1-1-2009

Model for amorphous aggregation processes

Samual Stranks
University of Adelaide

Heath Ecroyd
University of Wollongong, heathe@uow.edu.au

Steve Van Sluyter
Australian Wine Research Institute - Adelaide

Elizabeth J. Waters
Grape Wine Research Institute - Adelaide

John A. Carver

See next page for additional authors

Model for amorphous aggregation processes

Abstract

The amorphous aggregation of proteins is associated with many phenomena, ranging from the formation of protein wine haze to the development of cataract in the eye lens and the precipitation of recombinant proteins during their expression and purification. While much literature exists describing models for linear protein aggregation, such as amyloid fibril formation, there are few reports of models which address amorphous aggregation. Here, we propose a model to describe the amorphous aggregation of proteins which is also more widely applicable to other situations where a similar process occurs, such as in the formation of colloids and nanoclusters. As first applications of the model, we have tested it against experimental turbidimetry data of three proteins relevant to the wine industry and biochemistry, namely, thaumatin, a thaumatinlike protein, and alpha-lactalbumin. The model is very robust and describes amorphous experimental data to a high degree of accuracy. Details about the aggregation process, such as shape parameters of the aggregates and rate constants, can also be extracted.

Keywords

amorphous, aggregation, processes, model, CMMB

Disciplines

Life Sciences | Physical Sciences and Mathematics | Social and Behavioral Sciences

Publication Details

Stranks, S.D., Ecroyd, H., Van Sluyter, S., Waters, E.J., Carver, J.A., von Smekal, L. (2009). Model for amorphous aggregation processes. Physical Review E, 80, 051907.

Authors

Samual Stranks, Heath Ecroyd, Steve Van Sluyter, Elizabeth J. Waters, John A. Carver, and Lorenz von Smekal

Model for amorphous aggregation processes

Samuel D. Stranks,[1,2] Heath Ecroyd,[1,*] Steven Van Sluyter,[2,3] Elizabeth J. Waters,[2]
John A. Carver,[1] and Lorenz von Smekal[1,4,†]

[1]*School of Chemistry and Physics, The University of Adelaide, South Australia 5005, Australia*
[2]*The Australian Wine Research Institute, P.O. Box 197, Glen Osmond, South Australia 5064, Australia*
[3]*School of Botany, The University of Melbourne, Victoria 3010, Australia*
[4]*Institut für Kernphysik, Technische Universität Darmstadt, Schlossgartenstr. 9, 64289 Darmstadt, Germany*

(Received 23 January 2009; published 9 November 2009)

The amorphous aggregation of proteins is associated with many phenomena, ranging from the formation of protein wine haze to the development of cataract in the eye lens and the precipitation of recombinant proteins during their expression and purification. While much literature exists describing models for linear protein aggregation, such as amyloid fibril formation, there are few reports of models which address amorphous aggregation. Here, we propose a model to describe the amorphous aggregation of proteins which is also more widely applicable to other situations where a similar process occurs, such as in the formation of colloids and nanoclusters. As first applications of the model, we have tested it against experimental turbidimetry data of three proteins relevant to the wine industry and biochemistry, namely, thaumatin, a thaumatinlike protein, and α-lactalbumin. The model is very robust and describes amorphous experimental data to a high degree of accuracy. Details about the aggregation process, such as shape parameters of the aggregates and rate constants, can also be extracted.

DOI: 10.1103/PhysRevE.80.051907 PACS number(s): 87.15.km, 87.10.Ed, 87.15.nr

I. INTRODUCTION

Amorphous or disordered aggregation is a process that occurs often in nature. For example, it is associated with several age-related diseases such as cataract [1] and can result in the formation of protein haze in white wines [2]. It is also the bane of protein expression and purification for many biochemists, leading to protein precipitation and subsequent loss of protein, data, and time. Despite amorphous protein aggregation being such a significant problem, it has received little research attention and is often overshadowed by studies of protein aggregation leading to highly structured linear amyloid fibrils associated with neurodegenerative disorders, such as Alzheimer's and Parkinson's diseases [3].

Crystallin proteins in the eye lens are thought to be destabilized due to a lifetime of oxidative and radiative stress and unfold into aggregation-prone species that are precursors to cataract formation [4]. Cataract affects many aged people and is a particular problem in developing countries where, as a result, an estimated 24 million people are blind [5]. Currently, the principal treatments for cataract are removal of the opacified lens and replacement with an artificial lens [6].

Protein wine haze presents a significant problem to the wine industry. Currently, bentonite clay is used to adsorb the troublesome proteins from white wine which is then filtered back off the clay [2]. While this is effective in removing the proteins, it adversely affects the quality and quantity of the treated wine and also presents waste disposal issues [7]. It is estimated that the cost of bentonite fining to the wine industry worldwide is in the order of $300–500 million per annum [7]. Studies that describe the mechanism of protein wine haze formation are highly desirable as they will lead to improved technologies, maintaining wine quality and decreasing the costs of production [2].

In order to find solutions to these problems, there is a need to understand and model the processes and kinetics of the amorphous aggregation phenomena. There are many mathematical models in the literature for linear aggregation mechanisms, as reviewed by Morris *et al.* [8]. A unique approach was undertaken by Flyvbjerg *et al.* [9,10], who studied the self-assembly of microtubules, a linear aggregation mechanism, and derived a mathematical model to describe the phenomenon. However, there are few, if any, mathematical models of amorphous aggregation. In this paper, we propose a model for amorphous aggregation processes and demand that certain constraints, which are found by performing a data analysis similar to that by Flyvbjerg *et al.* [9], are imposed. The data shown here are turbidity measurements for the aggregation of heat-stressed thaumatin, a protein structurally very similar to thaumatinlike (TL) proteins, which play a major role in protein wine haze formation [11]. We fit these data using our amorphous aggregation model. The analysis is also performed on α-lactalbumin (α-LA), a well-characterized protein that amorphously aggregates upon reduction of its disulfide bonds [12], and is often used as a test protein to monitor the activity of molecular chaperone proteins [12,13]. In addition, we monitor the aggregation of, and apply our model to, a *Vitis vinifera* TL grape protein. These three diverse proteins give very similar results, indicating the broad application of the model and its applicability to other systems which behave in a similar manner, such as the formation of colloids and nanoclusters [14,15] and vapor condensation [16].

II. EMPIRICAL DATA ANALYSIS

Figure 1(a) shows a set of experimental turbidity (A) data for a range of initial concentrations of thaumatin (between

*Present address: School of Biological Sciences, University of Wollongong, Wollongong, NSW 2522, Australia.
†Corresponding author; lorenz.smekal@physik.tu-darmstadt.de

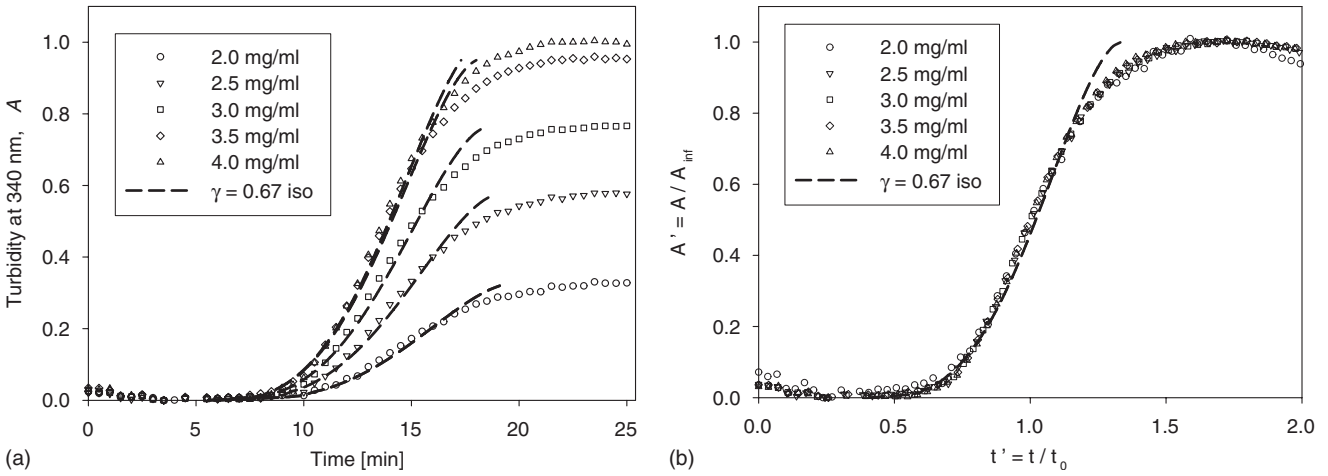

FIG. 1. (a) Typical experimental data for thaumatin at various initial concentrations, incubated at 60 °C in a model wine solution (see Appendix A for details). The dashed line overlaid shows the fits to the data by the model derived in this paper with a spherical exponent, $\gamma=2/3\approx0.67$. (b) The experimental data from (a), with the time and turbidity axes scaled by the half time, t_0, and asymptotic turbidity, A_∞, respectively, for each series. The data sets "collapse" to a single curve, thus obeying a scaling law. The dashed curve represents the same unique fit by the model to this scaled data (with $\gamma=0.67$) as the family of curves in (a).

2.0 and 4.0 mg/ml) upon heating at 60 °C. Overlaid are fits by the model to the experimental data. Data and fits for the TL protein and α-LA are shown in Figs. 7 and 8 of Appendix A, respectively.

The data points follow similar sigmoidal curves, with initial lag times, followed by a steep rise and subsequent plateau. The curves differ only by the magnitude of the increase in turbidity due to different initial protein concentrations; the lag times are very similar. Therefore, the issue of whether they differ only through the overall time and turbidity axes is addressed following the reasoning of Flyvbjerg *et al.* [9]. If that is the case, then the curves are said to "scale" over one another, following a *scaling law*.

We do this by finding the *asymptotic turbidity*, A_∞, and the time taken to reach half this value, the *half time*, t_0, for each data set. Upon scaling the turbidity and time axes with their respective characteristic parameters, the data collapse to a single curve [Fig. 1(b) and Figs. 7 and 8 in Appendix A]. The interpretation of this is that, independent of the initial monomer concentration, a single mechanism of protein aggregation is present over the concentration range considered. This allows use of a single curve for the modeling which may be unscaled to reproduce the original data.

Furthermore, we seek to obtain a relationship between the characteristic parameters A_∞ and t_0. Following the analysis in Ref. [9], a double-logarithmic plot of t_0 versus A_∞ yields a straight line (Fig. 2 and Figs. 9 and 10 in Appendix B), which indicates a *power law* relationship of the form

$$t_0 = \tau A_\infty^{-m}, \tag{1}$$

where $(-m)$ is the slope of the double-logarithmic plot and τ is some constant of proportionality.

For thaumatin, the mean result of the three replicates is $m=0.077\pm0.002$, where the quoted error is the SE of the mean. For the grape TL protein and α-LA, we obtain $m=0.039\pm0.011$ and 2.04 ± 0.39, respectively (see Figs. 9 and 10 in Appendix B).

Consequently, without the use of any theory, we have empirically obtained two independent constraints on our model from the experimental observations—the scaling and power laws.

III. MODEL

A. Linear versus amorphous aggregation

In simplistic terms, the aggregation of monomeric entities (e.g., proteins) to form linear chains such as amyloid fibrils and the self-assembly of microtubules, as modeled by [9], are linear mechanisms; the aggregating "monomers" can

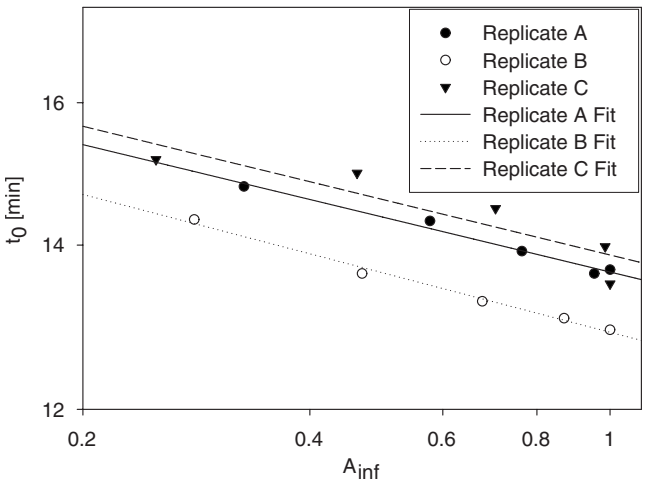

FIG. 2. Double logarithmic plots of t_0 against A_∞ for thaumatin, showing all three replicates. The linear fits of $\log_{10}(t_0/\tau)$ versus $\log_{10}(A_\infty)$ for replicates A–C give slopes $(-m)$ of −0.074, −0.080, and −0.075, R^2 values of 0.98, 0.99, and 0.82, and τ values of 13.65, 12.90, and 13.87 min, respectively, corresponding on these plots to the t_0 values with $A_\infty=1.0$. This gives mean values of $m=0.077\pm0.002$ and $\tau=13.48\pm0.29$ min, where errors are quoted as standard errors (SEs) of the mean.

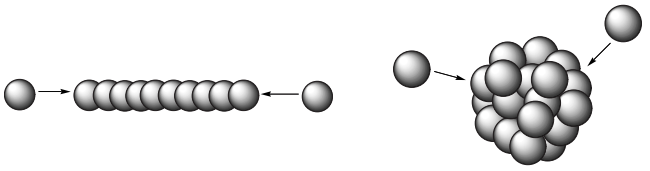

FIG. 3. *Left*: linear aggregation mechanisms are ordered one-dimensional processes and monomers can only add onto the ends of the aggregate chain. *Right*: amorphous aggregation is a disordered three-dimensional process, where monomers can add on to aggregates from any direction.

only add on to either end of the aggregate in a one-dimensional ordered manner. As the aggregate increases in length, the rate of aggregation should therefore not increase because there is no greater surface area available to the monomers.

Conversely, amorphous aggregation can be thought of as a disordered three-dimensional process, where monomers can add from any direction. Consequently, we must account for the fact that as the aggregates increase in size, the surface area available to the monomers increases and hence there should be a corresponding increase in the rate of aggregation. This will be the third constraint on the model, in addition to the scaling and power laws. Figure 3 compares the linear and amorphous aggregation mechanisms schematically.

B. General framework

A similar formalism to that introduced in Refs. [9,10] is followed here. We let $M(t)$ denote the total mass of monomers contained in the aggregates and $c(t)$ be the monomer concentration (both in units of arbitrary concentration, such as mass density). Clearly, conservation of mass gives us the relation [9,10]

$$M(t) + c(t) = c(0), \tag{2}$$

where $c(0) \equiv c_0$ is the initial protein monomer concentration (at time $t=0$). In addition, we see from Eq. (2) that M_∞, the asymptotic mass as time approaches infinity, is equal to the initial monomer concentration, c_0, because $c(t)$ goes to zero [9,10]. Furthermore, it is assumed that turbidity, $A(t)$, is proportional to the aggregate volume and hence mass in aggregates, $M(t)$.[1] Thus, we see that $A_\infty \propto M_\infty = c_0$.

Note that the asymptotic turbidities, A_∞, thus scale linearly with the initial monomer concentration, c_0. Inspection of Fig. 1 indicates that experimentally, the linear correlation between A_∞ and c_0 is only approximate and we attribute this primarily to experimental errors in preparing the monomer concentrations. In any case, the exact values for concentration are not important and are not required in the analysis at any point. It is only necessary to obtain data sets for *different* concentrations for a given protein.

The data shown are all relative to the asymptotic turbidity at the reference concentration, $c_0^{\text{ref}} = 4.0$ mg/ml.

The scaling law observed earlier gives motivation to introduce dimensionless "scaling variables." Therefore, we define t', A', M', and c' such that

$$t' = \frac{t}{t_0}, \quad A' = \frac{A}{A_\infty}, \quad M' = \frac{M}{M_\infty} = \frac{M}{c_0}, \quad c' = \frac{c}{c_0}, \tag{3}$$

where $t_0 = \tau A_\infty^{-m} = \tau (c_0/c_0^{\text{ref}})^{-m}$ with $c_0^{\text{ref}} = 4.0$ mg/ml and the exponent $m > 0$ is the parameter determined earlier for each of the proteins. This accounts for the power law constraint of Eq. (1).

C. Amorphous aggregation model

The mass in the aggregates, $M(t)$, will increase with time as monomers continue to add onto the cluster. The rate of this increase will depend on the concentration, c, of available monomers. One therefore expects that on one hand, this rate will obey a power law of the form

$$\frac{dM}{dt} \propto c^r \tag{4}$$

for some as yet undetermined exponent, $r \geq 0$. We keep this power general, allowing it to deviate from the classical value of 1 for simple first-order kinetics. First of all, our system is heterogeneous—the "reacting" monomer entities need to first be activated by being partially unfolded. The denaturation process, e.g., achieved by heating, results in a statistical distribution of monomer states, some of which are completely unfolded, others which remain in their native form, and many others in between, i.e., which are partially unfolded. Thus, only a certain proportion s is able to aggregate at any given time. Moreover, in the presence of density dependent mechanisms, such as collisions between monomers which may lead to activation or refolding, it is to be expected that this proportion also depends on their density, as modeled, for example, by another power law, $s \propto c^\delta$, leading to $r = 1 + \delta$. Hence we allow for deviation from $r=1$ in Eq. (4).

On the other hand, we must also take into account the requirement that the rate of aggregation increases as the aggregates grow in size. Assuming for simplicity that their average size grows in proportion to some other power, γ, of their mass, M, we require a factor of the form

$$\frac{dM}{dt} \propto M^\gamma, \tag{5}$$

where the exponent $\gamma \geq 0$ is also *a priori* unknown. Combining Eqs. (4) and (5), we obtain a rate equation of the form

$$\frac{dM}{dt} = f c^r M^\gamma, \tag{6}$$

where the constant f is to be determined experimentally.

[1] This relationship holds when treating the turbidity measurements as being due to Rayleigh scattering, where the size of the aggregates is much less than the wavelength used to analyze them [17]. Transmission electron microscope images (not shown) verified this, where only the largest protein aggregates just exceed the wavelength of the turbidity measurements, 340 nm. Nevertheless, the approximation is valid for the majority of the aggregation time series and is therefore assumed to hold for the entire process.

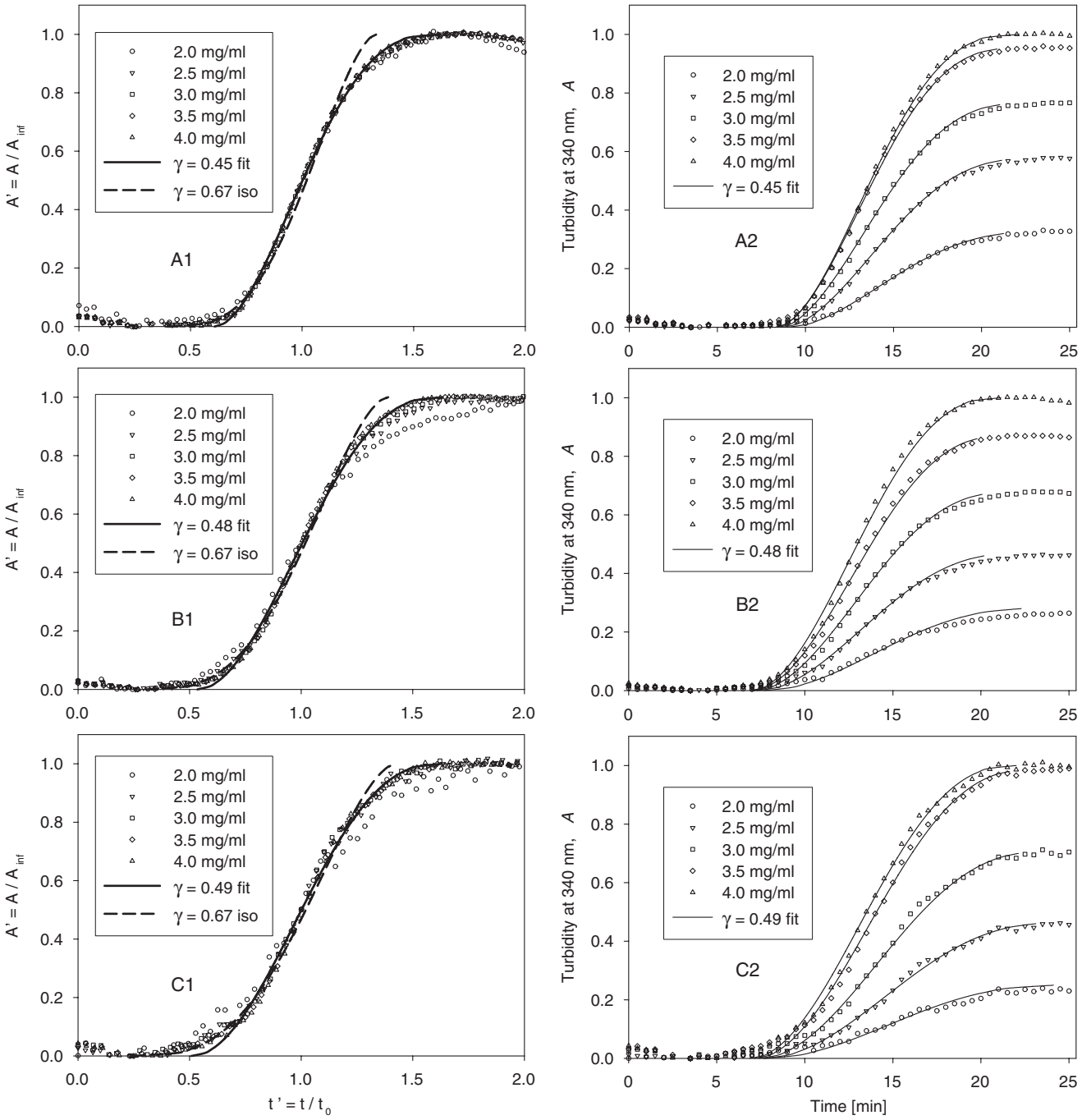

FIG. 4. *Left column*, A1–C1: thaumatin experimental data with the time and turbidity axes scaled with the half time, t_0, and asymptotic turbidity, A_∞, respectively. The three figures represent three repetitions of the experiment (replicates A–C). The solid curves are the fits by the model to the scaled data with γ_{fit} values and the dashed curves are fits by the isotropic (iso) model with $\gamma=0.67$. *Right column*, A2–C2: thaumatin experimental data from the left column unscaled to reproduce the original time series, overlaid with the fits by the model using γ_{fit} for each.

In order to account for the observed scaling, we first introduce the dimensionless variables of Eq. (3). In terms of these variables, Eq. (6) reads

$$c_0^{m+1}\frac{dM'}{dt'} = c_0^{r+\gamma} f' c'^{r} M'^{\gamma}. \tag{7}$$

Herein, $f' = \tau (c_0^{\mathrm{ref}})^m f$ is a scaled dimensionless rate constant to determine the overall rate of the process.

The observed scaling requires the rate equation in terms of the scaling variables [Eq. (7)] to be independent of the initial concentration, c_0, as in Refs. [9,10]. Therefore, an additional constraint is $r=m+1-\gamma$. With the power law in Eq. (1) and scaling, the amorphous aggregation model then predicts an exponent in Eq. (4) different from $r=1$ whenever $m \neq \gamma$.

Since Eq. (2) tells us that $c' = 1 - M'$, the final equation for

the amorphous aggregation model, including an appropriate initial condition, then reads

$$\frac{dM'}{dt'} = f'(1-M')^{m+1-\gamma}M'^{\gamma}, \quad M'(0)=0. \tag{8}$$

While the exponent m is obtained from the empirical data analysis discussed above, the exponent γ, introduced in Eq. (5), and the rate constant f' may be used to model different systems.

D. General solution

To allow for a potential lag time before the onset of aggregation, we introduce a characteristic parameter t'_{lag} associated with the sigmoidal curves by replacing $t' \to t' - t'_{\text{lag}}$. The general solution to Eq. (8) is then found, by integration, to be implicitly given by the equation

$$t' = \frac{M'^{1-\gamma}\,{}_2F_1(1-\gamma, 1+m-\gamma; 2-\gamma; M')}{f'(1-\gamma)} + t'_{\text{lag}}, \tag{9}$$

where ${}_2F_1$ is a hypergeometric function defined by the expansion

$${}_2F_1(a,b;c;z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!}, \tag{10}$$

in which $(a)_k = a(a+1)(a+2)\ldots(a+k-1) = \frac{\Gamma(a+k)}{\Gamma(a)}$ are the Pochhammer symbols and Γ is the gamma function [18]. For $z=M'$ in our case it is straightforward to verify that the series converges because $0 \le M' \le 1$.

Depending on the model assumptions, i.e., whether a lag phase is to be included and/or whether the exponent γ is fixed by a model assumption on the geometry of the amorphous aggregation process or used as a free parameter, we have various options to fit the implicit solution of Eq. (9) for $M'(t')$ to the scaled experimental data. We can either use the effective rate constant f' as the only parameter (with all others fixed) or include the aggregation geometry exponent γ and/or an additional lag time via t'_{lag} to perform the fits, thus including up to at most three free parameters.

E. Aggregation geometry

As discussed above, the exponent γ in the aggregation rate equation is introduced to account for an increase in the area available to the monomers to add on to the aggregate. This area can be the total surface area of the aggregates in some cases but it might be only some part of that in others. The exponent γ describes how the *effective area* relevant to the aggregation process (i.e., presented to the monomers) increases with the volume and hence the mass of the aggregates.

In linear aggregation such as amyloid fibril formation, for example, this effective area is independent of the volume which corresponds to $\gamma=0$ for one-dimensional aggregation.

Another interesting special case is that of an isotropic aggregation process in three dimensions as also sketched in Fig. 3. In this case, the effective area, A_{eff}, and hence rate of aggregation increase with the square of the aggregate's radius R, while the volume V increases as R^3. Therefore, at constant density,

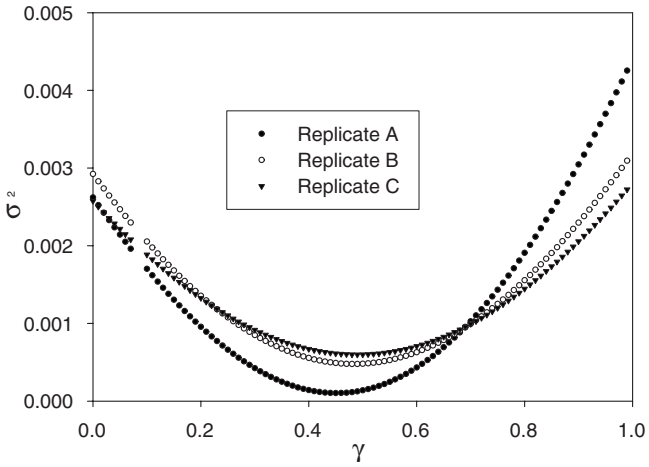

FIG. 5. Plot showing estimated variance between the fitted curve and thaumatin experimental data for values of γ between 0 and 1. Replicates A–C yield γ_{fit} values (i.e., minima of plots) of 0.45, 0.48, and 0.49, respectively, and a mean value of $\gamma_{\text{fit}} = 0.47 \pm 0.02$, where the quoted error is the standard deviation. The large jump in error near $\gamma=0.10$ corresponds to $\gamma \approx m$, which leads to a special case of the hypergeometric function solution.

$$A_{\text{eff}} \propto V^{2/3} \propto M^{2/3}. \tag{11}$$

We call $\gamma=2/3$ the exponent for *spherical* or *isotropic aggregation*.

Larger values, $\gamma > 2/3$, are possible when the area increases faster than that relative to the volume. This could be due to a roughening of the surface, for example.

When a deformation develops, on the other hand, there are preferential directions of aggregation singling out a certain fraction of the total surface, with the result that the relevant effective area, A_{eff}, will grow slower than the aggregate's total surface. Simple examples are prolate deformations which in the extreme case lead to linear aggregation again or oblate deformations. The latter might be modeled by a flat cylinder of increasing radius R but constant height. Then, only the outer wall of the cylinder attracts further monomers and the effective area increases only linearly with the radius, $A_{\text{eff}} \propto R$, while the volume increases as R^2. This gives

$$A_{\text{eff}} \propto V^{1/2} \propto M^{1/2}, \tag{12}$$

so that we call $\gamma=1/2$ the exponent for *cylindrical aggregation*, which of course is a two-dimensional process.

IV. THAUMATIN RESULTS

A. Isotropic model

We first assume isotropic aggregation and fix $\gamma=2/3 \approx 0.67$ as the spherical exponent in our fits. The dashed lines in Figs. 1(a) and 1(b) show the nonlinear regression fit to one of the replicates for thaumatin, fixing $\gamma=2/3$ and using f'

TABLE I. Summary of experimentally and computationally derived parameters for the three proteins analyzed.

Protein	m	τ (min)	γ_{fit}	r [a]	t'_{lag}	$\bar{f}=f'/\tau$ (min^{-1})
Thaumatin	0.077[b] $\pm$0.002 [c]	13.5 ± 0.3	0.47 ± 0.02 [d]	0.61 ± 0.02	0.55 ± 0.03	0.25 ± 0.01
TL protein	0.039 ± 0.011	12.1 ± 0.1	0.37 ± 0.09	0.67 ± 0.08	0.63 ± 0.03	0.27 ± 0.01
α-lactalbumin	2.04 ± 0.39	48.6 ± 13.2	0.56 ± 0.18	2.48 ± 0.21	0.11 ± 0.06	0.09 ± 0.03

[a] $r=1+m-\gamma_{\mathrm{fit}}$.
[b] m, τ, γ_{fit}, and t'_{lag} are each quoted as mean values across the repetitions.
[c] All quoted errors are SEs of the mean across the repetitions of the experiments unless stated otherwise.
[d] The errors in γ_{fit} are quoted as standard deviations (SDs) to give a more generous error estimate to account for the finite step size in the γ increment loops.

and t'_{lag} as fitting parameters.[2] This is the same fit as the dashed line for replicate A1 in Fig. 4. Almost identical results are obtained for replicates B1 and C1 (dashed lines), also shown in Fig. 4.

By inspection of the fits, the spherical exponent, $\gamma=2/3$, for isotropic (iso) aggregation is very well suited to describe the data for the earlier times, $t<t_0$. At later times, $t>t_0$, the quality of the fits deteriorates, however. The assumption of isotropic aggregation appears to break down in the final phase of the process typically after the half time, t_0.

B. Global fits

Since these observed deviations from the isotropic aggregation model could either be due to anisotropies in the aggregation or roughening of the surface, as explained above, we adopt an alternative model in which the exponent γ is varied to optimize the fits over a range of times around t_0, so chosen as to include the essential part of the aggregation process (excluding very early and very late times). To this end, we scan the range of exponents $0\leq\gamma<1$ in increments of 0.01.[3] This is done using a simple MATHEMATICA routine to obtain for each value of γ a nonlinear regression fit using f' and t'_{lag} as the two free parameters. The estimate of the variance is plotted for each value against γ in Fig. 5.[4]

The data range for the fits around t_0 is chosen as $2/3\leq t'\leq 4/3$.[5] This window includes the majority of the crucial central data and is mainly chosen to exclude the extremities. The fits are otherwise stable under variations of this range within reasonable limits.

For each of the three replicates we separately obtain a value of γ, termed γ_{fit}, minimizing the estimated variance, σ^2, as shown in Fig. 5. These values for the best fit over the transition range lead to a global average over the three sets of $\gamma_{\mathrm{fit}}=0.47\pm0.02$ for thaumatin. Figures 11 and 12 in Appendix B, respectively, show similar plots for the TL protein and α-LA, resulting in $\gamma_{\mathrm{fit}}=0.37\pm0.09$ and $\gamma_{\mathrm{fit}}=0.56\pm0.18$, respectively.

Replicates A1, B1, and C1 in Fig. 4 show the fits (solid lines) resulting for each of the three sets of scaled thaumatin experimental data (symbols) obtained with their respective optimized geometry exponents, γ_{fit}. An excellent agreement between fit and data is observed, in particular at later times, $t>t_0$, where the isotropic model (iso—dashed lines) starts to fail.

V. DISCUSSION

Our interpretation of these results is that they provide compelling evidence that the aggregation process for thaumatin is very well described by an isotropic three-dimensional process (corresponding to the spherical exponent, $\gamma=2/3$) in the early phases leading up to the half time, t_0. After that, better fits are obtained for a value of γ_{fit} near 1/2 which implies that the aggregates start to develop more of an oblate deformation eventually leading to an aggregation corresponding to the cylindrical exponent, $\gamma=1/2$. The two-dimensional model matches the data particularly well during the late stages of the aggregation process and, in this phase, it provides a very good description even considerably beyond the range used for the fits.

Replicates A2, B2, and C2 in Fig. 4 show the original thaumatin experimental data with the same fits using the γ_{fit} values, but after the original scaling from the left column, corresponding to Eq. (3), is undone.

Table I gives a summary of the values obtained for the various parameters of the three proteins. In order to compare different systems such as different proteins, it is furthermore useful to define a proper rate constant, $\bar{f}\equiv f'/\tau$, upon division by the characteristic time τ from the power law of Eq.

[2] The MATHEMATICA (Wolfram Research, Champaign, IL) function NonLinearRegress is used for all of the fitting procedures. In this procedure, the estimates of the model (fitting) parameters are chosen to minimize the χ^2 merit function, a function which measures the difference between data and the fitting model for a particular choice of the parameters given by the sum of squared residuals [19].

[3] Note that $\gamma=1$ is a limiting case in that it leads to an exponential growth at early times, while simple power laws $M'(t')\sim t'^{1/(1-\gamma)}$ are obtained at small t' for $0\leq\gamma<1$.

[4] The MATHEMATICA statistical output EstimatedVariance is equivalent to the squared sum of fit residuals (difference between data and fit estimate at any point) divided by the number of degrees of freedom [19].

[5] The fitting window midpoint for $\gamma=2/3$ was chosen to be the inflection point which is reached slightly before the half time, t_0. The difference is insignificant, however.

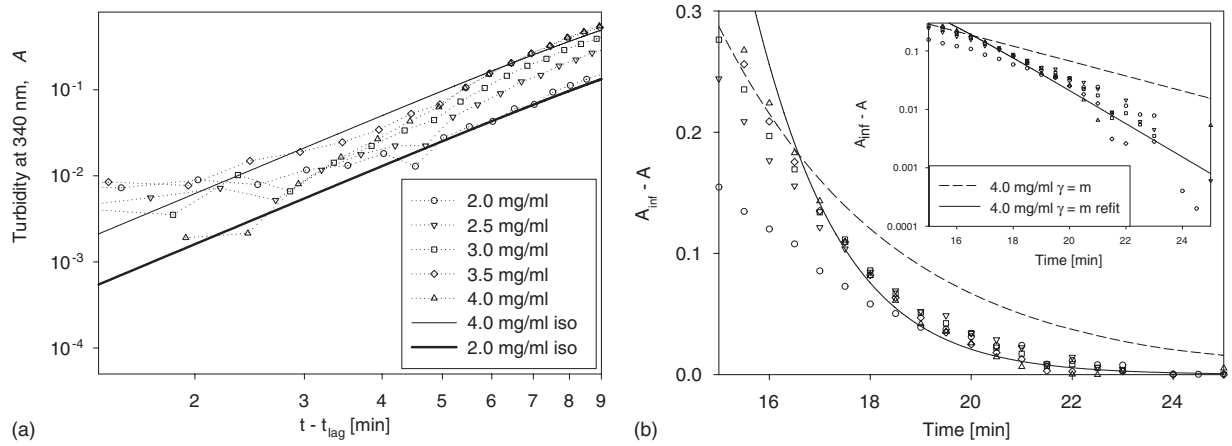

FIG. 6. (a) Double logarithmic plot of thaumatin experimental aggregation (turbidity) data from Fig. 1(a), showing early times. The data follow a straight line behavior, and this property can be readily captured by the model. (b) Thaumatin experimental aggregation (turbidity) data from Fig. 1(a) for late times, presented as monomer concentration ($c \propto A_\infty - A$). The plots exhibit asymptotic decay, and the inset shows the $A_\infty - A$ axis logarithmically, giving approximate straight line behavior. These late-time properties can be accounted for when setting $r=1$ which implies $\gamma=m$. The dashed line shows the original fit for $c_0=4.0$ mg/ml with $\gamma=m=0.08$ as obtained from the power law in Eq. (1). The solid line shows the result of an independent fit of this parameter to the late-time data resulting in $m=\gamma=0.45$ which is precisely the same value as obtained from the global fit for the geometry exponent γ to these data [replicate A in Fig. 4 (*see panel* A2)].

(1). This definition is independent of the reference concentration, c_0^{ref}, used in the turbidity measurements, and the results for $\bar{f}$ are also quoted in Table I. Details of corresponding analyses for the TL protein and α-LA are included in Appendixes A and B.

We found that the values of r for thaumatin (and the TL protein) were less than 1, the value expected from first-order kinetics. We rationalize this on the basis that heating is used to activate the monomers for aggregation. As a result, a statistical distribution of monomers is present, some of which are still in their native (folded) form, others which are partially unfolded, and others which are completely unfolded. Therefore, a certain proportion of the total monomers present in the system is unavailable for aggregation (e.g., those in the native form). In physical terms, when this proportion is smaller at larger densities, due to an increased number of collisions for example, this can give rise to an r value decreased from 1. Conversely, the r value for α-LA was found to be greater than 1. In this case, however, a reducing agent was used to induce aggregation by breaking the four disulfide bonds of α-LA to cause the protein to partially unfold. This chemical method results in practically all of the monomers being partially unfolded [12] and therefore activated for aggregation, and it can therefore not explain the observed r value larger than 1. Rather, the resulting value of r around 2.5 may indicate that a few nucleation steps, where more than one monomer is involved in forming the nuclei, might need to be included in the pathway to aggregation for α-LA. This is beyond the scope of the present model and will be left as a future extension.

Further verification of the validity of the model requires a thorough analysis of the early and late-time data. Figures 6(a) and 6(b) show the thaumatin experimental aggregation data from Fig. 1(a) for early and late times, respectively, emphasized on respective double-logarithmic and semilogarithmic axes (the same trends are seen for the TL protein and α-LA data, not shown here). For the early-time case, the monomer concentration changes very little ($c \sim c_0$), so the entire system is described by Eq. (5) only. Integrating this equation gives

$$M \propto t^{1/(1-\gamma)} \tag{13}$$

for early times. Therefore, our model predicts that double-logarithmic plotting of early-time aggregation data should result in straight lines (of slope $\frac{1}{1-\gamma}$). This behavior is observed very clearly as shown for thaumatin in Fig. 6(a) where we included the isotropic aggregation model fits with $\gamma=2/3$ for the limiting values of the initial monomer concentrations from Fig. 1 to indicate the band of lines expected for initial concentrations between 2.0 and 4.0 mg/ml. After about 2 min the data align with this band and we conclude that the three-dimensional isoptropic model with $M \propto t^3$ captures the early-time behavior of the aggregation process very well.

For the case of late times, one might intuitively expect a simple exponential decay of the monomer concentration. However, this requires $r=1+m-\gamma=1$. Otherwise the aggregation process would be fully completed in maybe large but strictly speaking finite time as can be seen from Eq. (9) which results in a finite $t'=t'_{\max}$ for $M'=1$ when $1+m-\gamma=r$ differs from 1. As this would appear to be rather unphysical we must assume that $r=1$ at least asymptotically at late times. Then, from Eq. (9),

$$M_\infty - M \propto \exp[-(c_0/c_0^{\mathrm{ref}})^m \bar{f} t], \tag{14}$$

where the constant of proportionality depends on $\gamma=m$.

The late-time data for $M_\infty - M$ of thaumatin are compared to such model solutions with $r=1$ and exponential decay in Fig. 6(b). The dashed line is obtained when using the value of the exponent m from the power law [Eq. (1)], i.e., with $\gamma=m=0.08$. The solid line is the result of an independent fit of this parameter to the late-time data which incidentally leads to a value $\gamma=m=0.45$ coinciding with the result of the

global fit for the geometry exponent γ_{fit} to these data, replicate A [see Fig. 4 (A2)], as described in Sec. IV B. Identical values are obtained for replicates B and C with very small errors. They are all consistent with the thaumatin average of the global fit parameter $\gamma_{\text{fit}}=0.47\pm0.02$ for approximately cylindrical aggregation, as listed in Table I.

The late-time data are then also well described, and we observe that the extracted value of the geometry exponent is insensitive to the necessary modification of r. We reiterate, however, that $r=1$ is required for the expected exponential decay of the monomer concentration at late times. The fact that this is in conflict with the empirical value of r used so far shows the limitation of the model. An obvious improvement would, of course, be to assume that the required deviations from $r=1$ in Eqs. (4) and (6) tend to zero, i.e., that $r\rightarrow1$ at late times. In physical terms such a modification would be very well justified. For thaumatin and the TL protein this is because one would expect the density dependence of the activation process to die off as monomer collisions do when their density approaches zero. Likewise, the compensation for potential nucleation steps in the case of α-LA, by adjusting the value of r to around 2.5, should really only be applied during the early phases with $r\rightarrow1$ after nucleation.

The overall effect of this modification is expected to be small, however, because the model describes the data very well over almost the entire aggregation process, including the depletion of monomers and the subsequent plateau region, as seen, e.g., in Fig. 4. Furthermore, the predicted geometry exponent is seen to be left unaffected. We therefore conclude that implementing $r\rightarrow1$ might be a rather cosmetic improvement. In doing so, the model would probably lose its mathematical simplicity which is one of its most appealing features.

Another limitation of the model in its present form is that there is no reference made to individual aggregates and their numbers. As a result, the model does not allow for the case where alternative growth pathways exist depending on the concentration of certain sizes of aggregates. For example, for large aggregate sizes, clumps of these big aggregates may begin to form rather than the simple addition of individual monomers to growing clusters. Therefore, further extensions of the model would need to account for a variable size distribution in order to rationalize and incorporate multiple aggregation pathways.

VI. CONCLUSIONS AND OUTLOOK

In this paper, we have proposed a model for aggregation involving a geometry exponent γ to model its various forms including the well-studied one-dimensional linear processes for $\gamma=0$, two-dimensional or cylindrical processes for $\gamma=1/2$, and isotropic three-dimensional processes with $\gamma=2/3$. Moreover, this exponent can provide an indication of the geometry of the process and changes in geometry of the aggregates over time.

While simple in its form, the model is widely applicable to a number of amorphous aggregation processes, not necessarily limited to protein aggregation. To use this model on other systems, the experimentally obtained data simply need to be proportional to the volume of the aggregates and obey a scaling law; the method is, therefore, not limited to turbidimetry data.

Nevertheless, this model is very well suited to describe the turbidimetry data for the thaumatin protein presented herein as a first example. In particular, for thaumatin we observe that the first half of the aggregation process is very well described by an isotropic three-dimensional form corresponding to $\gamma=2/3$, while the second half displays more of a two-dimensional cylindrical behavior with $\gamma\approx1/2$. Similar results are obtained for a TL protein and α-LA (whereby the rather limited scaling properties of the α-LA system lead to somewhat larger uncertainties—see Appendixes A and B).

A possible explanation for the geometry change at later times for the thaumatin protein may be that the hydrophobic interactions between aggregate and monomer are the primary modes of attraction [20] and, therefore, not all of the surface of an aggregate is available for a monomer to add. This is particularly the case for larger aggregates, where the hydrophobic regions will tend to point inward, exposing as little hydrophobicity as possible to the aqueous solution and hence reducing the number of sites available for monomers to add. Oblate deformations may result along with growth with reduced dimensionality in preferred directions. Whether it does or whether it leads to some other means to reduce the effective area relative to the volume is an open question which deserves further study.

An alternative explanation for the decrease in γ for later times could be that big aggregates might clump together to form huge aggregates. This would decrease the effective surface area available for monomer addition, consistent with the γ change. Furthermore, such a competing process would be more likely for later times when there are fewer monomers present and only big aggregates remain. Extensions of the model would need to account for such multiple growth pathways which depend on the size distribution of the aggregates at any time.

An analysis of the early-time behavior of the model shows that it is consistent with the experimental data and physical observations. On the other hand, we need to assume that $r\rightarrow1$ at late times to avoid the unphysical situation of a finite aggregation time. This is a minor limitation of the model which has little effect on the aggregation process as a whole. The model nevertheless reproduces the correct overall trends even at late times.

Future work will also include tests of the model in other systems. In addition, the findings could be reconciled with other data, such as time series of protein hydrodynamic radii as studied in [21]. Finally, one could incorporate the capacity to include a number of nucleation steps prior to aggregation in the model along the lines of the linear aggregation model developed in Ref. [9]. This could then be used to devise a criterion for whether nucleation is important or not in any given amorphous aggregation process.

ACKNOWLEDGMENTS

The authors thank Professor Henrik Flyvbjerg and Min-Chul Kim for their helpful comments and suggestions.

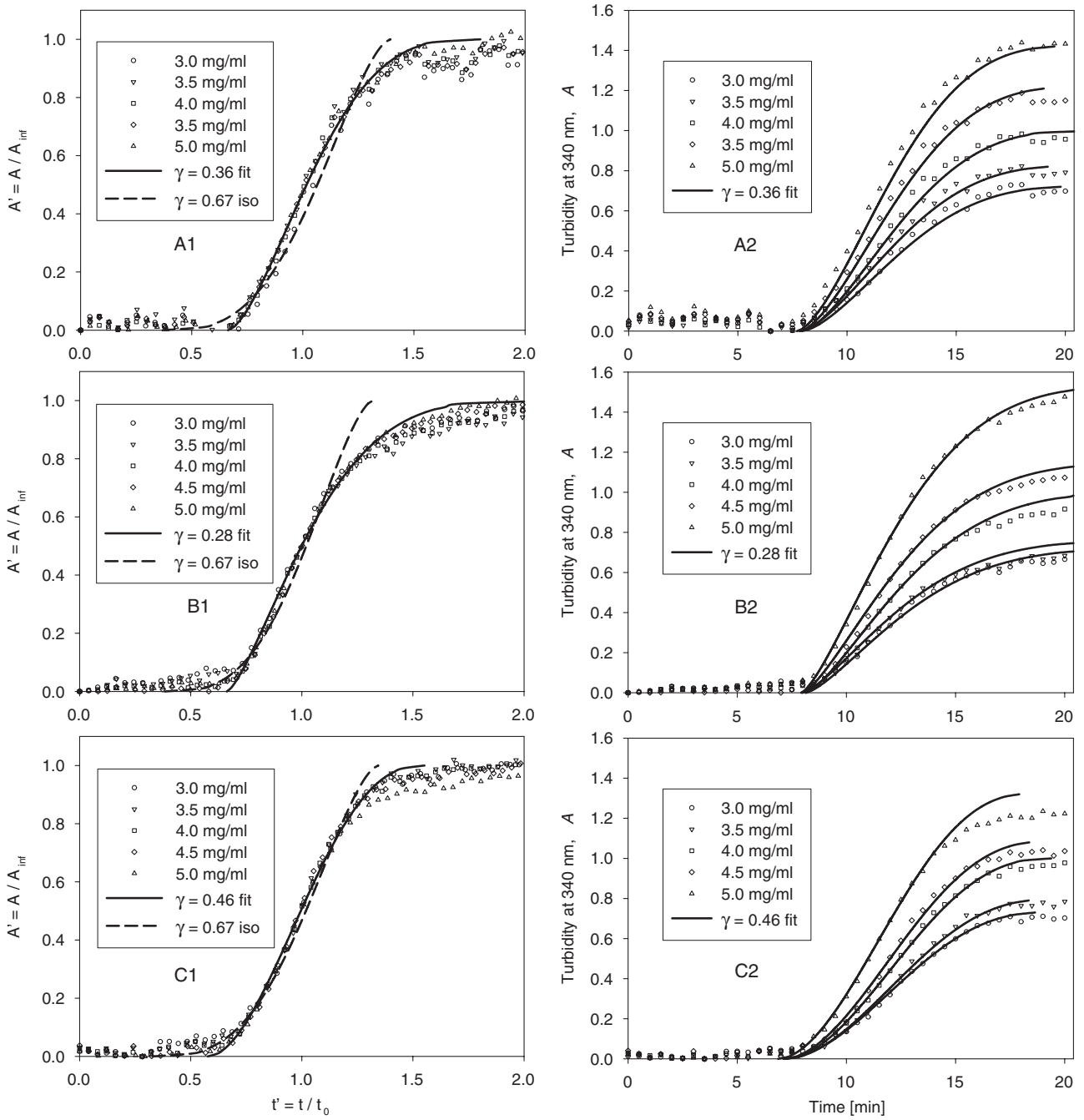

FIG. 7. Left column, A1–C1: TL protein experimental data with the time and turbidity axes scaled with the half time, t_0, and asymptotic turbidity, A_∞, respectively. The three figures represent three repetitions of the experiment (replicates A–C). The solid curves are the fits by the model to the scaled data with γ_{fit} values, and the dashed curves are fits by the (isotropic—iso) model with $\gamma=\frac{2}{3}$. Right column, A2–C2: TL protein experimental data from the left column unscaled to reproduce the original time series overlayed with the fits by the model using γ_{fit}.

S.D.S., S.V.S., and E.J.W. are supported by Australia's grapegrowers and winemakers through their investment body, the Grape and Wine Research and Development Corporation, with matching funds from the Australian Government. S.V.S. is supported by an Endeavor International Postgraduate Research Scholarship and a Melbourne International Research Scholarship, H.E. by the National Health and Medical Research Council Peter Doherty Foundation, and J.A.C. by an Australian Research Council grant.

APPENDIX A: MATERIALS AND METHODS

1. Materials

Thaumatin (mixture of I and II) from *Thaumatococcus daniellii* was purchased from MP Biomedicals (Solon, OH, USA) and calcium depleted, type III, α-LA was purchased from Sigma Chemical Co. (St Louis, MO, USA), both being used without further purification. A thaumatinlike (TL) protein was isolated (>98% purity) from Semillon grape juice

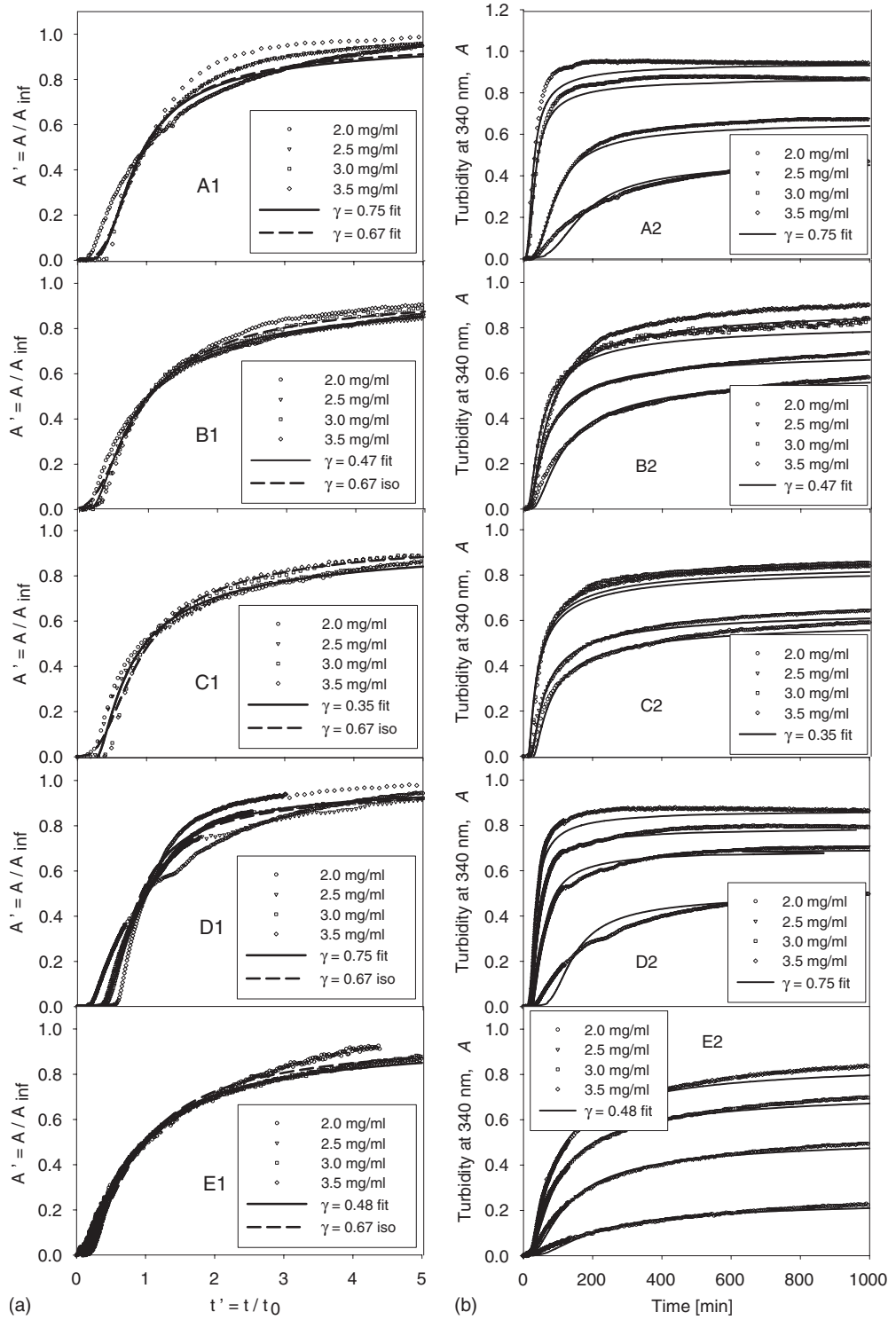

FIG. 8. Left column, A1–E1: α-LA experimental data with the time and turbidity axes scaled with the half time, t_0, and asymptotic turbidity, A_∞, respectively. The five figures represent five repetitions of the experiment (replicates A–E). The solid curves are the fits by the model to the scaled data with γ_{fit} values and the dashed curves are fits by the (isotropic—iso) model with $\gamma=\frac{2}{3}$. Right column, A2–E2: α-LA experimental data from the left column unscaled to reproduce the original time series overlayed with the fits by the model using γ_{fit}.

(Adelaide Hills, South Australia, 2005) by cation exchange and hydrophobic interaction chromatography [22]. Dithiothreitol (DTT) was obtained from Sigma. All buffers and solutions were filtered (0.45 μm) before use.

2. Determining protein concentrations

Concentrations of thaumatin (ϵ_{280}=27 755 M^{-1} cm^{-1}, mass=22 204 Da [23]) and α-LA (ϵ_{280}=28 540 M^{-1} cm^{-1}, mass=14 000 Da [24]) were determined spectrophotometri-

cally at 280 nm with a Cary 5000 UV-visible spectrophotometer (Varian, Melbourne, VIC, Australia).

For the TL protein, the concentration of the final purified protein was determined by high-performance liquid chromatography [25].

3. Monitoring aggregation by turbidimetry

Stock solutions of thaumatin (5.0 mg/ml) in a model wine buffer solution (*p*H 3.2, 12% *v*/*v* ethanol/water, 1.25 g/l potassium hydrogen tartrate, and 1 g/l Na_2SO_4) were prepared. Appropriate dilutions were made with the model wine buffer to give protein concentrations in 0.5 mg/ml increments from 2.0 to 4.0 mg/ml plus a control containing no protein (baseline). For the TL protein, the same model wine buffer was used except the *p*H was 3.5 and 100 mM potassium malate was included. The concentration range was 3.0–5.0 mg/ml in 0.5 mg/ml increments. Apo α-LA (10 mg/ml) in phosphate buffer (*p*H 7.2, 50 mM Na_3PO_4, 100 mM NaCl, and 2.5 mM ethylenediamine tetra-acetic acid) was used from 2.0–4.0 mg/ml in 0.5 mg/ml increments.

Incubation was at 60 °C with 2 s of shaking after each 30 s cycle for thaumatin and the TL protein. The α-LA aggregation experiments were performed at 37 °C (no shaking) and initiated by including DTT to a final concentration of 20 mM. The aggregation was monitored via turbidimetry at 340 nm using a Fluostar Optima microplate reader (BMG Labtechnologies, Melbourne, VIC, Australia). The data in each assay were averaged between duplicates and baseline subtracted, then normalized using the A_∞ value for 4.0 mg/ml (c_0^{ref}) for each repetition such that the A_∞ turbidity values for 4.0 mg/ml were always 1.0 (justifiable because turbidity is an arbitrary scale). This procedure produced the aggregation profiles in the right column of Fig. 4 and those of Figs. 7 and 8, whereby the thaumatin and TL protein experiments were repeated three times and five times for α-LA due to the greater variability in the data.

APPENDIX B: TL PROTEIN AND α-LA RESULTS AND DISCUSSION

Inspection of Figs. 7 and 8 (left columns) shows that the individual data for the aggregation of the TL protein and α-lactalbumin (α-LA), respectively, "scale" to an approximation. This allows us to use the scaling technique to analyze the data and fit the model to them in the same manner as described for thaumatin. It should be noted, however, that the scaling for α-LA is not as consistent as for thaumatin and the TL protein.

Once again, the double-logarithmic plots of the characteristic parameters produce straight lines for both the TL protein (Fig. 9, $m=0.039\pm0.011$) and α-LA (Fig. 10, $m=2.04\pm0.39$). The value of m for the TL protein is of the same order as thaumatin, but α-LA is two orders of magnitude greater than the other two. This is likely to be due to its dramatically different aggregation profile and that the conditions used to induce aggregation (i.e., reduction versus heating) lead to a much longer period of aggregation.

The fits by this amorphous aggregation model to the TL protein data produce strong correlations (Fig. 7) even though the data are slightly more noisy than for thaumatin. Replicates A1, B1, and C1 of Fig. 7 show that a γ value of 0.67 (the isotropic model) gives better fits for the early-time TL protein data, but the value which gives the best fit for mid to later times is $\gamma\approx0.40$ (i.e., γ_{fit} from Fig. 11). Therefore, the same interpretation can be used as with thaumatin—there appears to be a change in the aggregate geometry from isotropically aggregating entities for early times, $t<t_0$, to cylindrically aggregating entities for mid to later times, $t>t_0$. Since the change in γ is more significant for the TL protein

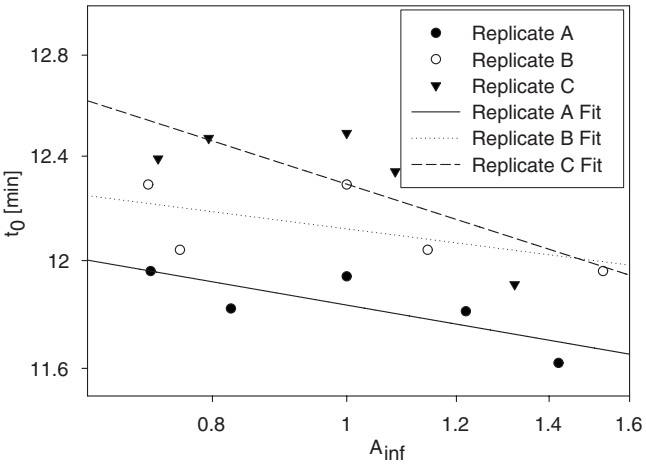

FIG. 9. Double logarithmic plots of t_0 against A_∞ for the TL protein, showing all three replicates. The linear fits of $\log_{10}(t_0/\tau)$ versus $\log_{10}(A_\infty)$ for replicates A–C, respectively, give slopes $(-m)$ of −0.033, −0.024, and −0.061, R^2 values of 0.62, 0.34, and 0.56, and τ values of 11.83, 12.12, and 12.29 min, corresponding on these plots to the t_0 values with $A_\infty=1.0$. This gives mean values of $m=0.039\pm0.011$ and $\tau=12.08\pm0.13$ min, where errors are quoted as SEs of the mean.

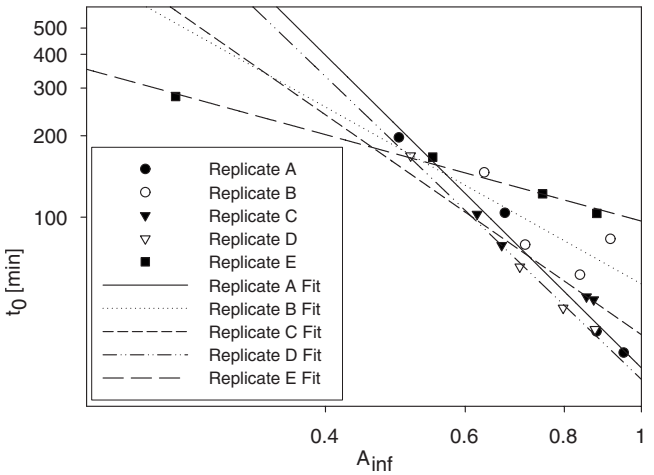

FIG. 10. Double logarithmic plots of t_0 against A_∞ for α-LA, showing all five replicates. The linear fits of $\log_{10}(t_0/\tau)$ versus $\log_{10}(A_\infty)$ for replicates A–E, respectively, give slopes $(-m)$ of −2.90, −1.64, −2.04, −2.81, and −0.80, R^2 values of 0.98, 0.53, 0.98, 0.99, and 0.99, and τ values of 27.7, 56.7, 36.8, 25.2, and 96.7 min, corresponding on these plots to the t_0 values with $A_\infty=1.0$. This gives mean values of $m=2.04\pm0.39$ and $\tau=48.6\pm13.2$ min, where errors are quoted as SEs of the mean.

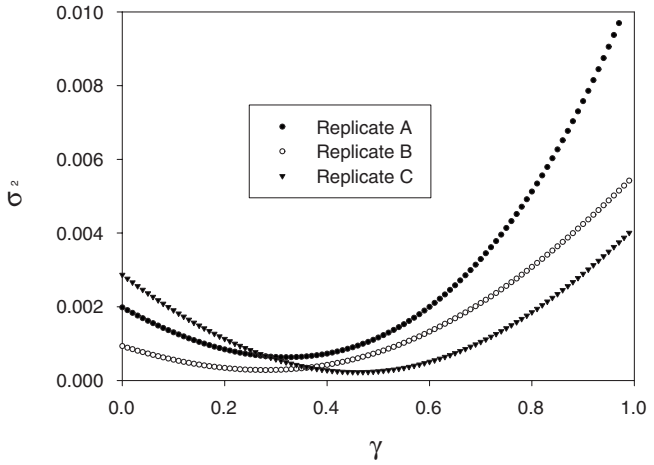

FIG. 11. Plot showing estimated variance between the fitted curve and the scaled TL protein experimental data for values of γ between 0 and 1. Replicates A–C yield γ_{fit} values (i.e., minima of plots) of 0.36, 0.28, and 0.46, respectively, and a mean value of $\gamma_{\text{fit}}=0.37\pm0.09$, where the quoted error is the standard deviation.

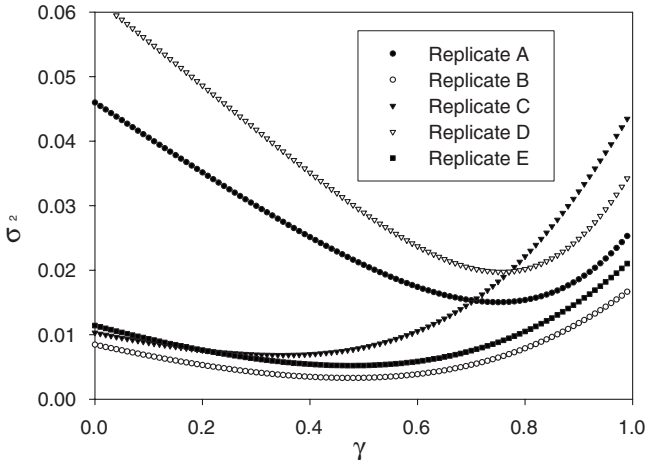

FIG. 12. Plot showing estimated variance between the fitted curve and the scaled α-LA experimental data for values of γ between 0 and 1. Replicates A–E yield γ_{fit} values (i.e., minima of plots) of 0.75, 0.47, 0.35, 0.75, and 0.48, respectively, and a mean value of $\gamma_{\text{fit}}=0.56\pm0.18$, where the quoted error is the SD. Replicates A and D do not scale as well (Fig. 8, sets A1 and D1) and consequently give larger errors overall and also give much higher γ_{fit} values. Therefore, these γ_{fit} values are not as reliable as those for B, C, and E.

than thaumatin, it suggests that these effects are more exaggerated in the TL case, perhaps leading to greater oblateness.

The greater approximation to scaling for the α-LA data makes it difficult to draw any firm conclusions about whether the fits are improved by changing the geometry exponent, γ. Figure 8 (replicates A1–E1) shows the fits to the scaled data by the model and it is difficult to distinguish whether the fit is improved by $\gamma=0.67$ (iso) or γ_{fit} (from Fig. 12). While the latter gives a lower overall error in the fit, we cannot conclude with greater confidence whether $\gamma=0.67$ gives better fits at early times. Therefore, we cannot interpret whether a change in aggregate shape occurs for α-LA. Nevertheless, the same overall trend is observed, that is γ_{fit} is near 0.5.

The less consistent scaling of α-LA could be due to the longer time period of aggregation exaggerating the effect of scaling with t_0. It could also be due to the greater variation of the profiles between experiments. Finally, it may be that different mechanisms are involved for the concentration range examined and therefore scaling would not be expected to occur. In fact, higher concentrations do not scale nearly as well (data not shown), indicating that the latter may be the case.

While the fits by the model are reasonable, improvements could be made for α-LA, for example, by adding several nucleation steps to the model. This would be a reasonable approach since it is thought that α-LA aggregation occurs via a nucleation-dependent pathway.

[1] R. J. W. Truscott, Exp. Eye Res. **80**, 709 (2005).
[2] E. J. Waters, G. Alexander, R. Muhlack, K. F. Pocock, C. Colby, B. K. O'Neill, P. B. Hoj, and P. Jones, Aust. J. Grape Win. Res. **11**, 215 (2005).
[3] J. A. Carver, R. A. Lindner, C. Lyon, D. Canet, H. Hernandez, and C. M. Dobson, J. Mol. Biol. **318**, 815 (2002).
[4] S. L. Flaugh, M. S. Kosinski-Collins, and J. King, Protein Sci. **14**, 569 (2005).
[5] I. Kaur, Y. Ghanekar, and S. Chakrabarti, Int. J. Hum. Genet. **8**, 161 (2008).
[6] Health at a Glance: OECD Indicators, OECD Online Bookshop, The Organisation for Economic Cooperation & Development (OECD), 2007.
[7] P. B. Hoj, D. Tattersall, K. Adams, K. F. Pocock, Y. Hayasaka, R. van Heeswijck, and E. J. Waters, Proceedings of the American Society of Enology and Viticulture, 50th Anniversary Meeting, Seattle, Washington, USA, 2000, pp. 149 (2000).
[8] A. M. Morris, M. A. Watzky, J. N. Agar, and R. G. Finke, Biochemistry **47**, 2413 (2008).
[9] H. Flyvbjerg, E. Jobs, and S. Leibler, Proc. Natl. Acad. Sci. U.S.A. **93**, 5975 (1996).
[10] H. Flyvbjerg and E. Jobs, Phys. Rev. E **56**, 7083 (1997).
[11] E. J. Waters, N. Shirley, and P. J. Williams, J. Agric. Food Chem. **44**, 3 (1996).
[12] R. A. Lindner, A. Kapur, and J. A. Carver, J. Biol. Chem. **272**, 27722 (1997).
[13] T. M. Treweek, R. A. Lindner, M. J. Walker, A. J. Aquilina, C. V. Robinson, J. Horwitz, M. D. Perng, R. A. Quinlan, and J. A. Carver, FEBS J. **272**, 711 (2005).
[14] E. E. Finney and R. G. Finke, J. Colloid Interface Sci. **317**, 351 (2008).
[15] A. F. Schmidt and V. V. Smirnov, Top. Catal. **32**, 71 (2005).
[16] F. F. Abraham, *Homogeneous Nucleation Theory: The Pretransition Theory Of Vapor Condensation* (Academic Press, New

York, 1974).
[17] L. P. Hills and T. O. Tiffany, Clin. Chem. **26**, 1459 (1980).
[18] I. S. Gradshtein and I. M. Ryzhik, *Table of Integrals, Series and Products*, 4th ed. (Academic Press, New York, 1980).
[19] S. Wolfram, *The Mathematica Book*, 4th ed. (Cambridge University Press, Cambridge, 1999).
[20] J. K. Cheung and T. M. Truskett, Biophys. J. **89**, 2372 (2005).
[21] Y. Panyukov, I. Yudin, V. Drachev, E. Dobrov, and B. Kurganov, Biophys. Chem. **127**, 9 (2007).
[22] S. C., Van Sluyter, M. Marangon, S. D. Stranks, K. A. Neilson, Y. Hayasaka, P. A. Haynes, R. I. Menz, and E. J. Waters, J. Agric. Food Chem. (in press).
[23] C. Charron, R. Giege, and B. Lorber, Acta Crystallogr., Sect. D: Biol. Crystallogr. **60**, 83 (2004).
[24] M. Engel, C. van Mierlo, and A. Visser, J. Biol. Chem. **277**, 10922 (2002).
[25] Z. Peng, K. Pocock, E. Waters, I. Francis, and P. Williams, J. Agric. Food Chem. **45**, 4639 (1997).