

# Size Evolution and Composition of the Intermediate Phase During Nonclassical Protein Crystal Growth from Solution

D. A. Barlow <sup>† \*</sup>, Jan Gregus <sup>‡</sup>

<sup>†</sup>Alderman Barlow Labs, PO Box 1394 Trenton, Fla. 32693, USA

<sup>‡</sup>Department of Science and Mathematics, Abraham Baldwin Agricultural College, Tifton, Ga. 31793, USA

## Abstract

We propose here that the intermediate nucleation phase identified in a certain case of protein crystal growth actually consists of two distinct parts; a low density and higher density phase. A theory for crystal growth is utilized to study the formation and growth of each phase. Within the framework of this theory the low density phase is shown to obey a fourth order kinetic law while the high density phase is zeroth order. The combination of these two phases is shown to be a good match for X-ray diffraction data which is indicative of its presence. The crystal growth rate is then given in terms of the kinetic behavior of the intermediate nucleation phase. From this, the crystal radius is estimated and shown to compare favorably with reported size data. A method is proposed for determining the conditions that lead to protein crystals of largest possible size.

**Keywords:** Two-step nucleation; proteins; crystal size; kinetics; theories of crystal growth

---

\*doug.barlow@aldermanbarlow.com

# 1 Introduction

For over 20 years now groups have been reporting that for protein crystal growth from solution there can develop in certain cases an intermediate nucleation phase which perhaps acts as a pre-catalytic for the eventual growth of crystals [1–5]. The process has been identified in certain inorganic growth systems as well [6, 7]. In a recent report on the kinetics of protein crystal nucleation Barlow and Gregus proposed that in one particular case the solute rich associates of the intermediate phase seemed to form around seed nuclei that are already present at or near the instant in time when the supersaturated solution is prepared [8]. Nanev and co-workers had previously suggested that initial seed particles, of biological origin, must be present in all supersaturated protein solutions [9]. It was further hypothesized in Ref. [8] that these early stage associates do not go on to form crystals themselves but are consumed by crystals growing in the solution formed upon homogeneous nuclei. There is significant experimental evidence for solute rich associates of an intermediate phase, during lysozyme crystal growth, being consumed by nearby growing crystals [3, 10, 11]

Reported X-ray diffraction data [2], for a case of isothermal  $\beta$ -lactoglobulin (BLG) crystal growth, indicating the presence of the intermediate phase, leads us to propose here that the intermediate phase is actually composed of two distinct parts, a low density and high density phase. This hypothesis is consistent with a recent experimental report for lysozyme nucleation and crystal growth where two distinct types of amorphous protein particles were seen to facilitate nucleation [12]. In this report a standard theory for crystal nucleation and growth is employed to study the kinetics of the solute rich intermediate associates. The nucleation rate of the intermediate nuclei is taken to be given by an expression determined in an earlier report. Using this model for the intermediate phase nucleation rate, along with equations of continuity and mass balance, expressions for the associate linear growth rate and distribution of associate sizes are derived. Each of these phases appears to have a distinct kinetic behavior, or reaction order. Using the derived kinetic expressions the time dependent concentration of each associate type can be computed. The sum of these two concentrations is shown to be a good match for reported X-ray diffraction data indicative of its presence.

To further verify the model used for the concentration of intermediate phase we utilize it as part of an expression for the crystal growth rate where the crystal growth rate is taken to be directly proportional to the supersaturation not involved in the transient intermediate phase. The resulting expression for the crystal radius as a function of time compares favorably with size data reported for the BLG growth run mentioned above. Further, this result points to there being a maximum crystal size in the equilibrium distribution which is directly proportional to the cube root of the initial supersaturation.

It is important to note that in this report the intermediate nucleation phase should not be confused with the intermediate stage of crystal growth, both of which will be referred to. In the next two sections we develop the theoretical models required for our description of the intermediate phase and crystal size data reported for the BLG crystallization case mentioned above.

## 2 Zeroth Order Theory

Consider a size distribution function for spherical associates  $f$  which is a function of the associate radius  $r$  and time  $t$ . Assuming there is no growth rate dispersion in the distribution we let the linear growth rate of the associate be  $G = dr/dt$ . We employ the well-known method of population balance and mass conservation, variants of which have often been used to study intermediate stage crystal growth and ripening [13–17]. Population balance leads to

$$\frac{\partial f}{\partial t} + \frac{\partial(Gf)}{\partial r} = 0 . \quad (1)$$

The associate propagation mechanism suggested in Ref. [8] is used again here. There is an initial seed density of  $n_1$  upon which associates form. These seed nuclei then decay away as the associates are consumed by crystals growing on homogeneous nuclei and/or the associate acts as a heterogeneous seed for crystal growth. Either way, at this point the associate's seed nuclei is considered transferred to the crystal and thus consumed. So the nucleation rate for associate seeds is negative and is denoted by  $J$  where  $J = -Gf(t, 0)$  since  $G > 0$  and  $f > 0$ . We let the time dependent normalized supersaturation be denoted by  $s(t)$ . A common model for the growth rate is then  $G = as$  where  $a$  is a constant that depends upon the interface transfer coefficient. However, here we take a more general view and let  $G = G(t)$  that is, the associate growth rate is some yet to be determined function of time.

We assume that all associates are born at  $t = 0$  and that all will remain the same size throughout the run so that there will be no size dispersion in the distribution. In this case it will be that  $df/dt < 0$ , i.e. the total number of associates will decrease with time and with the number of particles decreasing as they increase in size we will have that  $df/dr < 0$ . By inspection of Eq. (1), which is used for cases where  $J > 0$ , we see that it must be adjusted by a sign so that population balance for our system becomes

$$\frac{\partial f}{\partial t} - \frac{\partial(Gf)}{\partial r} = 0 . \quad (2)$$

In addition to Eq. (2), a second independent equation describing our system can be written by assuming conservation of mass,

$$\frac{ds}{dt} = -K \int_0^\infty r^2 G(t) f(t, r) dr , \quad (3)$$

where  $K$  is a constant which includes a shape factor.

An essential part of the proposed propagation mechanism is for the associate seed nuclei density  $n$  to be given as [8]

$$n(t) = \frac{n_o}{\frac{n_o - n_1}{n_1} + e^{kt}} , \quad (4)$$

where  $n_1$  is the initial seed nuclei density in the solution and  $n_o$  is the final nuclei density. From Eq. (4) it can be seen, as mentioned above, that the initial seed nuclei density will decay to zero as the intermediate nucleation phases passes. Therefore, in this model, the final nuclei density,  $n_o$ , is due to homogeneous nucleation.

The intermediate phase nucleation rate is  $J = dn/dt$ , which leads to

$$J = -\frac{n_o k e^{kt}}{\left(\frac{n_o - n_1}{n_1} + e^{kt}\right)^2} . \quad (5)$$

Given the simple nature of the mono-disperse size population studied here we propose that the distribution function be separable i.e.,

$$f = T(t)R(r) . \quad (6)$$

While non-separable distributions have been studied extensively [14, 18–20], separable distributions have been shown to lead to a reasonable theoretical description of the supersaturation decay during lysozyme and insulin bulk crystal growth [16, 17] and have been shown to describe the crystal size distribution during bulk insulin crystal growth from solution [21].

The associate nucleation rate is now given as,

$$J = -G(t)T(t)R(0) , \quad (7)$$

Solving this for  $T$  gives,

$$T(t) = \frac{-J}{G(t)R_o} , \quad (8)$$

where we have let  $R(0) = R_o$ . Now, Eq. (2) will be used to determine  $G(t)$  and  $R(r)$ . With these functions known, the distribution function  $f$  is assembled and the volume of the intermediate phase as a function of time can be studied.

Using Eq. (6) and Eq. (8) in Eq. (2) gives,

$$-\frac{R}{R_o} \frac{d(J/G(t))}{dt} + \frac{J}{R_o} \frac{dR}{dr} = 0 . \quad (9)$$

Expanding the derivative and rearranging leads to

$$\frac{1}{G^2} \frac{dG}{dt} - \frac{1}{GJ} \frac{dJ}{dt} = -\frac{1}{R} \frac{dR}{dr} . \quad (10)$$

Letting the separation constant be  $\lambda$ , where  $\lambda > 0$  leads to the following solution for  $R(r)$ :

$$R(r) = R_o e^{-\lambda r} . \quad (11)$$

Now, the time dependent part of Eq. (10) is considered. Rearranging leads to the Bernoulli equation,

$$\frac{dG}{dt} - \frac{1}{J} \frac{dJ}{dt} G - \lambda G^2 = 0 . \quad (12)$$

Eq.(12) is converted into linear form by using the substitution  $z = 1/G$  so that it becomes,

$$\frac{dz}{dt} + \frac{1}{J} \frac{dJ}{dt} z = -\lambda . \quad (13)$$

Using Eq. (5) an integrating factor for Eq. (13) is found to be

$$\exp \left[ \int \frac{1}{J} \frac{dJ}{dt} dt \right] = -\frac{J}{n_o k} . \quad (14)$$

Since  $G = 1/z$  we have then that

$$G = \frac{-J}{n_o k \left[ -\lambda \int \frac{-J(t')}{n_o k} dt' + C_1 \right]} . \quad (15)$$

Here  $C_1$  is a constant. Now  $J = dn/dt$  so that Eq. (15) is written as

$$G = \frac{-J}{\lambda \int dn + C_1} , \quad (16)$$

and therefore,

$$G = \frac{-J}{\lambda n(t) + C_1} . \quad (17)$$

Letting  $C_1 = 0$  and using Eqs. (4) and (5)  $G$  becomes,

$$G(t) = \frac{(k/\lambda)e^{kt}}{\beta + e^{kt}} . \quad (18)$$

Here we have set  $\beta = (n_o - n_1)/n_1$ .

Now,  $T(t)$  can be found from Eq. (7) and the distribution function written. However, we first use Eq. (3) to determine  $s(t)$  and in doing so arrive at a relation for  $\lambda$  in terms of other parameters. Additionally, the resulting expression for  $s(t)$  will be useful for defining an undetermined parameter that will appear in the derivation of the following section. It should be noted that the normalized supersaturation determined here and in the following section no longer refer to the bulk supersaturation but rather to some subset of the bulk supersaturation which drives each intermediate nucleation phase.

Using Eqs. (6), (8) and (11) in Eq. (3) leads to

$$\frac{ds}{dt} = KJ \int_0^\infty r^2 e^{-\lambda r} dr . \quad (19)$$

Computing the integral in Eq. (19) yields

$$\frac{ds}{dt} = \frac{2KJ}{\lambda^3} . \quad (20)$$

Using Eq. (5) in eq. (20) and separating leads to

$$\int ds = -\frac{2K}{\lambda^3} \int \frac{kn_o e^{kt}}{(\beta + e^{kt})^2} dt . \quad (21)$$

Evaluating the integrals in Eq. (21) leads to the following for  $s$ :

$$s(t) = \frac{2Kn_o}{\lambda^3} \left( \frac{1}{\beta + e^{kt}} \right) + C_2 , \quad (22)$$

where  $C_2$  is a constant of integration. The initial condition for  $s$  is that  $s(0) = 1$ . The final condition is that  $s \rightarrow 0$  as  $t \rightarrow \infty$ . These conditions are achieved when  $C_2 = 0$  and

$$s(t) = \frac{2Kn_o}{\lambda^3} \left( \frac{1}{\beta + e^{kt}} \right) \quad \text{for } \lambda = \sqrt[3]{\frac{2Kn_o}{\beta + 1}}. \quad (23)$$

Now, using Eqs. (6), (8), (11) and (18) the distribution function can be written as

$$f(t, r) = \left( \frac{\lambda n_o}{\beta + e^{kt}} \right) e^{-\lambda r} \quad \text{for } r = r_m(t). \quad (24)$$

Here the distribution function is restricted to radii of  $r = r_m(t)$  where  $r_m(t)$  is the radius at time  $t$  since in this case all associates were born at  $t = 0$ .

An expression for  $r_m(t)$  can now be determined. By definition the linear growth rate is  $G = dr/dt$ . Using this with Eq. (18) we have

$$\frac{dr_m}{dt} = \frac{(k/\lambda)e^{kt}}{\beta + e^{kt}}. \quad (25)$$

Separating and integrating with the initial condition  $r(0) = 0$  gives

$$r_m(t) = \frac{1}{\lambda} \ln \left( \frac{\beta + e^{kt}}{\beta + 1} \right). \quad (26)$$

Now, assuming that the associate is spherical in shape, the volume of associate phase per unit volume of solution is given as

$$V(t) = \frac{4}{3}\pi[r_m(t)]^3 n(t). \quad (27)$$

With Eqs. (4) and (26) this becomes

$$V(t) = \frac{4\pi}{3\lambda^3} \left[ \ln \left( \frac{\beta + e^{kt}}{\beta + 1} \right) \right]^3 \left[ \frac{n_o}{(\beta + e^{kt})} \right] \quad (28)$$

### 3 General Order Theory

A more general version of the derivation carried out in the previous section is now considered. Here the growth rate will depend upon the associate radius as  $r^p$ . Here  $p$  could be any positive number but is typically an integer. Therefore, the growth rate  $G$  is modeled as

$$G = \frac{g(t)}{\alpha r^p}. \quad (29)$$

Here  $\alpha$  is a constant. As in the previous section we have that  $f > 0$ ,  $G > 0$  and  $J < 0$ . Expressions for growth rates in the form of Eq. (29) have been discovered during theoretical

studies of colloidal aggregation [22] and Ostwald ripening [23]. We assume a separable distribution function as in Eq. (6) and that the nucleation rate  $J$  is modeled as

$$J = -\frac{g(t)}{\alpha r^p} T(t) R_o . \quad (30)$$

Now, using this in the separable distribution function and inserting into Eq. (2) leads to

$$-\alpha r^p R \frac{d(J/g)}{dt} + J \frac{dR}{dr} = 0 \quad (31)$$

Rearranging and separating yields,

$$-\frac{1}{J} \frac{d(J/g)}{dt} = -\frac{1}{\alpha r^p R} \frac{dR}{dr} . \quad (32)$$

The time dependent side of Eq. (32) leads to the same result as found in the previous section. Letting the separation constant be  $\lambda$ , the right side can be rearranged to give,

$$\frac{1}{R} dR = -\lambda \alpha r^p dr . \quad (33)$$

Integrating both sides applying initial conditions yields,

$$R = R_o \exp \left[ -\frac{\lambda \alpha r^{(p+1)}}{p+1} \right] . \quad (34)$$

Using  $T(t)$  from Eq. (30), along with Eq. (34), the entire distribution function can be written for the general order case.

$$f(t, r) = \left( \frac{\lambda n_o}{\beta + e^{kt}} \right) \alpha r^p \exp \left[ -\frac{\lambda \alpha r^{(p+1)}}{p+1} \right] \quad \text{for } r = r_m(t) . \quad (35)$$

Using Eqs. (29) and (35) in Eq.(3), the normalized supersaturation can be determined for the generalized case.

$$\frac{ds}{dt} = KJ \int_0^\infty r^2 \exp \left[ -\frac{\lambda \alpha r^{(p+1)}}{p+1} \right] dr = KJ \frac{\Gamma \left( \frac{3}{p+1} \right)}{p+1} \left( \frac{p+1}{\alpha \lambda} \right)^{3/(p+1)} , \quad (36)$$

where  $\Gamma$  denotes the gamma function. Using Eq. (5) in Eq. (36), separating, integrating and applying the final condition that  $s \rightarrow 0$  as  $t \rightarrow \infty$  leads to

$$s(t) = Kn_o \frac{\Gamma \left( \frac{3}{p+1} \right)}{p+1} \left( \frac{p+1}{\alpha \lambda} \right)^{3/(p+1)} \left( \frac{1}{\beta + e^{kt}} \right) . \quad (37)$$

In Eqs. (23) and (37) it must be that  $s(0) = 1$ . This implies that

$$\frac{2}{\lambda^3} = \frac{\Gamma \left( \frac{3}{p+1} \right)}{p+1} \left( \frac{p+1}{\alpha \lambda} \right)^{3/(p+1)} , \quad (38)$$

thus leading to,

$$\alpha = \lambda^p(p+1) \left[ \frac{\Gamma\left(\frac{3}{p+1}\right)}{2(p+1)} \right]^{(p+1)/3}. \quad (39)$$

Now an expression for  $r_m(t)$  in the general order case is sought. Using Eq. (29) we have for the linear growth rate,

$$\frac{dr_m}{dt} = \frac{g(t)}{\alpha r_m^p}. \quad (40)$$

Using Eq. (18) for  $g(t)$  in Eq. (40) re-arranging and taking the integral of both sides leads to

$$\int_0^{r_m} \alpha (r'_m)^p dr'_m = \int_0^t \frac{(k/\lambda)e^{kt'}}{\beta + e^{kt'}} dt'. \quad (41)$$

Integrating and solving Eq. (41) for  $r_m$  yields,

$$r_m(t) = \left[ \frac{p+1}{\lambda\alpha} \ln\left(\frac{\beta + e^{kt}}{\beta + 1}\right) \right]^{1/(p+1)}. \quad (42)$$

The form of Eq. (42), for large times, is comparable to similar expressions derived by Sugimoto in his study of Ostwald ripening [23, 24]. That is for large  $t$ ,  $r_m \propto t^{1/(p+1)}$ , where Sugimoto refers to  $p$  as the order of the reaction, a terminology that is also used in this report.

Following the reasoning used in the previous section we use Eqs. (4) and (42) to define the associate volume per unit volume of solution as

$$V(t) = \frac{4}{3}\pi \left[ \frac{p+1}{\lambda\alpha} \ln\left(\frac{\beta + e^{kt}}{\beta + 1}\right) \right]^{3/(p+1)} \left( \frac{n_o}{\beta + e^{kt}} \right). \quad (43)$$

As expected, Eq. (43) reduces to eq. (28) when  $p = 0$ .

Additional insight into the effect of reaction order on the kinetic behavior of the intermediate phase can be gained by using Eq. (39) in Eq. (43) and generating a surface plot of  $V$  vs.  $p$  and  $t$ . In Fig 1. it can be seen that higher reaction orders lead to higher peak values of total associate volume which appear earlier in the crystal growth run. It will be demonstrated in the next section that the intermediate phase activity investigated here can be described as the sum of an early time, or fourth order reaction, phase and a time delayed zeroth order phase each with a distinct protein concentration.

## 4 Comparison with Experiment

Previously in Refs. [2] and [8] a model was proposed to describe the kinetic behavior of the intermediate phase in a reported case of BLG crystal growth. In both cases the resulting curve giving the concentration of the intermediate phase was always smooth and concave down, that is, the curve has one hump. However, the X-ray diffraction data indicative of its presence, reported in Ref. [2], clearly indicates otherwise. This data is given by the area of X-ray intensity peaks vs. scattering vector for two cases of protein crystal growth from

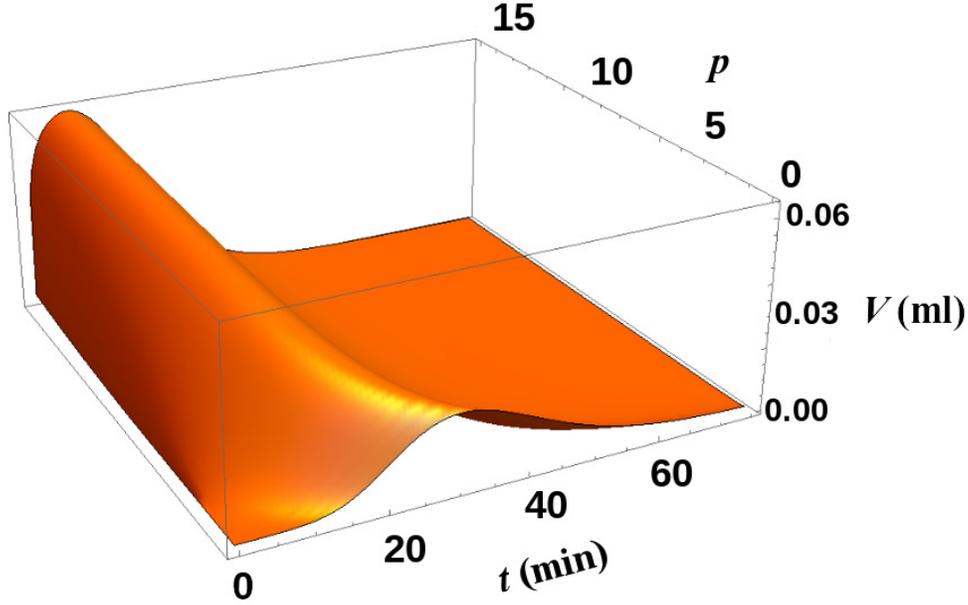


Figure 1: Surface plot of associate volume per ml of solution vs. time and reaction order, i.e. using Eq. (39) in Eq. (43). Rate constant was set to  $0.71 \text{ min}^{-1}$ .  $n_o$ ,  $n_1$  and  $\lambda$  are set to the values given in Table 1 and  $\beta = (n_o - n_1)/n_1$ .

solution and appears to have two humps. To address this discrepancy we propose that the observed intermediate phase is due to two distinct intermediate phases growing in the same associate: A low density solute rich phase and a higher density solute rich phase each having a distinct reaction order.

It is now shown how the expressions derived in the previous sections can be used to describe this phenomenon. We assume that the reported X-ray area data is proportional to the scattered intensity. From Glatter and Kratky [25] we find that the scattered intensity is directly proportional to a mass density squared times a volume squared. Therefore we consider the square of the mass of the intermediate phase per unit volume of solution and compare this with the reported X-ray data. For the low density phase mass per unit volume of solution is given by

$$M_I(t) = \rho_I V_I(t) , \quad (44)$$

and likewise for the higher density phase  $M_{II}(t)$ . Then the net mass per unit volume of the intermediate phase is

$$M_T = \rho_I V_I(t) + \rho_{II} V_{II}(t) . \quad (45)$$

$V_1$  and  $V_2$  are given by Eq. (43). We assume that the two phases have different rate constants which will be left as adjustable parameters so that  $k$  will be replaced by  $k_I$  in  $M_I$  and by  $k_{II}$  in  $M_{II}$ .  $\rho_I$  and  $\rho_{II}$  give the mass of protein with respect to associate volume for the two phases respectively; both are assumed to remain constant.

$M_T^2$  is fitted to X-ray diffraction data indicative of the presence of the intermediate phase taken from Ref. [2] and shown in Fig. 2. This curve fit is achieved when the low density phase is order  $p = 4$  and the high density phase is order  $p = 0$ . The rate constants

are  $k_I = 0.17 \text{ min}^{-1}$  and  $k_{II} = 0.126 \text{ min}^{-1}$ . The remaining parameter values used in this application of Eq. (45) are listed in Table 1 and further described in the following section. The estimated concentration of the low and high density intermediate phase for this growth case are also plotted individually in Fig. 3. Though Sauter et. al. [2] reported area data for two distinct cases (different initial supersaturations and salt concentrations) of BLG crystal growth where the intermediate phase is present, we study only one here as it was also reported with accompanying kinetic data for crystal density and size.

The previous kinetic models for an intermediate phase of uniform density, given in Refs. [2] and [8], were written in terms of one rate constant  $k$ . It is therefore the case here that the two distinct rate constants and thus different nucleation rates involved in the creation of  $M_I$  and  $M_{II}$  can be thought of as competing rate-limiting steps and that by the law of kinetic resistances the rate constants combine as [26]

$$\frac{1}{k} = \frac{1}{k_I} + \frac{1}{k_{II}}, \quad (46)$$

to yield the overall rate constant for intermediate phase nucleation. We speculate here that the zeroth order phase is a kinetically limited phase, that is, limited by the rate of incorporation at the associate surface, while the fourth order phase is diffusion limited.

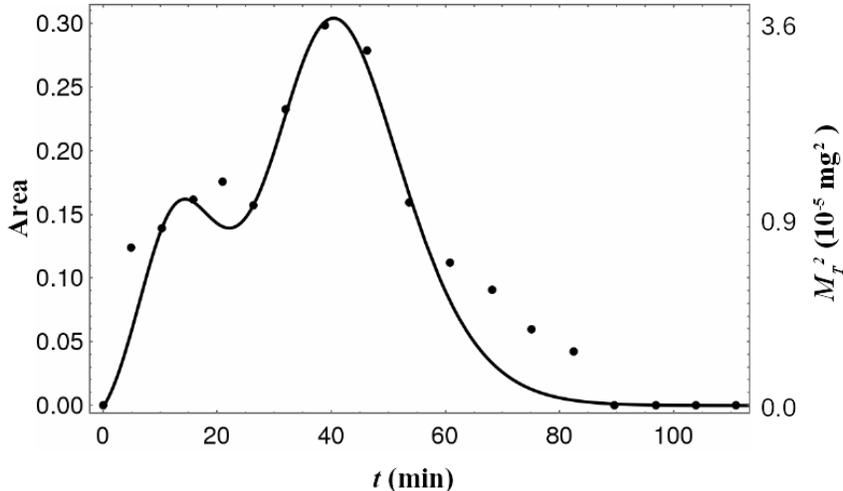


Figure 2: The square of intermediate phase mass,  $M_T$  from Eq. (45), in one ml of solution fitted to X-ray area data for the intermediate nucleation phase for protein crystals grown from a 20 mg/mL BLG solution with 15 mM  $\text{CdCl}_2$  taken from Ref. [2].  $T = 293 \text{ K}$ .

Crystal size data vs. time was reported in Ref. [2] for the crystallization case depicted in Fig. 2. Also, in this same report, kinetic data was given for the density of crystals in solution. It is clear from these two sets of data that the crystal growth rate does not have the same kinetic behavior as the nucleation rate the growth rate being significantly retarded at early times. It is possible that the presence of the intermediate phase affects the crystal growth rate somehow. We suggest here that the growth of the intermediate phase creates a diffusion limited situation for crystal growth and thus lowers the solute diffusion coefficient with respect to the crystal which in turn acts to lower the crystal growth rate. This notion

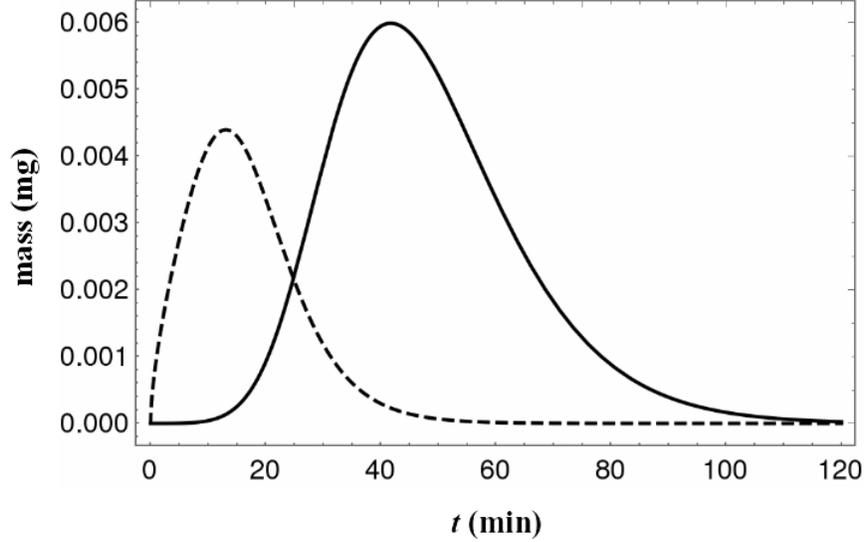


Figure 3: A figure showing each term on the right of Eq. (45) plotted separately showing the estimated mass in each of the intermediate nucleation phases per ml of solution for protein crystals grown from a 20 mg/mL BLG solution with 15 mM CdCl<sub>2</sub> taken from Ref. [2]. Dashed curve shows the lower density, fourth order phase, solid curve gives the higher density zeroth order phase.  $T = 293$  K.

is incorporated within a model for the linear crystal growth rate which is used to estimate the crystal radius vs. time and then compared with the reported data mentioned above.

To find an expression for the crystal radius,  $r_c$ , we invoke a commonly used model for the linear growth rate for crystals grown from solution where the rate is directly proportional to the supersaturation [9, 19, 27, 28]—a model typically valid for cases where there are uniform attachments to a surface which is rough on the molecular scale [29]. Specifically, a version of the model suggested by Laudise [30] is used here where

$$\frac{dr_c}{dt} = \frac{k_d}{\sigma} (c(t) - c_s). \quad (47)$$

Here,  $c(t)$  is solute concentration,  $c_s$  the equilibrium solubility,  $k_d$  the diffusion velocity constant and  $\sigma$  the width of the diffusion layer. Letting the initial solute concentration be given by  $c_o$  we define  $\Delta c = c_o - c_s$ . Now we write  $dr_c/dt$  as the sum of two terms involving the concentration of solute converted to crystal  $M_c$  and the solute within the intermediate phases  $M_I$  and  $M_{II}$ .

$$\frac{dr_c}{dt} = k_d \left[ \frac{1}{\sigma_c} (\Delta c - M_c(t)) - \frac{1}{\sigma_i} (M_I(t) + M_{II}(t)) \right], \quad (48)$$

where it is assumed that the crystal has the diffusion layer  $\sigma_c$  and the intermediate associate has diffusion layer  $\sigma_i$ . Eq. (48) will have the initial conditions that  $dr_c/dt = k_d \Delta c / \sigma_c$  and  $r_c = 0$  at  $t = 0$  and the final conditions of  $dr_c/dt = 0$  and  $\Delta c = M_c(t)$  at equilibrium. Using the model of Eq. (27) we define

$$M_c(t) = \rho \frac{4}{3} \pi r_c^3 n_h(t) , \quad (49)$$

where  $n_h$  is the homogeneous nucleation rate taken from Ref. [8]:

$$n_h(t) = \frac{n_o e^{kt} - n_o}{\frac{(n_o - n_1)}{n_1} + e^{kt}} . \quad (50)$$

Using Eq. (49) in Eq. (48) we arrive at Abel's differential equation of the first kind. An exact solution for this nonlinear differential equation has been reported [31]. However, as this solution is elaborate and transcendental, a numerical approximation for Eq. (48) is considered here for purposes of comparison with experimental size data.

The crystal size data reported in Ref. [2] will now be compared with the numerical solution of Eq. (48). In this report crystal size is given as length which we take to be twice the crystal radius  $r_c$ . Using expressions for  $M_I$  and  $M_{II}$ , used previously in the curve fit of Fig. 2 and letting  $k_d/\sigma_c$  and  $k_d/\sigma_i$  be adjustable parameters, the solution for Eq. (48) is fitted to the size data from experiment. This result is shown in Fig. 4.

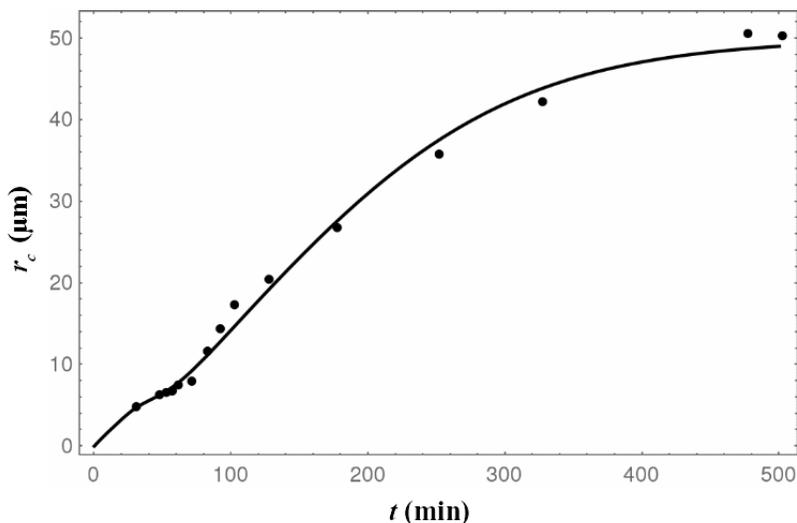


Figure 4: Radius vs. time data points for BLG crystals grown from a 20 mg/mL solution with 15 mM  $\text{CdCl}_2$  taken from Ref. [2]. Curve is from the numerical solution to Eq. (48) using the parameters listed in Table 1.  $T = 293$  K.

The invariable parameter values used in the implementation of Eqs. (48) and Eq. (45) were arrived at as follows. The parameter  $k$  was estimated recently in Ref. [8] for the growth run studied here.  $n_o$  and  $n_1$  are estimated using the dimensional scale on the video included with the supplemental information for Ref. [2]. A slide and thus image thickness of 300  $\mu\text{m}$ , as reported in Ref. [2], was assumed. The density of BLG in the crystalline state  $\rho$  is given the value reported by Heijna et al. for lysozyme [10]. To determine the initial supersaturation  $\Delta c$ , it is assumed that at equilibrium all of the excess solute has gone into crystal so that  $\Delta c \simeq \frac{4}{3} \rho \pi r_M^3 n_o$  where  $r_M$  is the maximum crystal radius at equilibrium.  $r_M$  is estimated from the kinetic size data reported in Ref. [2] to be 50  $\mu\text{m}$ .  $1/\lambda$  gives the

mean radius for the spherical associate in the zeroth order model and fulfills a similar role in all higher order cases. Vekilov [3], Heijna et al. [10] and Liu et al. [11] have reported for lysozyme crystallization from solution that regions of solute rich material initially form and yet do not go on to form crystals. They do grow to detectable size before disappearing. From optical micrographs in these reports we judge the median radius of solute rich associates to be around 10  $\mu\text{m}$ . We therefore set  $\lambda = 1000 \text{ cm}^{-1}$ .

The density for the intermediate phase has been estimated by Pan and co-workers for lysozyme [32]. They report a range of density, for the temperature at which the data in Fig. 2 was collected, of 0.1 g/ml to 0.4 g/ml. The fit in Fig 1 was obtained by letting the ratio of  $\rho_I$  to  $\rho_{II}$  be 4:11. We therefore let  $\rho_I = 0.1 \text{ g/ml}$  and  $\rho_{II} = 0.275 \text{ g/ml}$ . Values for all of the parameters used in the generation of the curves of Figures 1, 2, 3 and 4 are listed in Table 1.

Table 1: Values for parameters used in the generation of the curves in Figures 1, 2, 3 and 4.

$\frac{k_d}{\sigma_c} \left( \frac{\text{cm}^4}{\text{gs}} \right)$	$\frac{k_d}{\sigma_i} \left( \frac{\text{cm}^4}{\text{gs}} \right)$	$n_o \text{ (cm}^{-3}\text{)}$	$n_1 \text{ (cm}^{-3}\text{)}$	$\Delta c \text{ (mg/ml)}$	$\rho \text{ (g/cm}^3\text{)}$	$r_M \text{ (\mu m)}$
$11.5 \times 10^{-4}$	1.5	$3.87 \times 10^4$	$3.87 \times 10^3$	16.4	0.81	50.0
$k \text{ (min}^{-1}\text{)}$	$k_I \text{ (min}^{-1}\text{)}$	$k_{II} \text{ (min}^{-1}\text{)}$	$\rho_I \text{ (g/ml)}$	$\rho_{II} \text{ (g/ml)}$	$\lambda \text{ (cm}^{-1}\text{)}$	
0.071 Ref. [8]	0.17	0.126	0.1	0.275	1000	
0.072 Eq. (46)						
0.05 Ref. [2]						

Additionally, from the values used for  $k_d/\sigma_c$  and  $k_d/\sigma_i$  we find that  $\sigma_c \simeq 1304\sigma_i$ . That is, the diffusion width for the crystal is predicted to be about 1300 times wider than for the intermediate associate.

## 5 Maximum Crystal Size

Useful information concerning the maximum crystal size in the distribution at equilibrium can be obtained by considering Eq. (48) at large times when the second term on the right side is essentially zero. This leads to

$$\frac{dr_l}{dt} = \frac{k_d}{\sigma_c} (\Delta c - \rho \frac{4}{3} \pi r_l^3 n_o). \quad (51)$$

where from Eq. (50) the fact that  $n_h \rightarrow n_o$  as  $t \rightarrow \infty$  is used.  $r_l$  then denotes the crystal radius at large times. Eq. (51) can be separated and integrated leading to,

$$\frac{r_M^3 \sigma_c}{k_d \Delta c} \left\{ \frac{1}{r_M^2 \sqrt{3}} \arctan \left( \frac{2r_l + r_M}{r_M \sqrt{3}} \right) - \frac{1}{6r_M^2} \ln \left[ \frac{(r_M - r_l)^3}{r_M^2 - r_l^3} \right] - \frac{1}{r_M^2 \sqrt{3}} \arctan \left( \frac{1}{r_M \sqrt{3}} \right) \right\} = t, \quad (52)$$

where

$$r_M^3 = \frac{\Delta c}{\rho_3^{\frac{4}{3}} \pi n_o} . \quad (53)$$

Interestingly, Eq. (52) points to there being a theoretical maximum for the crystal radius,  $r_M$ . This is given by

$$r_M = \left( \frac{3\Delta c}{4\rho\pi n_o} \right)^{1/3} . \quad (54)$$

This result is consistent with no fewer than three other estimates for the maximum crystal size during protein crystal growth from solution reported previously in the literature. First, Nanev and co-workers gave for insulin crystal growth [33]

$$r_M = \left( \frac{\Omega\Delta c}{n_o} \right)^{1/3} , \quad (55)$$

where  $\Omega$  is a density defined in Ref. [33]. Barlow derived for insulin crystal growth that [17]

$$r_M = \frac{1}{\lambda} \cosh^{-1} \left( B\lambda^3 \sqrt[3]{\Delta c} + 1 \right) , \quad (56)$$

where the constant  $B$  is described in Ref. [17] and in this case  $1/\lambda$  refers to the mean size of the insulin crystals at equilibrium. If  $r_M$  is within 25% of the mean size in the equilibrium distribution then  $B\lambda^3 \sqrt[3]{\Delta c} < 1$  so that the above is approximately given by

$$r_M \sim B\lambda^2 (\Delta c)^{1/3} . \quad (57)$$

Recently, work was reported by Braun et al. [34] where BLG clusters in solution were studied using X-ray scattering and neutron backscattering methods. While suggesting that the native form of BLG in solution is that of the dimer they find that larger clusters form in BLG solutions where the cluster hydrodynamic radius is directly proportional to the cube root of the number of dimers present. In other words, they find that the radii of clusters increase in direct proportion to the cube root of the solute concentration.

Of course Eq. (54) does not necessarily imply that conditions can be brought about whereby the largest crystal size at equilibrium is unlimited. It is known that for most protein solutions there is an upper limit to the initial supersaturation whereby protein crystallization can occur; above this concentration aggregation takes place. This point is at the high concentration side of the so called *crystallization gap* in the temperature-concentration phase diagram [35]. More specifically, Eq. (54) predicts that the maximum crystal size will occur when the ratio  $\Delta c/n_o$  is a maximum. Here  $\Delta c$ , restricted to be within the crystallization gap, could in fact be anywhere within the gap where the maximum of  $\Delta c/n_o$  occurs. Therefore, the development of a diagram involving  $\Delta c$  and  $n_o$ , which denotes the crystallization gap, might be useful for identifying the true maximum possible crystal size.

## 6 Conclusion

In this report an expression for the nucleation rate of the non-classical intermediate phase for protein crystal growth from solution is used within a classic population-mass balance model

to study the kinetics of concentration of the intermediate phase associates for a case of BLG crystallization. This leads to an expression for the associate growth rate as a function of time. Modifying this growth rate by adding the factor of an inverse power law of the radius allows for a more general result in terms of the positive integer exponent of the radius, or the Sugimoto order of the reaction. With this result an expression for the over all concentration of the intermediate phase is written. These expressions are used to study a case of non-classical BLG protein crystal growth where the intermediate phase was confirmed by real-time X-ray diffraction data.

It was hypothesized here that in this case the intermediate phase is composed of two distinct parts: a low density phase and a higher density phase. The high density phase is fit by a zeroth order law while the low density phase by a fourth order law. We should note here, that given the fact that we were only able to analyze one set of data, there is no reason to believe that the Sugimoto reaction orders adopted here are universal. The sum of these two time dependent mass densities are shown to accurately describe the overall kinetic behavior of the intermediate phase as given in the experimental report.

The crystal growth rate was taken to be directly proportional to the supersaturated solute not involved in the intermediate phase nor yet taken up into the crystal. Using this model, along with the expression for kinetics of the intermediate phase concentration, we are able to estimate the crystal radius as a function of time and compare this with experimental size data for the crystal growth run mentioned above.

These results lead us to propose that the intermediate phase particles do not go on to form crystals themselves but are rather consumed by crystals growing from homogeneous or heterogeneous seeds. This proposal contrasts with the two-step mechanism given by Sauter et al. [2]. In their worked it was assumed that the presence of an intermediate phase would not sufficiently restrict the solute flow to a growing crystal so that the observed retardation of the crystal growth rate must be due to the maturing of some essential seed within the intermediate phase. We demonstrated here, via Eq. (48) and Fig. 4, that the presence of an intermediate phase can temporarily restrict solute flow to the crystals in such a way as to yield the kinetic size data reported in Ref. [2]. Our alternative proposal is consistent with recent experimental results reported for lysozyme crystallization by Yamazaki et al. [12]. Here they report the presence of two distinct precursor intermediate phases that possibly act as heterogeneous seeds but do not go on to form crystals themselves.

Finally, these results point to there being a maximum possible crystal size which is directly proportional to the cube root of the initial supersaturation and inversely proportional to the cube root of the final number density of homogeneous nuclei.

## References

- [1] ten Wolde, P. R.; Frenkel, D. Enhancement of Protein Crystal Nucleation by Critical Density Fluctuations, *Science*, **1997**, 277, (5334), 1975-1978.
- [2] Sauter, A.; Roosen-Runge, F.; Zhang, F.; Lotze, G.; Jacobs, R. M. J.; Schreiber, F. Real-Time Observation of Nonclassical Protein Crystallization Kinetics, *J. Am. Chem. Soc.*, **2015**, 137, 1485-1491.
- [3] Vekilov, P. G. Two-Step Mechanism for the Nucleation of Crystals from Solution, *J. Crystal Growth*, **2005**, 275, 65-76.
- [4] Savage, J. R.; Dinsmore, A. D. Experimental Evidence for Two-Step Nucleation in Colloidal Crystallization, *Phys. Rev. Lett.*, **2009**, 102, 198302.
- [5] Lutsko, J.; Nicolis, G. Theoretical Evidence for a Dense Fluid Precursor to Crystallization, *Phys. Rev. Lett.*, **2006**, 96, 046102.
- [6] Chakraborty, D.; Patey, G. N. Evidence that Crystal Nucleation in Aqueous NaCl Solution Occurs by the Two-Step Mechanism, *Chem. Phys. Lett.*, **2013**, 587, 25-29.
- [7] Matsukawa, Y.; Takeuchi, T.; Kakubo, Y.; Suzudo, T.; Watanabe, H.; Abe, H.; Toyama, T.; Nagai, Y. The Two-Step Nucleation of G-phase in Ferrite, *Acta Mater.*, **2016**, 116, 104-113.
- [8] Barlow, D. A.; Gregus, J. The Kinetics of Homogeneous and Two-Step Nucleation during Protein Crystal Growth from Solution, *Int. J. Chem. Kinet.*, **2019**, 51, 840-847.
- [9] Nanev, C. N.; Hodzhaoglu, F. V.; Dimitrov, I. L. Kinetics of Insulin Crystal Nucleation, Energy Barrier and Nucleus Size, *Crys. Growth Des.*, **2011**, 11, 196-202.
- [10] Heijna, M. C. R.; van Enckevort, W. J. P.; Vlieg, E. Crystal Growth in a Three-Phase System: Diffusion and Liquid-Liquid Phase Separation in Lysozyme Crystal Growth, *Phys. Rev. E*, **2007**, 76, 011604.
- [11] Liu, Y.; Wang, X.; Ching, C. B. Toward Further Understanding of Lysozyme Crystallization: Phase Diagram, Protein-Protein Interaction, Nucleation Kinetics and Growth Kinetics, *Crys. Growth Des.*, **2010**, 10, 548-558.
- [12] Yamazaki T.; Kimura Y.; Vekilov P. G.; Furukawa E.; Shirai M.; Matsumoto H.; Van Driessche A. E. S.; Tsukamoto K. Two Types of Amorphous Protein Particles Facilitate Crystal Nucleation, *Proc. Natl. Acad. Sci. USA*, **2017**, 114, (9), 2154-2159.
- [13] Lifshitz, I. M.; Slyozov, V. V. The Kinetics of Precipitation from Supersaturated Solid Solutions, *J. Phys. Chem. Solids*, **1961**, 19, 35-50.
- [14] Buyevich, Yu. A.; Mansurov, V. V. Kinetics of the Intermediate Stage of Phase Transition in Batch Crystallization, *J. Crystal Growth*, **1990**, 104, 861-867.

- [15] Iggländ, M.; Mazzotti, M. Population Balance Modeling with Size-Dependent Solubility: Ostwald Ripening, *Crys. Growth Des.*, **2012**, 12, 1489-1500.
- [16] Barlow, D. A. Theory of the Intermediate Stage of Crystal Growth with Applications to Protein Crystallization, *J. Crystal Growth*, **2009**, 311, 2480-2483.
- [17] Barlow, D. A. Theory of the Intermediate Stage of Crystal Growth with Applications to Insulin Crystallization, *J. Crystal Growth*, **2017**, 470, 8-14.
- [18] Barlow D. A.; Baird J. K.; Su C.-H. Theory of the von Weimarn Rules Governing the Average Size of Crystals Precipitated from a Supersaturated Solution, *J. Crystal Growth*, **2004**, 264, 417-420.
- [19] Makoveeva, E. V.; Alexandrov, D. V. A Complete Analytical Solution of the Fokker-Planck and Balance Equations for Nucleation and Growth of Crystals, *Phil. Trans. Roy. Soc. A*, **2017**, 376, 20170327.
- [20] Alexandrov D. V.; Malygin A. P. Transient nucleation kinetics of crystal growth at the intermediate stage of bulk phase transitions, *J. Phys. A: Math. Theor.*, **2013**, 46, 455101-455108.
- [21] Schlichtkrull J. Insulin Crystals V. The Nucleation and Growth of Insulin Crystals, *Acta Chem. Scan.*, **1957**, 11, 439-460.
- [22] Buyevich, Yu. A.; Ivanov A. O. Kinetics of Phase Separation in Colloids, *Physica A*, **1993**, 192, 375-390.
- [23] Sugimoto, T. Kinetics of Reaction-Controlled Ostwald Ripening of Precipitates in the Steady State, *J. Colloid Interface Sci.*, **1978**, 63, (2), 369-377.
- [24] Sugimoto, T. General Kinetics of Ostwald Ripening of Precipitates, *J. Colloid Interface Sci.*, **1978**, 63, (1), 16-26.
- [25] Glatter O.; Kratky O. *Small Angle X-ray Scattering*; Academic Press: London, 1982; p. 23.
- [26] Ovchinnikov, A. A.; Timashev, S. F.; Belyy, A. A. *Kinetics of Diffusion Controlled Chemical Processes*; Nova Science Publishers: New York, 1989; p. 11.
- [27] Yürüdü, C.; Jones, M. J.; Ulrich, J. Modeling of Diffusion for Crystal Growth, *Soft Materials*, **2012**, 10, (1-3), 257-284.
- [28] Wilcox, W. R. The Role of Mass Transfer in Crystallization Processes. In *Preparation and Properties of Solid State Materials Vol. 1, Aspects of Crystal Growth*; R. A. Lefever Ed.; Marcel Dekker: New York, 1971; p. 97.
- [29] Durbin, S. D.; Feher, G. Crystal Growth Studies of Lysozyme as a Model for Protein Crystallization, *J. Crystal Growth*, **1986**, 76, 583-592.
- [30] Laudise, R. A. *The Growth of Single Crystals*; Prentice-Hall: New Jersey, 1970; p. 93.

- [31] Panayotounakos, D. E.; Zarmoutis, T. I. Construction of Exact Parametric or Closed Form Solutions of Some Unsolvability Classes of Nonlinear ODEs (Abel's Nonlinear ODEs of the First Kind and Relative Degenerate Equations), *Int. J. Math. Math. Sci.*, **2011**, 387429.
- [32] Pan, W.; Kolomeisky, A. B.; Vekilov, P. G. Nucleation of Ordered Solid Phases of Proteins via a Disordered High-Density State: Phenomenological Approach, *J. Chem. Phys.*, **2005**, 122, 174905.
- [33] Nanev, C. N.; Tonchev, V.D.; Hodzhaoglu, F. V. Protocol for Growing Insulin Crystals of Uniform Size, *J. Crystal Growth*, **2013**, 375, 10-15.
- [34] Braun, M. K.; Grimaldo, M.; Roosen-Runge, F.; Hoffmann, I.; Czakkel, O.; Sztucki, M.; Zhang, F.; Schreiber, F.; Seydel, T. Crowding-Controlled Cluster Size in Concentrated Aqueous Protein Solutions: Structure, Self- and Collective Diffusion, *J. Phys. Chem. Lett.*, **2017**, 8, 2590-2596.
- [35] Fusco, D.; Charbonneau, P. Soft Matter Perspective on Protein Crystal Assembly, *Colloids Surf. B*, **2016**, 137, 22-31.