau = \frac{1}{u_0(C,T)} + \frac{u_1(T)}{u_0(C,T)u_2(T)} + \frac{1}{u_2(T)}. \tag{2}$$

The rates $u_0$, $u_1$, and $u_2$ depend on temperature as $u_i(T)=U_i \exp(-E_i/k_BT)$, $i=0,1,2$. As a first simple approximation, the steady-state nucleation rate $J$ can be calculated as $J=\tau^{-1}$. We get for $J$

$$J = \frac{U_0U_2 \exp\left(\frac{E_0+E_2}{k_BT}\right)}{U_0 \exp\left(-\frac{E_0}{k_BT}\right) + U_1 \exp\left(-\frac{E_1}{k_BT}\right) + U_2 \exp\left(-\frac{E_2}{k_BT}\right)}. \tag{3}$$

This is a general expression for the nucleation rate with one intermediate state. Note that it differs significantly from a classical nucleation theory formula with a single temperature-dependent exponent.[27]

Next, we divide the numerator and denominator on the left side of Eq. (3) by $U_0 \exp(-E_0/k_BT)$, and define a Gibbs free-energy change for the formation of the intermediate $\Delta G=E_0-E_1$. Since the rate of nucleation of liquid droplets $O(10^7-10^9$ cm$^3$ s$^{-1})$ (Ref. 35) is significantly faster than the typical rates of protein crystal nucleation $O(10^{-2}-1$ cm$^3$ s$^{-1})$,[36] we assume that $u_2(T)\ll u_0(T)$, i.e., the structuring of the dense liquid droplets is a rate-limiting step. Equation (3) simplifies to

$$J = \frac{U_2 \exp\left(-\frac{E_2}{k_BT}\right)}{1+\frac{U_1}{U_0}\exp\left(\frac{\Delta G}{k_BT}\right)}. \tag{4}$$

Equation (4) shows that $J$ does not explicitly depend on the size of the barrier for decay of the dense liquid intermediate: even if $E_1=0$ and the intermediate is unstable (this corresponds to a typical density fluctuation), the expression for $J$ does not change.

### B. The maximum in the nucleation rate

Assuming that $U_1/U_0$, $U_2$, and $E_2$ are not functions of temperature, we solve the equation $(\partial J/\partial T)_{T^*}=0$ and find that the nucleation rate reaches a maximum at a temperature $T^*$ determined from

$$\frac{E_2}{k_B T^{*2}}\left[1+\frac{\exp\left(-\frac{\Delta G}{k_B T^*}\right)}{U_1/U_0}\right]=\left|\frac{\partial}{\partial T}\left(\frac{\Delta G}{k_B T}\right)\right|_{T^*}. \tag{5}$$

If $\Delta G$ is independent of temperature, this expressions yields a $T^*$ which is not physically meaningful.

For a more realistic description, we allow $E_0$ and $E_1$ to change with temperature. We assume that their difference $\Delta G=0$ at $T_{L\text{-}L}$, and $\Delta G/k_BT$ changes linearly with $T$ both above and below $T_{L\text{-}L}$. To find the increment of this dependence, we assume that it equals the temperature increment of the dependence of the free-energy change for $L$-$L$ separation $\Delta G_{L\text{-}L}$ below the critical temperature for $L$-$L$ separation $T_c$.[37] To find this $\Delta G_{L\text{-}L}$ increment, we use the known strong dependence of enthalpy of $L$-$L$ separation $\Delta h_{L\text{-}L}$ on temperature that yields $\Delta h_{L\text{-}L}=-40$ kJ/mol at $T_c-T\approx 10$ K.[37] Integrating numerically the Gibbs–Helmholtz equation with this $\Delta h_{L\text{-}L}(T)$, we find that $\Delta G_{L\text{-}L}$ is an increasing power-law function of $T_c-T$. Approximating this $\Delta G_{L\text{-}L}(T)$ dependence with $\Delta G_{L\text{-}L}/k_BT=AT+B$, we get $A=0.0666$. Using the same $A$ for $\Delta G/k_BT$ at $T$ both above and below $T_{L\text{-}L}$, we find that $B=-18.9$ ensures that $\Delta G=0$ at $T_{L\text{-}L}$ for $C=50$ mg/ml, while $B=-19.2$ for $C=80$ mg/ml, the two concentrations of the data in Fig. 1. Then, from Eq. (5) with $E_2\sim k_BT$, $T^*\sim 170$ K, far below that observed experimentally. More importantly, we find (model results not shown) that the nucleation rate $J$ egresses toward the maximum at $T^*$ upon temperature decrease not by a steep exponential as in Fig. 1, but rather by a weak sublinear function.

### C. Using the system parameters

The above considerations show that a model that could adequately describe the data in Fig. 1 should account for the physical specificity of the system and of the chosen kinetic model, as well as for the temperature dependence of $E_2$.

A crucial element of the model is the nature of the intermediate state. We assume that is akin to the dense liquid and exists at the same temperature as the dilute solution. This allows us use the above linear dependence $\Delta G(T)$ on both sides of $T_{L\text{-}L}$, and this is equivalent to assuming that $E_1\neq 0$ at $T>T_{L\text{-}L}$, i.e., the intermediate is not a mere density fluctuation. The alternative assumption would lead to a break in $\Delta G$ at $T_{L\text{-}L}$ and a discontinuity in the $J(T)$ dependence. The absence of such discontinuity in Fig. 1 suggests that the intermediate is metastable and has a finite lifetime. Experimental evidence of the existence of such a metastable intermediate has now been found for three protein, including lysozyme, data which are modeled here, using dynamic light scattering.[38]

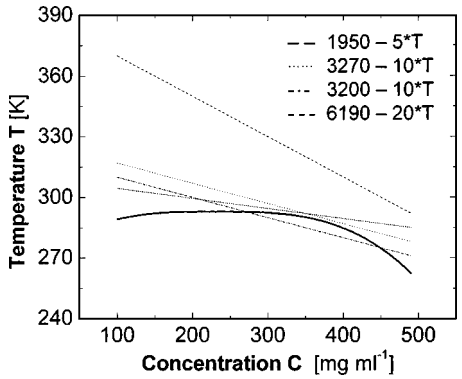


FIG. 3. Assumed value of the concentration of the dense phase intermediate. Dashed, dotted, dash-dotted, and short-dashed lines correspond to different assumptions of this concentration, as indicated in the plot. Solid line marks liquid-liquid coexistence line and is shown for comparison.

While these tests show that the size of the metastable liquid droplets is from a few tens to a few hundred nanometers, neither this nor any other technique can provide an estimate of the protein concentration in these droplets. We assume that at each temperature, an intermediate state with a concentration $C_1$ that does not depend on the dilute solution concentration is selected and that $C_1$ obeys

$$C_1=A_1T+B_1. \tag{6}$$

Below, we show that the nucleation rate law depends only weakly on the exact values of $A_1$ and $B_1$ as long as the resulting $C_1$ in the range $T<T_c$ is in the vicinity of the equilibrium concentration of the dense liquid. The $C_1(T)$ dependencies tested below are depicted in Fig. 3.

The preexponential factor $U_2$ in Eq. (4) accounts for the kinetics of growth of ordered clusters within the dense liquid droplet, and it should depend on the temperature, concentration, and viscosity within the droplet. In analogy to the preexponential factors for nucleation of solids in melts, we assume that $U_2$ is proportional to $C_1$ and $T$ and inversely proportional to viscosity in the dense liquid $\eta$ at concentration $C_1$ (Ref. 39),

$$U_2=k_2\frac{C_1T}{\eta(C_1,T)}, \tag{7}$$

where $k_2$ is a kinetic constant.

We assume that $\eta$ follows the same dependence in the dilute solution and in the dense liquid and represent $\eta(C,T)$ as the product of the Andrade–Eyring expression for the temperature dependence[40] and an empirical expression for the concentration dependence.[41] We get

$$\eta=\eta_0\{1+[\eta]C_1\exp(k_\eta[\eta]C_1)\}\exp(-E_\eta/k_BT), \tag{8}$$

where $\eta_0$ is the viscosity of the solvent, $[\eta]$ is the viscosity increment at low concentrations, $E_\eta$ is the Arrhenius-type temperature factor for the viscosity, and $k_\eta$ is a parameter typically determined by fitting $\eta(C)$ dependencies to Eq. (8); $k_\eta$ takes values form 1 to 10.[41] The former three parameters

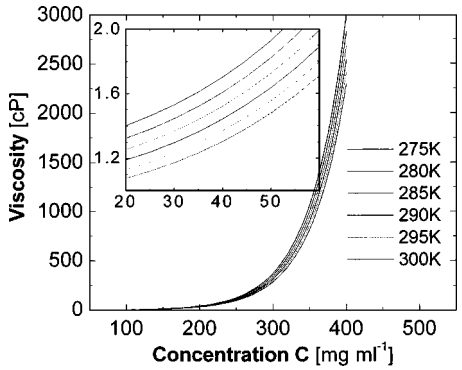


FIG. 4. The dependencies of the viscosity $\eta$ on protein concentration $C_{\mathrm{lys}}$ and temperature $T$ according to Eq. (8) assumed in the model evaluations. The dependencies at low concentrations shown in the inset were fitted to experimental determinations of $\eta(C_{\mathrm{lys}},T)$ from Ref. 42.

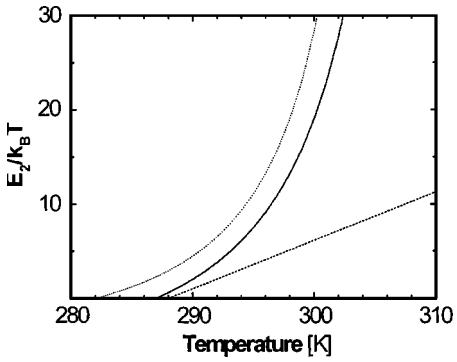


FIG. 5. The dependencies of the barrier $E_2$ for nucleation of crystals from the dense phase intermediate on temperature $T$, calculated according to Eq. (9) using solution-crystal equilibrium temperature $T_e$ and crystallization spinodal temperature $T_{\mathrm{sp}}$ from Table I, for three of the modeled cases: upper dotted line: 80 mg ml$^{-1}$ lysozyme, 5% glycerol; intermediate solid line: 50 mg ml$^{-1}$ lysozyme; lower dashed line: 80 mg ml$^{-1}$ lysozme.

can be determined by fitting Eq. (8) to the data on the viscosity of low-concentration lysozyme solutions at various temperatures.[42] Thus, in all cases $\eta_0=0.051$ cP, $[\eta]=0.0045$ mg$^{-1}$ ml, and $E_\eta=-7.3$ kJ/mol. The resulting $\eta(C,T)$ curves with $k_\eta=4$ are shown in Fig. 4. As expected, they diverge as $C$ approaches 600 mg ml$^{-1}$ with a corresponding volume fraction 0.49, the noninteracting hard sphere limit.[41]

The expression for $u_0$ in Eq. (4) above is the rate of nucleation of dense liquid droplets and thus the preexponential factor $U_0$ is a linear function of the dilute solution concentration.[43,44] Since we know nothing about $U_1$ we assume that it is constant, so that the ratio $U_1/U_0=1$ for $C=50$ mg ml$^{-1}$ and $50/C$ for other dilute solution concentrations $C$.

### D. The barrier $E_2$

The steep dependence of $J$ on the deviation of the temperature from the value at solution-crystal equilibrium suggests that the barrier $E_2$ should be a decreasing function of temperature. Furthermore, comparing the temperatures of the maxima in Fig. 1 at the two studied concentrations to locations in the $(C,T)$ plane, where the nucleus size decreases to one molecule and the nucleation barrier vanishes,[23,38] we find that they coincide. In analogy to the similar phase locations in phase diagrams of fluids,[24–26] we assign this temperature to the crystallization spinodal and denote it with $T_{\mathrm{sp}}$. Note that this spinodal is kinetically determined and denotes a location where the solution loses stability with respect to a combination of concentration and structure fluctuations. This differs from the spinodal (or pseudospinodal[45]) for *L-L* separation introduced in Ref. 37, which may be related to stability loss by concentration fluctuations alone.

We assume that between $T_e$ and $T_{\mathrm{sp}}$, $E_2$ smoothly decreases from infinity to zero, and in analogy to a recently introduced law,[46] follows a $\Delta T^{-2}$ dependence,

$$E_2(T)=\frac{E^*}{(T_e-T)^2}\left[1-\frac{(T_e-T)^2}{(T_e-T_{\mathrm{sp}})^2}\right]. \tag{9}$$

The parameter $E^*$ is determined from the following considerations. At $T$ higher than $T_{L\text{-}L}$ by several degrees, the nucleation barrier should be $\approx n^*\Delta\mu/2$, where $n^*$ is the nucleus size and $\Delta\mu$ is the thermodynamic supersaturation. Since $\Delta\mu/k_BT\approx 2-3$ and $n^*$ is between 4 and 10,[23] the barrier should be several $k_BT$ units. Using the values of $T_e$ and $T_{\mathrm{sp}}$ from Table I and assuming $E^*=15\,000$ kJ mol$^{-1}$, we get the $E_2(T)$ dependencies shown in Fig. 5, which fulfill this requirement. Note that $E^*$ corresponds to the nucleation barrier $E_2$ only in the immediate vicinity of $T_e$, where $E_2$ approaches infinity. Thus, the high value of $E^*$ allows a strong dependence of $E_2$ on $\Delta T$ via Eq. (9) that results in $E_2\approx 50$ kJ mol$^{-1}$ at $T=300$ K, where nucleation is still extremely slow, see Fig. 1.

## III. MODEL EVALUATION

### A. Pure solutions

Lysozyme solutions represent a unique protein system, for which the available data provide values or reasonable estimates for most of the parameters of the above model. The

TABLE I. Thermodynamic characteristics: temperatures of solid-liquid equilibrium $T_e$, liquid-liquid equilibrium $T_{L\text{-}L}$, and the crystallization spinodal (see text for definition) $T_{\mathrm{sp}}$ used in the model calculations, all three from Refs. 22 and 37. The kinetic parameters: kinetic coefficients for viscosity $k_\eta$ and structuring of intermediate $k_2$, and scaling factor for second free-energy barrier $E^*$ used in model calculations. In all cases here, the parameters in the $C_1(T)$ dependence in Eq. (6) are $A_1=-10$ and $B_1=3270$.

| | Model characteristics | $T_{L\text{-}L}$(K) | $T_{\mathrm{sp}}$(K) | $T_e$(K) | $k_\eta$ | $k_2$ | $E^*$(kJ/mol) |
|---|---|---|---|---|---|---|---|
| 1 | $C=50$ mg ml$^{-1}$ | 285 | 287 | 315 | 4 | 0.266 | 15 000 |
| 2 | $C=80$ mg ml$^{-1}$ | 287 | 288 | 319 | 4 | 0.266 | 15 000 |
| 3 | $C=50$ mg ml$^{-1}$, 5% glycerol | 278 | 281 | 311 | 4 | 1.862 | 10 000 |
| 4 | $C=50$ mg ml$^{-1}$, 0.2% PEG | 285 | 287 | 315 | 4 | 0.8911 | 15 000 |

TABLE II. Values of the parameters used in model calculations probing the effects of the choice of concentration of intermediate $C_1$ on the nucleation rate law; thermodynamic characteristics $T_{L\text{-}L}$, $T_{sp}$, and $T_e$ as in case 1 in Table I.

| | | $C_1$ | | | | |
|---|---|---|---|---|---|---|
| | Model characteristics | $A_1$ | $B_1$ | $k_\eta$ | $k_2$ | $E^*$ (kJ/mol) |
| 5 | $C=50$ mg ml$^{-1}$ and $C=80$ mg ml$^{-1}$, $C_1$ follows $T_{L\text{-}L}(C)$ | −10 | 3200 | 7 | 42.56 | 15 000 |
| 6 | $C=50$ mg ml$^{-1}$ and $C=80$ mg ml$^{-1}$ | −10 | 3200 | 4 | 0.077 | 18 000 |
| 7 | $C=50$ mg ml$^{-1}$ and $C=80$ mg ml$^{-1}$ | −5 | 1950 | 7.5 | 1284 | 15 000 |
| 8 | $C=50$ mg ml$^{-1}$ and $C=80$ mg ml$^{-1}$ | −20 | 6190 | 2 | 0.0114 | 15 000 |

only materials parameter missing is $k_\eta$, accounting for high-concentration effects in the dependence of the viscosity of the dense liquid intermediate on concentration and temperature in Eq. (8). Thus, $k_\eta$ remains an adjustable parameter of the model. The other inevitably (due to our low understanding of nucleation in dense liquids) adjustable parameter is the coefficient $k_2$ in the preexponential factor for the ordering of the dense liquid intermediate $U_2$ in Eq. (7). Using the values of the respective characteristic temperatures $T_e$, $T_{L\text{-}L}$, and $T_{sp}$ and the parameters listed in Table I, we fit the model predictions to the data at $C=50$ mg ml$^{-1}$ in Fig. 1. The values of $k_2$ and $k_\eta$ yielding the best fit are shown in Table I. Note that all values of $k_\eta$ are within the range 1–10 found with other solutions for which the viscosity at high concentration has been studied.[41]

Using the characteristic temperatures for $C=80$ mg ml$^{-1}$ and the same kinetic parameters as those for $C=50$ mg ml$^{-1}$, we evaluated the model for the former protein concentration. Figure 1 shows that the correspondence between the model prediction and the actual data for $C=80$ mg ml$^{-1}$ is remarkably good.

### B. The concentration $C_1$ in the dense liquid intermediate

To evaluate the significance of the accuracy of the guess of the concentration within the dense liquid intermediate, we evaluated the dependence $J(C_1)$ stemming from Eqs. (4), (7), and (8) at several $T=$const. We got very shallow maxima at $C_1$'s near and to the right of the $L$-$L$ coexistence line. The increase in $J$ to the left of these maxima is due to the dependence in Eq. (7), while the decrease to the right of maxima, to Eq. (8). The shallowness of the maxima indicates that the exact selection of $C_1$ is not crucial for the resulting value of the nucleation rate. Since the nucleation follows the fastest pathway, we chose the parameters $A_1=-10$ and $B_1=3270$ in Eq. (6) that ensure $C_1$ values in the region of these maxima.

As a second test of the significance of $C_1$, we evaluated the model with different values of the parameters $A_1$ and $B_1$ in Eq. (6). If the tested $C_1(T)$ dependence runs near to the case modeled above, such as cases 6 and 7 in Table II and Fig. 3, slight adjustments of the values of the parameters $k_\eta$ and $E^*$ within their acceptable physical ranges ensure a perfect fit of the model to the experimental results for both $C=50$ mg ml$^{-1}$ and $C=80$ mg ml$^{-1}$.

However, if the tested $C_1(T)$ dependence lay far from case modeled above, such as case 8 in Table II and Fig. 3, the model could not be fit to any of the data sets because of the high $\eta$ values in the high-concentration region: compare Figs. 3 and 4. In a special case 5 in Table II, $C_1(T)$ was chosen to cross the $L$-$L$ coexistence line and the model calculations were carried out assuming that $C_1$ belongs to that line if the linear dependence lies beneath it. The model yielded a second maximum in the $J(T)$ dependence at $T$ where $C_1(T)$ crosses the $L$-$L$ coexistence line, which, apparently, is absent in the experiential data in Fig. 1.

This latter strong deviation indicates that the intermediate differs from the macroscopic dense liquid phase. On the other hand, the dependence of $C_1(T)$ was introduced assuming that the intermediate is similar to the dense liquid. These are not necessarily contradictory statements. We note that small clusters of a phase may have different properties than the macroscopic phase itself, as suggested by Gibbs.[46–48] Thus, it is likely that if the intermediate clusters are allowed to grow, they will become macroscopic dense liquid droplets and the suggestion that the intermediate is similar to the dense liquid phase is still valid.[30]

### C. The effects of glycerol and polyethylene glycol on the nucleation rate

The above model provides a framework for the understanding of the experimentally observed effects of two additives, glycerol and polyethylene glycol 8000 (PEG) on the nucleation kinetics:[22] it allows discrimination between the effects of the additives on the solution thermodynamics from those on the nucleation kinetics. The experimental results show that glycerol shifts the temperatures of solution-crystal equilibrium and of $L$-$L$ coexistence and delays the nucleation rate. The low amount of PEG used, 0.2%, does not affect the phase diagram, however, it leads to significant acceleration of the nucleation rate.[22]

The experimental results in the presence of glycerol are compared to the model predictions with three sets of parameters. The dashed line in Fig. 6(a) shows that if the effects of glycerol were limited to the effects on solution thermodynamics, the nucleation rate would have been suppressed very significantly. However, the increase of $k_2$ necessary to fit the model predictions to the experimental data, see Table I, suggest that glycerol significantly accelerates the kinetics of growth of the ordered clusters within the dense liquid intermediate. Furthermore, the scaling factor for $E_2$ was reduced by 33% indicating that glycerol leads to lower barriers for the nucleation of the ordered phase.

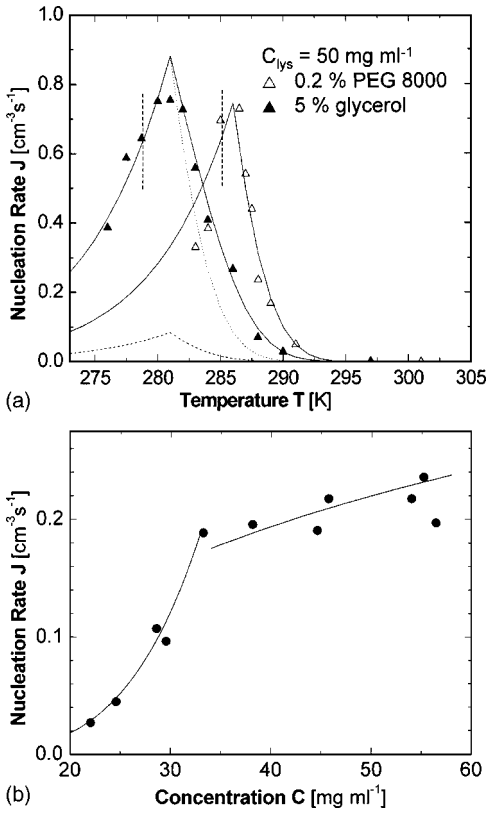


FIG. 6. Additional evaluations of the rate of homogeneous nucleation $J$ of lysozyme crystals. (a) The dependence on temperature $T$ at $C=50$ mg ml$^{-1}$ and in the presence of 5% glycerol or 0.2% polyethylene glycol 8000 as indicated in the plot. Symbols represent experimental results from Ref. 22. Lines are results of the model. For glycerol: dashed line is for $T_e$ and $T_{sp}$ are as in case 3 in Table I, all other parameters are as in case 1 in Table I; dotted line is for $T_e$ and $T_{sp}$ are as in case 3 in Table I, $k_2=1.862$, all other parameters are as in case 1 in Table I; solid line is for all parameters as in case 3 in Table I. For PEG: solid line is for case 4 in Table I. The temperatures of $L$-$L$ separation are 278 K in the presence of glycerol and 285 K in the presence of PEG, and are marked with vertical dashed lines. (b) The dependence on protein concentration at $T=285.7$ K. Symbols represent experimental results from Ref. 23. Lines are results of the model.

The effects of PEG are well modeled by a higher $k_2$ again suggesting an acceleration of the growth rate of the critical clusters.

We refrain from offering a molecular-level interpretation of the decrease in the barrier for nucleation of ordered clusters within the dense liquid intermediate in the presence of glycerol. However, the accelerated kinetics of growth of the clusters in the presence of both glycerol and PEG is very interesting. We draw an analogy to the ~ ten-fold acceleration of the kinetics of step growth on insulin crystals in the presence of acetone.[49] This acceleration was attributed to the destruction of the water shells around the insulin molecules in solution by acetone evidenced independently in Ref. 50. The link between the two is the finding that the kinetics of growth of ordered solid phases from solution is limited by the necessity to push structured waters out of the interstice between the incoming molecule and those already in the crystal.[37] Glycerol and PEG are similar to acetone in that all three are amphiphylic molecules. Thus, we conclude that the only effect of PEG and the strongest effect of glycerol on the nucleation kinetics occur via the destruction of the water shells of the molecules in the dense liquid precursor.

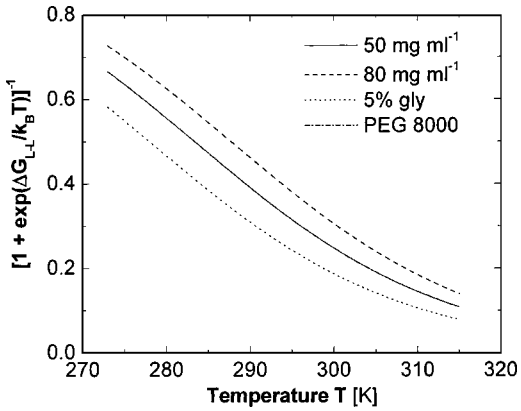


FIG. 7. Evaluation of the contribution of the intermediate to the increase in the nucleation rate at lower temperatures as for the four main model cases as marked in the plot.

### D. The dependence of the nucleation rate on protein concentration at constant temperature

The data on $J(C)$ at $T=285.7$ K in Fig. 6(b) is split into two regions. At $C<33$ mg ml$^{-1}$, $J$ increases quasiexponentially and the nucleus size $n^*=4$ molecules. At $C>33$ mg ml$^{-1}$, $J$ is a weak function of $C$ and $n^*=1$. In analogy to the spinodal $T_{sp}$ above, we introduce $C_{sp}=33$ mg ml$^{-1}$. To model the dependence $J(C)$, we use the values of $\Delta G$, $C_1$, and $\eta$ at this temperature. We use $U_1/U_0=50/C$ and $E_2=n^*\Delta\mu/2$, with $\Delta\mu=k_BT\ln(C/C_e)$ at $C<C_{sp}$ and $E_2=0$ at $C>C_{sp}$. A good fit of the model results to the experimental data required a slight increase of $k_2$ from 0.266, see Table I, to 0.2793 at $C<C_{sp}$. Because of the abrupt change of $E_2$ at $C_{sp}$, $k_2$ is needed to be changed to $1.596\times10^{-6}$ for $C>C_{sp}$.

It seems reasonable to assign the necessity of the increase of $k_2$ at $C<C_{sp}$ to the accumulation of inaccuracies in the values of $\Delta G$, $C_1$, and $\eta$. Furthermore, the change of $k_2$ at $C_{sp}$ is due to the jump in the size of the nucleus, reflected in the $J(C)$ data. Such jumps are also a part of the $J(T)$ data in Fig. 1; however, the driving force for the $n^*$ jumps is the decreasing $\Delta\mu$, which is a more sensitive function of $T$ than $C$. The density of $J(T)$ data points is insufficient to reveal the interruptions in the $n^*(T)$ dependence, and the use of a smooth $E_2(T)$ is appropriate. This smooth $E_2(T)$ allowed the use of a single value of the adjustable parameters $k_2$ even below $T_{sp}$. Thus, the good fit of the model results of the data in Fig. 6(b) represent another success of the model.

### E. The contribution of the dense liquid intermediate to faster nucleation rates at lower temperatures

As a rough estimate of this contribution, we evaluated the denominator in the nucleation rate law Eq. (4) in Fig. 7; this denominator would not be present in a single step nucleation. Due to our arbitrary choice of $U_0$ and $U_1$, the absolute values of this expression are meaningless. However, we see that as $T$ is lowered, the denominator contributes factors of $\sim5-7$ in the response of $J$ to $T$. This contribution enhances at $T>T_{sp}$ the effect of lower $E_2$. At $T<T_{sp}$, the acceleration due to lower $\Delta G$ is overwhelmed by the deceleration due to higher viscosity within the intermediate.

## IV. SUMMARY AND CONCLUSIONS

We have developed a simple phenomenological model describing the kinetics of nucleation of protein ordered solid phases from solution via a metastable intermediate. The model allows accurate predictions of the complex dependencies of the nucleation rate of crystals of the protein lysozyme using a single adjustable kinetic parameter and a reasonable guess of one materials parameter. The good correspondence between the model results and the experimental data supports the following features of the nucleation mechanism.

(i) The intermediate is similar to the dense liquid existing in lysozyme solutions.
(ii) The intermediate is metastable with respect to the dilute solution, i.e., a barrier for its decay exists and it is not a simple density fluctuation.
(iii) The rate-determining step in the nucleation mechanism is the formation of an ordered cluster within the dense liquid intermediate.
(iv) The viscosity within the metastable dense liquid droplet is a crucial parameter in the kinetics of nucleation of ordered solid phases.

The model allows discrimination between the thermodynamic and kinetic effects of additives that slow down or accelerate nucleation. It suggests that the amphiphylic character of the two tested additives accelerate the kinetics of growth of the near-critical ordered clusters in the dense liquid droplets by destroying the water structures around the protein molecules.

Future theoretical work should address the case of two step nucleation, in which the second step, ordering, is not the rate limiting one. Other potential avenues include understanding of the means to enhance the ordering employed by experimentalists: external fields, shear flow, etc. Experimentalists should provide data on the behavior of protein solution viscosity at high solution concentrations, as well as, of course, data on the nucleation rates with other protein and nonprotein materials.

## ACKNOWLEDGMENTS

The authors thank B. Widom and M. E. Fisher for valuable comments. This work was supported by the Office of Biological and Physical Research, NASA (Grant No. NAG8-1854 to P.G.V.), the Camille and Henry Dreyfus New Faculty Award Program (Grant No. NF-00-056 to A.B.K.), the Welch Foundation (Grant No. C-1559 to A.B.K.), and the Alfred P. Sloan Foundation (Grant No. BR-4418 to A.B.K.).

[1]S. K. Burley, S. C. Almo, J. B. Bonanno *et al.*, Nat. Genet. **23**, 151 (1999).
[2]K. Wütrich, Nat. Struct. Biol. **5**, 492 (1998).
[3]*Macromolecular Crystallography, Part C*, edited by C. W. Carter, Jr., and R. M. Sweet (Academic, San Diego, 2003), Vol. 368.
[4]W. A. Eaton and J. Hofrichter, in *Advances in Protein Chemistry*, edited by C. B. Anfinsen, J. T. Edsal, F. M. Richards, and D. S. Eisenberg (Academic, San Diego, 1990), Vol. 40, p. 63.
[5]J. W. Kelly, Nat. Struct. Biol. **7**, 824 (2000).
[6]M. Bucciantini, E. Giannoni, F. Chiti *et al.*, Nature (London) **416**, 507 (2002).
[7]J. Brange, *Galenics of Insulin* (Springer, Berlin, 1987).
[8]P. Reichert, C. McNemar, N. Nagabhushan, T. L. Nagabhushan, S. Tindal, and A. Hruza, U.S. Patent Patent No. 5,441,734 (1995).
[9]A. P. Gast, C. K. Hall, and W. R. Russel, Faraday Discuss. Chem. Soc. **76**, 189 (1983).
[10]D. F. Rosenbaum, A. Kulkarni, S. Ramakrishnan, and C. F. Zukoski, J. Chem. Phys. **111**, 9882 (1999).
[11]D. F. Rosenbaum, P. C. Zamora, and C. F. Zukoski, Phys. Rev. Lett. **76**, 150 (1996).
[12]D. Leckband and J. Israelachvili, Q. Rev. Biophys. **34**, 105 (2001).
[13]D. N. Petsev and P. G. Vekilov, Phys. Rev. Lett. **84**, 1339 (2000).
[14]P. G. Vekilov and A. A. Chernov, in *Solid State Physics*, edited by H. Ehrenreich and F. Spaepen (Academic, New York, 2002), Vol. 57, p. 1.
[15]N. Asherie, A. Lomakin, and G. B. Benedek, Phys. Rev. Lett. **77**, 4832 (1996).
[16]M. Hagen, E. Meijer, G. Mooij, D. Frenkel, and H. Lekkerkerker, Nature (London) **365**, 425 (1993).
[17]J. A. Thomson, P. Schurtenberger, G. M. Thurston, and G. B. Benedek, Proc. Natl. Acad. Sci. U.S.A. **84**, 7079 (1987).
[18]C. Liu, N. Asherie, A. Lomakin, J. Pande, O. Ogun, and G. B. Benedek, Proc. Natl. Acad. Sci. U.S.A. **93**, 377 (1996).
[19]M. Muschol and F. Rosenberger, J. Chem. Phys. **107**, 1953 (1997).
[20]O. Galkin, K. Chen, R. L. Nagel, R. E. Hirsch, and P. G. Vekilov, Proc. Natl. Acad. Sci. U.S.A. **99**, 8479 (2002).
[21]Q. Chen, P. G. Vekilov, R. L. Nagel, and R. E. Hirsch, Biophys. J. **86**, 1702 (2004).
[22]O. Galkin and P. G. Vekilov, Proc. Natl. Acad. Sci. U.S.A. **97**, 6277 (2000).
[23]O. Galkin and P. G. Vekilov, J. Am. Chem. Soc. **122**, 156 (2000).
[24]J. D. van der Waals, *Nobel Lectures, Physics 1901-1921* (Elsevier, Amsterdam, 1979).
[25]J. W. Cahn and J. E. Hilliard, J. Chem. Phys. **28**, 258 (1958).
[26]K. Binder and P. Fratzl, in *Phase Transformation in Materials*, edited by G. Kostorz (Wiley, New York, 2001).
[27]D. Kashchiev, *Nucleation. Basic Theory with Applications* (Butterworth, Oxford, 1999).
[28]P. R. ten Wolde and D. Frenkel, Science **277**, 1975 (1997).
[29]V. Talanquer and D. W. Oxtoby, J. Chem. Phys. **109**, 223 (1998).
[30]P. G. Vekilov, Cryst. Growth Des. **4**, 671 (2004).
[31]V. J. Anderson and H. N. W. Lekkerkerker, Nature (London) **416**, 811 (2002).
[32]G. Nicolis and C. Nicolis, Physica A **323**, 139 (2003).
[33]A. Lomakin, N. Asherie, and G. B. Benedek, Proc. Natl. Acad. Sci. U.S.A. **100**, 10254 (2003).
[34]N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, New York, 1992), Chapt. XII.
[35]M. Shah, O. Galkin, and P. G. Vekilov, J. Chem. Phys. (in press).
[36]O. Galkin and P. G. Vekilov, J. Mol. Biol. **336**, 43 (2004).
[37]D. N. Petsev, X. Wu, O. Galkin, and P. G. Vekilov, J. Phys. Chem. B **107**, 3921 (2003).
[38]O. Gliko and L. Filobelo (unpublished).
[39]I. Gutzow and J. Schmeltzer, *The Vitreous State* (Springer, Berlin, 1995).
[40]R. B. Bird, W. E. Stewarr, and E. N. Lightfoot, *Transport Phenomena* (Wiley, New York, 1960).
[41]C. W. Macosco, *Rheology Principles, Measuremnts and Applications* (VCH, New York, 1994).
[42]W. J. Fredericks, M. C. Hammonds, S. B. Howard, and F. Rosenberger, J. Cryst. Growth **141**, 183 (1994).
[43]D. Kashchiev, in *Science and Technology of Crystal Growth*, edited by J. P. v. d. Eerden and O. S. L. Bruinsma (Kluwer Academic, Dordrecht, 1995), p. 53.
[44]B. Mutaftschiev, in *Handbook of crystal growth*, edited by D. T. J. Hurle (Elsevier, Amsterdam, 1993), Vol. I, p. 189.
[45]P. G. Debenedetti, *Metastable Liquids* (Princeton University Press, Princeton, 1996).
[46]D. Kashchiev, J. Chem. Phys. **118**, 1837 (2003).
[47]J. W. Gibbs, Trans. Conn. Acad. Arts Sci. **3**, 108 (1876).
[48]J. W. Gibbs, Trans. Conn. Acad. Arts Sci. **16**, 343 (1878).
[49]I. Reviakine, D. K. Georgiou, and P. G. Vekilov, J. Am. Chem. Soc. **125**, 11684 (2003).
[50]L. Bergeron, L. Filobelo, O. Galkin, and P. G. Vekilov, Biophys. J. **85**, 3935 (2003).