Published in final edited form as:

J Phys Chem B. 2009 May 14; 113(19): 7020–7032. doi:10.1021/jp8083088.

A Lumry-Eyring Nucleated-Polymerization (LENP) Model of Protein Aggregation Kinetics 2. Competing Growth via Condensation- and Chain-Polymerization

Yi Li and Christopher J. Roberts *

Department of Chemical Engineering, 150 Academy St., Colburn Laboratory, University of Delaware, Newark, DE 19716

Abstract

The Lumry-Eyring with nucleated-polymerization (LENP) model from part 1 (Andrews and Roberts, *J. Phys. Chem. B* **2007**, *111*, 7897–7913) is expanded to explicitly account for kinetic contributions from aggregate-aggregate condensation polymerization. Experimentally accessible quantities described by the resulting model include monomer mass fraction (m), weight-average molecular weight (M_w), and ratio of M_w to number-average molecular weight (M_n) as a function of time (t). Analysis of global model behavior illustrates ways to identify which steps in the overall aggregation process are kinetically important, based on the qualitative behavior of m , M_w , and M_w/M_n vs. t , and based on whether bulk phase separation or precipitation occurs. For cases in which all aggregates remain soluble, moment equations are provided that permit straightforward numerical regression of experimental data to give separate time scales or inverse rate coefficients for nucleation and for growth by chain and condensation polymerization. Analysis of simulated data indicates that it may be possible to neglect condensation reactions if only early-time data are considered, and also highlights difficulties in conclusively distinguishing between alternative mechanisms of condensation even when kinetics are monitored with both m and w_M .

Keywords

non-native aggregation; mathematical modeling; protein stability

1. Introduction

Non-native aggregation commonly refers to the process of forming protein aggregates in which the constituent monomers have significantly altered secondary structure compared to the native or folded state.^{1–3} Aggregates may be soluble or insoluble, with soluble aggregates potentially ranging in size from dimers to so called high molecular weight species ($\sim 10 - 10^3$ or more monomers per aggregate).^{4–6} Formation of non-native aggregates is problematic for protein based pharmaceuticals and other biotechnology products due to increased manufacturing costs, regulatory concerns, and product marketability.^{3,6–8} Non-native aggregates are also implicated in a number of chronic diseases^{9,10} and are suspected immunogenic agents in biopharmaceuticals.^{11,12}

Because non-native aggregation (hereafter referred to simply as aggregation) is typically net irreversible under the conditions that aggregates form, elucidating key mechanistic details that control aggregation kinetics is of general importance for these systems. However, even

*corresponding author; email: E-mail: cjr@udel.edu; tel: 302-831-0838; fax: 302-831-1048.

apparently simple experimental kinetics can be a convolution of multiple stages.² These may include: (partial) monomer unfolding; reversible self association or pre nucleation; nucleation of the smallest irreversible aggregates; and subsequent aggregate growth via chain polymerization and/or aggregate self association or phase separation. Furthermore, many of the kinetically relevant intermediates are often too poorly populated or transient to be directly characterized with available experimental methods.^{2,6,13,14} As a result, proper deconvolution of different stages of the aggregation process requires qualitative and quantitative comparison with mechanistic mathematical models that are couched in experimentally accessible quantities such as mass-percent loss of monomer and time-dependent scattering data.^{2,4,15-19}

A large majority of available mathematical models for aggregation kinetics can be categorized in terms of which stage or stages in the overall aggregation process that they treat explicitly or implicitly. Currently, no available model treats all of the above stages with equally detailed descriptions for natively folded proteins. Rather, most models fall in one of two categories.² Those in the spirit of Lumry-Eyring treatments primarily consider only unfolding and folding in mechanistic detail, and use phenomenological or empirical treatments for assembly steps. Alternatively, polymerization models typically ignore conformational transitions and treat only assembly steps in detail.²

A previous report²⁰ presented a first generation Lumry-Eyring Nucleated-Polymerization (LENP) model that included thermodynamics of monomer conformational stability and prenucleation, along with dynamics of nucleation and of growth via chain polymerization. It is also one of only two models²⁰⁻²² that consider the effects of aggregate (in)solubility on experimental kinetics of monomer loss or soluble-aggregate size distributions. In this context, soluble aggregates are those that are available to consume additional protein monomers,¹⁵⁻²⁴ rather than being defined in terms of a particular size range.²⁵

The previous LENP model did not include detailed treatments of the kinetics and mechanism of aggregate-aggregate coalescence or condensation leading to soluble and/or insoluble aggregates.^{2,20,22} Incorporating details of condensation is important if one is interested in quantifying the resulting aggregate size distribution, but can lead to significant added mathematical complexity.²⁶ This may explain why, for non-native protein aggregation, there are relatively few experimentally tested kinetic models that describe condensation in considerable detail, and those models have typically been system-specific.

For example, Pallitto and Murphy incorporated size-dependent, diffusion limited lateral and end to end association to describe soluble filament and insoluble fibril formation based on *a priori* knowledge of stoichiometry and geometry in A β aggregation.¹⁶ In simpler treatments, Modler et al¹⁷ and Speed et al¹⁸ considered irreversible condensation polymerization to form soluble aggregates, with rate coefficients that were assumed to be independent of polymer size (degree of polymerization). In each case, kinetic models were regressed against time-dependent measurements of one or more aspects of the aggregate size distribution, e.g., weight-average molecular weight¹⁶⁻¹⁸ or z-average hydrodynamic radius.¹⁶ Condensation was determined to be an important or even dominant contribution in each case. However, in each case the models were developed for only a specific protein system, without considering global model behavior.

Furthermore, it is also common practice to fit monomer loss data to models in which condensation is inherently neglected,^{15,23,27,28} even though corroborating structural evidence to support such an assumption may be available in only a fraction of reported cases.^{2,6} Overall, this highlights a need for more general analysis of aggregation kinetics within a mechanistic framework that can easily distinguish which contributions are important, and that

can also provide a means to quantify those contributions by regression against experimental kinetics.

The present report extends the previous LENP model to include explicit and detailed descriptions of condensation. Particular questions that are addressed include: (1) what experimental signatures easily allow one to qualitatively determine whether neglecting condensation^{20,23,24,27-29} is appropriate? (2) can one quantitatively separate contributions from condensation, chain-polymerization, and nucleation without detailed *a priori* knowledge¹⁶ of the association mechanism or aggregate morphology? (3) how sensitive are experimentally accessible kinetics to mechanistic details such as size-dependent vs. size-independent condensation steps? (4) how are the answers for (1) to (3) altered if one considers only early-time data (i.e., only the first few percent loss of monomer)? These questions are important for deconvoluting the effects of chemical additives or protein stabilization strategies on different stages of aggregation,^{2,30-32} inferring mechanistic details of aggregate-aggregate assembly,¹⁶ and in applications such pharmaceutical product stability that typically focus on only small extents of reaction or percent loss of monomer.^{3,6} Finally, this report provides the global behavior of the improved LENP model, and illustrates an application of the model to experimental data using recently reported results for aggregation of α -chymotrypsinogen A (aCgn).⁵

2. Model Description & Derivations

Table 1 summarizes key symbols and definitions used throughout this report. Fig. 1 schematically shows the six stages of nonnative aggregation that are included in the model developed and analyzed here. Stages 1 to 4 are the same as those employed in the previous LENP model.²⁰ Briefly, the six stages in Fig. 1 are: (1) conformational transitions of monomers between folded (F) and unfolded (U) states, with the possibility for stable folding intermediates (I). The monomer conformational state (e.g., F , I , or U) that is most prone or reactive with respect to aggregation is denoted R ; (2) association of R monomers to form reversible prenuclei or oligomers (R_i) composed of i molecules; (3) nucleation of the smallest aggregate species that is effectively irreversible (A_x) by a conformational rearrangement step ($R_x \rightarrow A_x$),^{16,20} (4) growth of soluble aggregates via chain polymerization; (5) soluble aggregate growth due to aggregate-aggregate association such as condensation polymerization;^{5,16} (6) removal of aggregates via phase separation to form macroscopic particles or precipitates.^{21,33,34} In stage 6, all aggregates composed of n^* or more monomers are treated as insoluble.²⁰⁻²²

As in the previous report,²⁰ stages 1 and 2 are assumed to be fast and thus preequilibrated compared to stages 3-6. As a result, only equilibrium constants for unfolding (K_{FI} , etc.) and prenucleation (K_i , $i = 2, \dots, x-1$) appear in stages 1 and 2, respectively. The kinetics of conformational rearrangement as part of nucleation in stage 3 are treated by assuming a concerted, unimolecular rate-limiting step with rate coefficient $k_{r,x}$.²⁰ The balance of rearrangement ($R_x \rightarrow A_x$) and association ($R + R_{x-1} \rightarrow R_x$) steps in stage 3 is treated with a local steady-state approximation. For association, $k_{a,x}$ and $k_{d,x}$ denote forward and reverse rate coefficients. Similar considerations and nomenclature are included for growth via chain polymerization (stage 4).²⁰ R monomers can reversibly self associate with pre-existing soluble aggregates, followed by a conformational rearrangement step that makes monomer addition effectively irreversible. The rate coefficients k_g , k_d , k_r and equilibrium constant K_{RA} in stage 4 are the same as in the earlier LENP model.²⁰ In stage 5, $k_{i,j}$ denotes the rate coefficient for irreversible association of aggregates composed of i and j monomers to form a soluble aggregate of $i + j$ monomers. Stage 6 is effectively instantaneous phase separation of any aggregate that contains n^* or more monomers.

2.1 LENP model equations

The following derivations are based on the reaction scheme in Fig. 1, and employ the same nomenclature as previous work²⁰ to the extent possible here. Characteristic time scales are

defined for nucleation ($\tau_n \equiv \tau_n^{(0)} f_R^{-x} (C_{ref}/C_0)^{x-1}$), growth via monomer addition

($\tau_g \equiv \tau_g^{(0)} f_R^{-\delta} (C_{ref}/C_0)^\delta$), and condensation ($\tau_c \equiv \frac{\tau_c^{(0)} C_{ref}}{C_0}$) (see also Appendix). In these definitions, $f_R = [R]/([M]+[I]+[U])$ is the mole fraction of monomer that is in the aggregation prone conformational state. C_{ref} is a reference state concentration that defines the concentration scale of the standard state for association free energies and equilibrium constants. The respective intrinsic time scales (denoted with superscript (0)) are defined as

$\tau_n^{(0)} \equiv (k_{nuc} K_{x-1} C_{ref}^{x-1})^{-1}$, $\tau_g^{(0)} \equiv (k_g K_{RA}^{\delta-1} C_{ref}^\delta)^{-1}$, and $\tau_c^{(0)} \equiv (k_{x,x} C_{ref})^{-1}$. They are termed intrinsic because they are independent of initial monomer concentration and the free energy of monomer conformational transitions. $k_g \equiv k_d k_r / (k_d + k_r)$ is the effective rate coefficient for chain polymerization, and $k_{nuc} \equiv k_{a,x} k_{r,x} / (k_{d,x} + k_{r,x})$ is that for nucleation.²⁰

The above definitions along with the derivations elsewhere²⁰ and in the Appendix show that although there are numerous parameters in Fig. 1 and Table 1, the assumptions of preequilibration for stages 1 and 2, and local steady state for stages 3 and 4 reduce the total to only seven distinguishable parameters or functions: τ_n and x account for stages 1, 2, and 3; τ_g and δ account for stage 4; and n^* accounts for stage 6. Stage 5 is accounted for by τ_c and $\kappa_{i,j} \equiv k_{i,j} C_0 \tau_c$. $\kappa_{i,j}$ may be a function of i and j , but its (i,j) dependence is uniquely set by the choice of mechanistic model describing size-dependent condensation (see also below and Sec. 2.3). Therefore, there are six adjustable model parameters once the condensation mechanism is selected.

The Appendix provides additional details regarding derivations of the kinetic working equations for monomer and all soluble aggregates. Eqs. A1, A4, and A5 are the dynamic material balances based on Figure 1 and mass action kinetics. They can be rewritten in nondimensional form by defining $\theta = t/\tau_n$, $\beta_{gn} = \tau_n/\tau_g$, and $\beta_{cg} = \tau_g/\tau_c$ to give

$$\frac{dm}{d\theta} = -xm^x - \delta\beta_{gn}m^\delta\sigma \quad (1)$$

$$\frac{da_x}{d\theta} = m^x - \beta_{gn}a_xm^\delta - \beta_{cg}\beta_{gn}\kappa_{x,x}a_x^2 - \beta_{cg}\beta_{gn}a_x \sum_{j=x}^{n^*-1} \kappa_{x,j}a_j \quad (2)$$

$$\left. \frac{da_i}{d\theta} \right|_{\substack{x < i < n^* \\ i \in \{\text{even}\}}} = \beta_{gn}(a_{i-1} - a_i)m^\delta - \beta_{cg}\beta_{gn}\kappa_{i,i}a_i^2 - \beta_{cg}\beta_{gn}a_i \sum_{j=x}^{n^*-1} \kappa_{i,j}a_j + \beta_{cg}\beta_{gn} \sum_{j=x}^{i/2} \kappa_{i-j,j}a_{i-j}a_j \quad (3)$$

When i in Eq. 3 is odd, the right-most summation runs from x to $(i-1)/2$ instead of $i/2$. The dimensionless monomer concentration is $m \equiv ([M]+[I]+[U])/C_0$, with contributions from $[R_i]$ neglected for $K_i C_0 < 1$.²⁰ Dimensionless concentrations for nuclei ($a_x \equiv [A_x]/C_0$) and larger irreversible aggregates ($a_i \equiv [A_i]/C_0$) are similarly defined. The dimensionless condensation rate coefficient is defined as $\kappa_{i,j} \equiv k_{i,j}/k_{x,x}$. If there is no size dependence for condensation, $\kappa_{i,j} = 1$ is a constant for all (i,j) pairs. The model parameters that determine the characteristic

behavior of the solutions to Eq. 1-3 are $[x, \delta, \kappa_{ij}, \beta_{gn}, \beta_{cg}, n^*]$. Eq. 1 is identical to the previous version of the LENP model.²⁰ Eq. 2-3 are more complex than in the earlier model, as they include contributions from condensation (i.e., the terms in which β_{cg} appears). If one neglects condensation ($\beta_{cg} = 0$, $\tau_c \rightarrow \infty$), Eq. 2-3 are equivalent to the condensation free model in ref. 20.

In general, Eqs. 1-3 cannot be solved exactly in analytical form. They can be numerically integrated to simulate the time profiles for monomer concentration on a mass fraction basis (m), as well as the size distribution of aggregates and all associated moments of that distribution. The former quantity is often experimentally accessible by techniques such as size exclusion chromatography (SEC) and field flow fractionation (FFF).^{13,14,35} Indirect measures of m might also be useful, provided they can be properly calibrated against direct measurements.⁶ Examples include dye binding,^{36,37} changes in beta sheet content monitored spectroscopically,^{15,38,39} and turbidity or optical density (provided all aggregates are insoluble).³⁴ In contrast, the detailed or precise size distribution (a_j vs. j) is not usually accessible experimentally. However, techniques employing static or dynamic laser light scattering are able to provide exact or approximate values for the weight-average molecular weight (M_w) and the ratio of M_w to the number-average molecular weight (M_n). The ratio M_w/M_n is the polydispersity of the size distribution.⁴⁰ Using techniques such as SEC or FFF with in-line light scattering detection,^{5,13,35,41} it is also possible to separately measure the weight-average molecular weight of the aggregate size distribution (M_w^{agg}), and to provide a lower bound on the polydispersity of that distribution, M_w^{agg}/M_n^{agg} .⁵

Using the nomenclature here, the weight- and number-average molecular weights of soluble aggregates are

$$\frac{M_w^{agg}}{M_{mon}} = \frac{\sum_{j=x}^{n^*-1} j^2 a_j}{\sum_{j=x}^{n^*-1} j a_j} = \frac{\lambda_2}{\lambda_1} = \frac{\lambda_2}{1-m} \quad (4a)$$

$$\frac{M_n^{agg}}{M_{mon}} = \frac{\sum_{j=x}^{n^*-1} j a_j}{\sum_{j=x}^{n^*-1} a_j} = \frac{\lambda_1}{\sigma} \quad (4b)$$

with M_{mon} denoting the molecular weight of a monomer, and the superscript *agg* indicating that the summations are carried out over all soluble species that do not assay as monomers. For the present case, this makes the lower bound $j=x$ in the summations in Eq. 4. This is expected under conditions where prenuclei are thermodynamically disfavored (low values of $K_i C_0$).²⁰ Equivalent expressions can be derived if one can experimentally resolve smaller aggregates or if it is not convenient to separate monomer contributions in the assay being employed.^{16-18, 20}

Eq. 4 also relates M_w^{agg} and M_n^{agg} to the first and second moments of the soluble aggregate size distribution (λ_1 and λ_2 , respectively).

$$\lambda_1(t) = \sum_{j=x}^{n^*-1} j a_j(t) \quad (5a)$$

$$\lambda_2(t) = \sum_{j=x}^{n^*-1} j^2 a_j(t) \quad (5b)$$

For $n^* \rightarrow \infty$, λ_1 is equal to the fractional monomer loss ($1-m$) at a given time. The zeroth moment of the aggregate size distribution is equal to σ as it was defined previously²⁰ (see also Appendix). Physically, σ is the total number of aggregates per unit volume, scaled by the initial protein concentration on a monomer basis. These moments are not normalized (e.g., σ is not 1). It follows from Eq. 4-5 that the polydispersity of the aggregate size distribution can be expressed as

$$\frac{M_w^{agg}}{M_n^{agg}} = \frac{\lambda_2 \sigma}{\lambda_1^2} \quad (6)$$

2.2 Moment Equations for Soluble Aggregate Conditions

For cases where aggregates remain soluble throughout the course of an experiment (i.e., large n^*), Eqs. 1-3 present an essentially infinite set of coupled, non linear ODEs. These must be repeatedly solved numerically to regress model parameters from experimental data unless one instead employs approximate, analytical solutions. Examples of accurate analytical solutions when condensation can be neglected were the focus of ref. 20 and have been previously used for regression against experimental data.^{4,15,19,42} However, those treatments do not provide a means to describe changes in the aggregate size distribution when condensation is appreciable.²⁰ An alternative approach is to replace Eqs. 1-3 by summing across all aggregate sizes (j) to provide differential equations for the time dependence of m and different moments of the aggregate size distribution (see also Appendix). Under conditions where nucleation is slow compared to aggregate growth, Eqs. A4-A6 are accurate approximations, and with Eq. A1 lead to

$$\frac{dm}{dt} = -x \frac{m^x}{\tau_n} - \frac{\delta}{\tau_g} m^\delta \sigma \quad (7)$$

$$\frac{d\sigma}{dt} = \frac{m^x}{\tau_n} - \frac{\bar{\kappa}_n}{2\tau_c} \sigma^2 \quad (8)$$

$$\frac{d\lambda_2}{dt} = x^2 \frac{m^x}{\tau_n} + \frac{1}{\tau_g} m^\delta (\delta^2 \sigma + 2\delta(1-m)) + \frac{\bar{\kappa}_w}{\tau_c} (1-m)^2 \quad (9)$$

with

$$\bar{\kappa}_n \equiv \frac{\sum_{i=X}^{\infty} \sum_{j=X}^{\infty} \kappa_{i,j} a_i a_j}{\sum_{i=X}^{\infty} \sum_{j=X}^{\infty} a_i a_j} \quad (10a)$$

$$\bar{\kappa}_w \equiv \frac{\sum_{i=X}^{\infty} \sum_{j=X}^{\infty} \kappa_{i,j} (i a_i) (j a_j)}{\sum_{i=X}^{\infty} \sum_{j=X}^{\infty} (i a_i) (j a_j)} \quad (10b)$$

and with the first moment (λ_1 replaced by m . Eq. 10 defines number-average and weight-average $\kappa_{i,j}$ values ($\bar{\kappa}_n$ and $\bar{\kappa}_w$, respectively). In the most general case, $\bar{\kappa}_n$ and $\bar{\kappa}_w$ are not constant because they depend on the aggregate distribution $\{a_j\}$, and this distribution changes as aggregation proceeds. The simplest case mathematically is with $\bar{\kappa}_n$ and $\bar{\kappa}_w$ identical to unity at all times, and occurs when $\kappa_{i,j}$ is independent of size.

Eq. 7-10, along with Eq. 4 provide a numerically tractable means for parameter estimation based on experimental kinetic data for monomer loss and aggregate molecular weight. However they are applicable only when all aggregates remain soluble. If appreciable aggregate phase separation (precipitation) occurs, Eq. 1-3 or simplified limiting cases such as shown below (Sec. 3) and elsewhere²⁰ must instead be used.

2.3 A Simple Size-Dependent Condensation Model

As a test case to explore the effects of a physically plausible size dependence for $\kappa_{i,j}$, a diffusion-limited Smoluchowski model^{43,44} for aggregate association rates was selected (see also, Sec. 3.2).

$$k_{i,j} = 4\pi N_A (D_i + D_j) (R_i + R_j) f \quad (11)$$

N_A is Avogadro's number; D_i and D_j are the translational diffusion coefficients for aggregates composed of i and j monomers, respectively; R_i and R_j are the respective contact radii; and f is a steric factor that accounts for the fact that only a fraction of the surface of the aggregate (s) may be "reactive" with respect to contacting another aggregate. For simplicity, f was assumed to be independent of i and j , and the Stokes Einstein equation was applied for the translational diffusion coefficients. The resulting expression is

$$k_{i,j} = \frac{2N_A k_B T}{3\eta} \left(\frac{1}{R_i} + \frac{1}{R_j} \right) (R_i + R_j) \quad (12)$$

where k_B is Boltzmann's constant, T is the absolute temperature, and η is the viscosity of the solvent. Analogous but more complex expressions can be derived by assuming different aggregate morphologies and/or details of the aggregate-aggregate association process.^{16,45} Using Eq. 12 in the definition of $\kappa_{i,j}$, and making the simplifying approximation that $R_j \sim j$ gives

$$\kappa_{i,j}=0.25\left(i^{-1}+j^{-1}\right)(i+j) \quad (13)$$

$\bar{\kappa}_n$ and $\bar{\kappa}_w$ are calculated based on the time-dependent aggregate size distribution $\{a_j\}$ as it is updated during numerical integration (see also Eq. 10). It is not possible to solve Eq. 7-10 with a size-dependent $\kappa_{i,j}$ unless one assumes or knows the relationship between $\{a_j\}$ and the moments of the distribution. For illustration purposes here, simple discrete probability distribution functions (pdf) were used to describe the aggregate size distribution with mean (μ) and variance (σ_μ):

$$\mu = \frac{\sum_{i=x}^{\infty} i a_i}{\sum_{i=x}^{\infty} a_i} = \frac{\lambda_1}{\sigma} = \frac{M_n^{agg}}{M_{mon}} \quad (14a)$$

$$\sigma_\mu = \frac{\sum_{i=x}^{\infty} i^2 a_i}{\sum_{i=x}^{\infty} a_i} - \left(\frac{\sum_{i=2}^{n^*} i a_i}{\sum_{i=2}^{n^*} a_i} \right)^2 = \frac{\lambda_2}{\sigma} - \left(\frac{\lambda_1}{\sigma} \right)^2 \quad (14b)$$

For under dispersed distributions ($\sigma_\mu < \mu$) the binomial pdf was used, while for equal or overdispersed distributions ($\sigma_\mu \geq \mu$) the negative binomial pdf was used.^{46,47} In each case, the (normalized) pdf is completely specified by the mean and the variance, and these in turn are set by σ , λ_1 (or m) and λ_2 .

Alternative models for the size dependence of $\kappa_{i,j}$ and for pdfs to approximate the aggregate size distribution were also considered. However, a systematic study of each was foregone, as there are many possible alternatives and the purpose of considering a size-dependent $\kappa_{i,j}$ in the present study was only to qualitatively assess the utility and limitations of using the simpler, size-independent $\kappa_{i,j}$ approximation that is commonly used.^{17,18,21}

3. Results & Discussion

Solutions to the LENP model (Eqs. 1-5) were simulated systematically over a wide range of model parameters, including: $x = 2-10$, $\beta_{gn} = 10^{-1}-10^3$, $\beta_{cg} = 0-10^3$, and $n^* = 10$ to 2×10^4 (effectively $n^* \rightarrow \infty$). The value of δ was set as 1 for all simulated results reported below. Results for $\delta > 1$ were tested for selected conditions, and all derivations and resulting working equations below do not require $\delta=1$ to be assumed. Additional parameter values beyond the extremes of the ranges listed above were also tested to confirm that no qualitative changes in behavior were observed by extending the parameter ranges. The initial conditions in each case were $m = 1$, $\sigma = 0$, $a_j = 0$ ($x \leq j < n^*$).

Four main outputs from the model solutions are (each as a function of time): (1) monomer loss kinetics on a mass fraction basis, $m(t)$ and dm/dt ; (2) the zeroth moment or total number concentration of the aggregate size distribution, $\sigma(t)$ and $d\sigma/dt$; (3) weight-average molecular weight of soluble aggregates, M_w^{agg}/M_{mon} ; (4) aggregate polydispersity, M_w^{agg}/M_n^{agg} . As noted in Sec. 2.1, outputs (1), (3), and (4) are directly or indirectly accessible in *in vitro* experiments. Typically, $\sigma(t)$ is not directly accessible via experiment, but its behavior is a useful indicator of qualitatively distinct kinetic regimes²⁰ (see also below).

3.1 Global Model Behavior

Numerical solutions to the LENP model (Eq. 1-5) across a broad range of model parameter values with $\kappa_{i,j} = 1$ displayed qualitatively distinct regimes or types of behavior in terms of experimental observables. Table 2 summarizes the different types or categories of limiting behavior, using nomenclature based on previous reports.^{20,21} The type of qualitative behavior the model exhibits is dictated mathematically by the values of the five key dimensionless groups or parameters noted above (n^* , x , δ , β_{gn} , β_{cg}). Figures 2 and 3 illustrate the qualitative behaviors in terms of $m(t)$, $\sigma(t)$, M_w^{agg}/M_{mon} vs. $(1-m)$, and M_w^{agg}/M_n^{agg} vs. $(1-m)$. Each of these quantities except σ can be experimentally determined quantitatively or semi quantitatively. The behavior of σ is included because it provides insight into the behavior of $m(t)$ in each case. In Figures 2 and 3, t is scaled by t_{50} in order to more easily compare profiles with greatly different absolute time scales; t_{50} is defined by $m(t = t_{50}) = 0.5$.

In Table 2, the scaling exponents correspond to limiting behaviors of the effective or observed rate coefficient for monomer loss (k_{obs}) and apparent reaction order (ν), defined by

$$\frac{dm}{dt} \cong -k_{obs}m^\nu \quad (15)$$

when $m(t)$ is considered over multiple half lives.²⁰ The scaling relationships were derived previously²⁰ for most entries in Table 2, and are included here for completeness when the new features are presented below. The primary new results are for the behavior of M_w^{agg}/M_{mon} when comparing conditions where condensation is negligible or appreciable. The key features of types **Ia**, **Ib**, **Ic**, **II**, and **IVa/IVb** are briefly reviewed below. Type **III** occurs only if aggregate solubility limits are reasonably large,²¹ and is not reviewed further here. For reference, Figure 4 provides illustrative state diagrams that show ranges of model parameter values over which each kinetic type occurs. Each choice of parameter values for simulated profiles in Fig. 2 and 3 correspond to a state point in Figure 4.

Type **Ia** denotes cases in which high molecular weight soluble aggregates form via a combination of nucleated-chain polymerization and condensation polymerization, and the rates of condensation are similar to or much greater than those for chain polymerization. Characteristic features of type **Ia** kinetics include: all aggregates remain soluble; $\nu \geq 2$ (Figs. 2A, 3A), k_{obs} scaling with C_0 to at least the first power, M_w^{agg}/M_{mon} increasing as $(1-m)$ raised to a power much greater than 1 (Figs 2B, 3B), and high polydispersity values (Figs. 2D, 3D). The relationships between the scaling parameters for type **Ia** depend on whether chain polymerization slow or fast compared to nucleation (low or high β_{gn} , respectively). In either case, σ shows a rapid initial increase, but declines rapidly before m declines much below 1 (Figs. 2C, 3C). This occurs because condensation rapidly decreases the number concentration of aggregates, as each condensation step consumes two aggregates (a_i and a_j) but produces only one (a_{i+j}). In terms of global model behavior (Figure 4), type **Ia** occurs for $n^* \rightarrow \infty$, and high β_{cg} values for a given value of β_{gn} . The approximate locations of boundaries between different types on the state diagrams are only weakly dependent on x and δ (not shown).

Type **Ib** denotes cases in which all aggregates that form are either insoluble (low n^*) or soluble aggregates grow so rapidly to n^* that they are present at levels that are too low to be easily detectable. Characteristic features of type **Ib** kinetics include: visible precipitates present at low extents of reaction (m near 1); $\nu = x \geq 2$ (Figs. 2A, 3A) and $k_{obs} \sim C_0^{x-1}$; and essentially undetectably low soluble aggregate concentrations (Fig. 2C). Little or no information regarding M_w^{agg}/M_{mon} or polydispersity is accessible because of the low total soluble aggregate

concentrations. In terms of global model behavior (Figure 4), type **Ib** occurs for low n^* , or for larger finite n^* values when values of β_{cg} and/or β_{gn} are large.

Type **Ic** denotes cases in which soluble aggregates nucleate but do not phase separate or grow to much larger sizes on the time scale of monomer loss. Characteristic features of type **Ic** kinetics include: all aggregates remain soluble; $\nu = x \geq 2$ (Figs. 2A, 3A) and $k_{obs} \sim C_0^{x-1}$; low values of M_w^{agg}/M_{mon} and M_n^{agg}/M_{mon} (Figs. 2D, 3D). σ increases monotonically to a relatively large plateau value (Figs. 2C, 3C) because aggregates do not grow by condensation and do not reach solubility limits. In terms of global model behavior (Figure 4), type **Ic** occurs for low n^* or high n^* , provided that β_{cg} and β_{gn} are both small.

Type **II** denotes cases in which soluble aggregates nucleate and then grow predominantly via chain polymerization. Characteristic features of type **II** kinetics include: all aggregates remain soluble; $\nu = \delta \geq 1$ (Figs. 2A, 3A) and $k_{obs} \sim C_0^{(x+1-1)/2}$; M_w^{agg}/M_{mon} scales linearly with $(1-m)$ once m is significantly less than 1 (Figs. 2B, 3B, and discussion below); low polydispersity that depends only weakly on extent of reaction (Figs. 2D, 3D) σ increases monotonically to a plateau value (Figs. 2C, 3C) because aggregates do not grow by condensation and do not reach solubility limits. The plateau value is relatively low because chain polymerization is fast compared to nucleation, and therefore only a small number of nuclei form before the monomer pool is depleted due to chain polymerization. In terms of global model behavior (Figure 4), type **II** occurs for high n^* , with low β_{cg} and high β_{gn} .

When all aggregates remain soluble (limit of large n^*), M_w^{agg} can be formally expressed as

$$M_w^{agg} = \frac{M_w^{agg}}{M_n^{agg}} M_n^{agg} = \left(\frac{M_w^{agg}}{M_n^{agg}} \right) \frac{1-m}{\sigma} \quad (16)$$

The second equality in Eq. 16 follows from Eq. 4b and the identity $\lambda_1 = (1-m)$ for large n^* . Eq. 16 shows that M_w^{agg} is linear in $(1-m)$ with a positive, non-zero slope for conditions where the polydispersity (M_w^{agg}/M_n^{agg}) and the number concentration of aggregates (σ) do not change appreciably as monomers are consumed. Physically, this is the case for type **II** kinetics as summarized above. Analogous but less general relationships were derived phenomenologically in ref. 20. Eq. 16 also applies for types **Ia** and **Ic**, and shows the mathematical basis for the scaling behavior of M_w^{agg} with $(1-m)$ summarized above and Table 2. Eq. 16 is not valid for type **Ib** because M_n^{agg} is not equal to $(1-m)$ once insoluble aggregates form.

Type **IV** (a or b) behavior is effectively an intermediate or transition between limiting case behaviors (types **Ia**, **Ib**, **Ic**, **II**). As such, type **IV** behavior has some features in common with each of the limiting case behaviors that bound it in the model parameter space. Type **IV** is included for the sake of completeness in Table 2 and Figures 2 - 4. The subcategories (**IVa** vs. **IVb**) distinguish which limiting-case behaviors bound the type **IV** transition region in the state diagrams below. For reference, type **IVb** is mathematically equivalent to what was simply termed type **IV** behavior in the earlier LENP model.²⁰

Aggregation kinetics for α ,¹⁶ phosphoglycerate kinase,¹⁷ and P22 tailspike¹⁸ have each been successfully described by models that are similar to or formally the same as type **Ia**. α -chymotrypsinogen A displays type **II** behavior under sufficiently acidic (pH < ca. 4) and low ionic strength buffer conditions,^{5,15,19} but shifts to **Ia** behavior at higher ionic strengths,⁵ or shifts to **Ib** behavior at pH closer to neutral (unpublished results). Illustrative data for aCgn are included explicitly in Sec. 3.3.

A number of experimental systems qualitatively behave like type **Ib** or **IVb** (or **III**, see ^{ref. 39}), in that aggregates precipitate,^{26,34,42,48} but to best of our knowledge only bG-CSF has been explicitly modeled as type **Ib** or **III**, and shown to exhibit the quantitative scaling behaviors listed in Table 2.²¹ Unfortunately, it is often the case that published reports do not explicitly indicate whether and/or when precipitation was observed during the course of measurements of $m(t)$. Therefore it is difficult to determine whether additional systems may be well-described by the LENP model with finite n^* . To the best of our knowledge, no previous models other than the direct precursors to this work²⁰⁻²² explicitly account for the effects of aggregate insolubility on monomer loss kinetics or soluble aggregate size distributions.

Finally, Eq. 1-3 allow simulation of the complete aggregate size distribution, as shown in Fig. 5A-B (conditions same as in Fig. 2). Evolution of the aggregate size distribution as monomer loss progresses is illustrated for type **II** ($\beta_{cg} = 0$, Fig. 5A) and type **Ia** ($\beta_{cg} = 10$, Fig. 5B). As expected, condensation results in a broader size distribution and decreased total number of aggregates. If one uses only moment based kinetic equations (e.g., Eq. 7-10 in the present case), it is necessary to assume the relationship between the aggregate size distribution and the particular moments. Sec. 2 described a simple way to estimate the aggregate size distributions from moment based simulations including only zeroth, first, and second moments. Fig. 5C shows that the resulting size distributions are semi quantitatively in agreement with those from the full model (Eq. 1-3, Fig. 5A) under conditions where condensation is negligible. Comparing Fig. 5D with Fig. 5B shows that the moment-based simulations correctly predict that the distribution greatly broadens with time when condensation is appreciable. However, there are qualitative differences in the shape of the distributions from the full model that cannot be captured without assuming a more complex form for the underlying distributions. This highlights a potential limitation of moment-based models if only a limited number of moments are experimentally accessible (see also discussion below).

3.2 Parameter Estimation with the LENP model

For aggregates that remain soluble and are able to grow rapidly compared to nucleation, Eqs. 7-10 combined with Eq. 4 provide a computationally simple means to quantify separate characteristic time scales of nucleation, chain polymerization, and condensation using data regression against $m(t)$ and $M_w^{agg}(t)$ simultaneously.

To assess the accuracy of τ_n , τ_g , and τ_c values regressed with moment equations, simulated $m(t)$ and $M_w^{agg}(t)$ data over a common time range ($4 \times t_{50}$) were generated using Eqs. 1-5 with $\beta_{gn} = 1000$, $x = 6$, $\delta = 1$, and $\kappa_{i,j} \equiv 1$ ($\kappa_n = \kappa_w = 1$), with β_{cg} systematically increased from 0 to 10. Only data points at selected time intervals were used for regression, so as imitate typical experimental data without *in situ* measurements. The results below do not change substantially if a larger number and finer spacing of data points are used. The simulated data sets were nonlinearly regressed against Eqs. 7-9 and Eq. 4.

As a test of whether models that neglect condensation can reasonably fit data in which condensation is appreciable, The same simulated data were also regressed with $\tau_c \rightarrow \infty$; Therefore, fitting only τ_g and τ_n . Furthermore, simulated data sets from Eq. 1-5 were truncated at successively smaller extents of reaction (i.e., “early time” data only), and regression vs. Eq. 7-9 was repeated. The latter two cases help to address the question of whether $\{m, M_w^{agg}\}$ data can reliably differentiate between aggregation models that do not include condensation steps, depending on whether one uses data over multiple half lives^{5,16-20} or only under early-time conditions.^{23,28,49}

Figure 6 compares The regressed time constant values ($\tau_{i,fit}$, $i = g, n, c$) to the true values ($\tau_{i,true}$, $i = g, n, c$) for The cases described above. The 95% confidence intervals of the fitted parameters and coefficient of determination (R^2) are included in Fig. 6 to illustrate the quality

of the fit in each case. The size and distribution of residuals were also examined to evaluate the quality of each fit (not shown), and were found to be consistent with the magnitude of confidence intervals and R^2 values reported below. The model parameters x and δ are necessarily integers in the LENP model, and so were held constant to avoid unnecessary complications of working with mixed-integer regression. Instead, The values of x and δ were systematically varied over physically plausible ranges ($x \geq 2$, $\delta \geq 1$) and regression of τ_n , τ_g , τ_c was repeated for each pair of x and δ values.

The best-fit results in Fig. 6 are for $\delta = 1$, as all other δ values produced clearly inferior fits (not shown). However, fits with different values of nucleus stoichiometry (x) were not statistically distinguishable unless very large x values ($> ca. 10$) were used. The large- x fits were clearly inferior to the small- x fits, but it was not possible to further distinguish a best-fit x value. This is not unexpected based on previous analysis that showed reliable determination of x values required kinetic data over a relatively wide range of initial protein concentrations (C_0).²⁰ For concreteness, the results in Fig. 6 are for $x = 6$, the same value of x used to generate the simulated data from Eq. 1-3. More generally, this result highlights inherent difficulties in determining nucleus size from data regression vs. kinetic models when the data are available at only one or a small range of C_0 values.

The results in Fig. 6A show that regression against Eq. 7-9 provides accurate parameter values for a given set of $m(t)$ and $M_w^{agg}(t)$ data. This includes conditions where condensation is negligible ($\beta_{cg} \ll 1$) and where it is the dominant mode of growth ($\beta_{cg} \gg 1$). In all cases, the accuracy of fitted parameters was within 5% of the true values, R^2 values were greater than 0.99, and residuals were small and evenly distributed. In contrast, Fig. 6B shows that fitting with a model in which condensation is neglected clearly produced poor fits and inaccurate fitted parameter values under conditions where condensation is appreciable ($\beta_{cg} \sim 1$) or dominant ($\beta_{cg} \gg 1$).

Figure 6C illustrates instead that if one is able to consider sufficiently early-time conditions ($m \ll 1$), it is possible to obtain reasonably accurate values of τ_g and τ_n with a model that neglects condensation. No values of τ_c are shown because $\tau_c \rightarrow \infty$ for the fits in Fig. 6C. The labels above each data set in Fig. 6C indicate the value of m at which the data were truncated for fitting. The truncation m value for a given data set was selected as the point at which the polydispersity first rose above a threshold value of $M_w^{agg}/M_n^{agg} = 1.1$ (cf. Fig. 2D and discussion below). The results in Fig. 6C are perhaps not surprising because the initial conditions considered here are ones in which aggregates are not present, and because condensation rates are proportional to the square of the total aggregate concentration (i.e., σ^2) while chain polymerization rates are linear in σ . Thus, condensation rates do not become appreciable until larger amounts of monomer have been consumed to create new aggregates. One can reach the same conclusion via an analytical perturbation solution (results not shown), such as applied previously to a condensation-free model.²³ The above arguments notwithstanding, even with early-time data it is not possible to deconvolute τ_g and τ_n unless both $m(t)$ and $M_w^{agg}(t)$ data are employed.

In practical terms, it is unlikely that one will know *a priori* whether experimental data are collected for sufficiently early times to assure condensation can be neglected. The results in Fig. 6C, when compared to those in Fig. 2D, support the empirical practice of considering condensation to be negligible if the sample polydispersity remains relatively low

$(M_w^{agg}/M_n^{agg})^{1.1-1.2}$.^{4,5} The results in Fig. 2C suggest an additional criterion for neglecting condensation is that M_w^{agg} scales linearly with $(1-m)$. Ideally, however, it seems most prudent to instead consider models that include growth via both monomer addition and aggregate-aggregate condensation when attempting to regress accurate and mechanistically sound

parameter values from experimental kinetics. An example of this approach applied to experimental data for aCgn aggregation is provided below (Sec. 3.3).

For simplicity, all preceding examples in this section used only the case of size-independent rate coefficients for condensation ($\kappa_{i,j} = 1$). From a practical standpoint, it also is often convenient to assume size-independent condensation so as to reduce the computational burden and complexity of models for regression.^{17,18} Furthermore, it is not clear *a priori* that typical experimental kinetic measurements provide sufficient information to reliably distinguish between different condensation-mediated growth mechanisms. This motivates the question, can experimental $m(t)$ and $M_w^{agg}(t)$ data robustly distinguish between different models for condensation-mediated growth?

In order to address this question, Eq. 7-10 were solved with a simple diffusion-limited Smoluchowski model for $\kappa_{i,j}$ (cf., Section 2) to provide simulated kinetic data that were then regressed against Eq. 7-9 with the size-independent condensation model used above. Illustrative results are shown here for simulated data (size-dependent $\kappa_{i,j}$) with $\beta_{gn} = 1000$, $\beta_{cg} = 1, 10, 20$. Figure 7A shows results for $\beta_{cg} = 20$. The size-independent model provided excellent fits to size-dependent simulated data in all cases, with $R^2 > 0.99$ and small, evenly distributed residuals (not shown). Despite the seemingly high quality fit for m and M_w^{agg} in Fig. 7A, the true value of $\bar{\kappa}_n$ increases dramatically as aggregation proceeds, although $\bar{\kappa}_w$ remains reasonably close to 1 throughout (data not shown). Thus, although the size-independent model fits the simulated $\{m, M_w\}$ data well to within the precision of typical experimental data, the fitted value for τ_c is only a rough approximation to its true value.

Fig. 7B further shows that for $\beta_{cg} = ca. 10$ or higher, deviations are found not only in τ_c , but in all three fitted parameters (τ_g, τ_n, τ_c). Thus, although the fits appeared to be good in all test cases, the fitted values of (τ_g, τ_n, τ_c) were inaccurate except when condensation was not dominant over chain polymerization ($\beta_{cg} \sim 1$ or smaller). The last two columns in Fig. 7B are for fits using a size-independent model of condensation, but with data truncated at low extents of reaction. In this case, accurate (τ_g, τ_n, τ_c) were obtained even when condensation is dominant (high β_{cg}). Intuitively, this is reasonable because at low extents of reaction the aggregate size distribution will lie relatively close to the nucleus size (x), and the assumption that all $k_{i,j}$ values are the same as $k_{x,x}$ is reasonable.

The above results clearly illustrate that aggregation kinetics monitored experimentally in terms of m and M_w can qualitatively identify whether condensation steps are appreciable, but that obtaining good fits to a kinetic model will not necessarily provide fitted parameter values that accurately reflect the true values for the system. Of course, true values of model parameters cannot be known *a priori* for an experimental system, and so it would not be possible to statistically distinguish these mechanisms in such a situation. As a result, it cannot be generally concluded that m and M_w kinetic data on their own will be sufficient to conclusively distinguish between alternative models for aggregate condensation. Preliminary results (not shown) indicate that this limitation might be overcome if one can experimentally measure higher moments of the distribution, as well as if one can accurately quantify sample polydispersity. In practice, this may remain an outstanding challenge because these quantities are difficult if not impossible to accurately quantify with currently available commercial equipment for the typical size ranges of soluble protein aggregates ($\sim 1 - 10^2$ nm). Qualitatively, however, it may be possible to distinguish between different condensation mechanisms with information regarding aggregate morphology. For example, different types of condensation mechanisms may result in aggregates with different characteristic fractal structures.⁵¹ In such cases, this argues for the importance of using additional data, such as aggregate structure or morphology, when elucidating mechanistic details of aggregation.^{16,51}

3.3 LENP model applied to aggregation of aCgn

Figure 8 illustrates fits of the LENP model (Eq. 7-10) to experimental aggregation kinetics for α -chymotrypsinogen A (aCgn) monitored by size exclusion chromatography with inline static laser light scattering.⁵ The data are from two different solution conditions (summarized in the figure caption; additional details in ref. 5), and are plotted in the same format as Figures 2 and 3. In both cases the aggregates are soluble throughout the experimental time scale, and therefore $n^* \rightarrow \infty$ for fitting with the LENP model. As was done in section 3.2, τ_n , τ_g , and τ_c were regressed for a range of integer values of δ and x to obtain the best least-squares fits to $m(t)$ and $M_w^{agg}(t)$ data simultaneously. The best-fit values for each case, along with 95% confidence intervals are given in the caption to Figure 8.

Qualitative comparison with Fig. 2 and 3 shows that the selected conditions correspond to type **II** (squares) and type **Ia** (triangles) behavior. The qualitative features for the type **Ia** conditions cannot be produced without including condensation steps in the model (stage V, Fig. 1): for example, the pronounced upturn of M_w^{agg} vs. $1 - m$ in Figure 8B, and a concomitant, large increase in polydispersity⁵ (results not shown here). In quantitative terms, the best fit parameter values give $\beta_{gn} \sim 10^3$ in both cases. They give $\beta_{cg} \sim 10$ and $\beta_{cg} \ll 1$, respectively, for the type **Ia** and **II** cases. These results are qualitatively and quantitatively consistent with the analysis and discussion in section 3.1. Finally, the different experimental conditions for aCgn in Figure 8 correspond to aggregates with qualitatively different morphology; the aggregates for the type **II** conditions in Figure 8 are linear polymers,^{4,5} while those for the type **Ia** conditions are more globular and compact.⁵ These morphological differences are consistent with qualitative differences in growth mechanisms for limiting cases **Ia** and **II** in the LENP model. However, they do not provide sufficient information to discern additional details of the condensation mechanism (e.g., size dependent vs. size-independent $k_{i,j}$). A more global search of solution conditions that give rise to behaviors other than types **II** and **Ia** for aCgn is currently underway, and will be included as part of a future report.

4. Summary

This report presents an LENP model of nonnative protein aggregation that explicitly includes the contributions of aggregate-aggregate association or condensation. The model improves upon the previous LENP model²⁰ while maintaining its strengths and ability to capture a wide variety of experimental behaviors. The global behavior and application to simulated data are illustrated primarily using a size-independent condensation mechanism similar to that employed previously,^{17,18} and to a lesser extent using a simple Smoluchowski, diffusion-limited condensation mechanism. Illustrative examples are also included via application of the LENP model to experimental aggregation kinetics of α -chymotrypsinogen A.⁵

The results illustrate a number of ways to qualitatively determine whether soluble aggregate growth occurs via chain polymerization, aggregate-aggregate condensation, or a combination of both. It is shown that this assessment is easily done by measuring both monomer loss (or mass percent conversion to aggregate) and weight-average molecular weight when monitoring aggregation kinetics. When high molecular weight aggregates remain soluble, moment-based kinetic equations provide a means to quantitatively separate the time scales or inverse rate coefficients for nucleation (τ_n), growth by chain polymerization (τ_g), and condensation (τ_c). This requires time dependent data on aggregate molecular weight M_w , and cannot be done with only data for monomer concentration m . However, even regression against both m and M_w is not necessarily sufficient to distinguish between alternative models for condensation. Use of early time data to provide accurate values of τ_n and τ_g was also evaluated and found to provide reasonable estimates even when details of a condensation mechanism are unknown. The current LENP model is also easily adaptable to include more complex aggregate growth mechanisms.

Acknowledgements

Financial support from Merck & Co. (YL) and the National Institutes of Health (CJR; grant no. R01 EB006006) is gratefully acknowledged.

5. Appendix

The dynamic material balances of monomer (m), nuclei (a_x) and larger aggregates ($a_i, i > x$) for the reaction scheme in Fig. 1 are given by Eq. A1-A3, assuming each step in Fig. 1 is an elementary reaction obeying mass-action kinetics, and stages 1 and 2 are pre-equilibrated. Rate coefficients and equilibrium constants are defined in Fig. 1 and are consistent with more detailed descriptions given in ref. 20.

$$\frac{dm}{dt} = -xm^x/\tau_n - \delta m^\delta \sigma / \tau_g \quad (\text{A1})$$

$$\frac{da_x}{dt} = m^x/\tau_n - a_x m^\delta / \tau_g - k_{x,x} C_0 a_x^2 - \left(a_x \sum_{j=x}^{n*-1} k_{x,j} C_0 a_j \right) \quad (\text{A2})$$

$$\left. \frac{da_i}{dt} \right|_{\substack{x < i < n* \\ i \in \text{even}}} = (a_{i-\delta} - a_i) m^\delta / \tau_g - k_{i,i} C_0 a_i^2 - a_i \sum_{j=x}^{n*-1} k_{i,j} C_0 a_j + \sum_{\substack{j=x \\ i-j \geq x}}^{i/2} k_{i-j,j} C_0 a_{i-j} a_j \quad (\text{A3})$$

Eqs. A1-3 are similar to expressions that were derived previously²⁰ except that terms are included to account for the consumption of nuclei by condensation steps, as well as formation and consumption of other aggregates through condensation. Symbols in the above equations are explained in Section 2.1, and are consistent with previous work.²⁰ The corresponding moment equations follow by taking weighted sums over da_i/dt from $i = x$ to ∞ , along with the model parameters defined in Sec. 2: *Zeroth Moment*:

$$\frac{d\sigma}{dt} = \sum_{i=x}^{\infty} \frac{da_i}{dt} = \frac{m^x}{\tau_n} - \frac{1}{2\tau_c} \left(\sum_{i=x}^{\infty} a_i \sum_{j=x}^{\infty} \kappa_{i,j} a_j + \sum_{i=x}^{\infty} \kappa_{i,i} a_i^2 \right)$$

$$\frac{d\sigma}{dt} \cong \frac{m^x}{\tau_n} - \frac{1}{2\tau_c} \left(\sum_{i=x}^{\infty} \sum_{j=x}^{\infty} \kappa_{i,j} a_i a_j \right) \quad (\text{A4})$$

First Moment:

$$\frac{d\lambda_1}{dt} = \sum_{i=x}^{\infty} i \frac{da_i}{dt} = \frac{d(1-m)}{dt} \quad (\text{A5})$$

Second Moment:

$$\frac{d\lambda_2}{dt} = \sum_{i=x}^{\infty} i^2 \frac{da_i}{dt}$$

$$\frac{d\lambda_2}{dt} = \frac{x^2 m^x}{\tau_n} + \frac{m^\delta}{\tau_g} \sum_{i=x}^{\infty} ((i+\delta)^2 - i^2) a_i + \frac{1}{\tau_c} \left(\sum_{i=x}^{\infty} i a_i \sum_{j=x}^{\infty} \kappa_{i,j} j a_j + \sum_{i=x}^{\infty} \kappa_{i,i} (i a_i)^2 \right)$$

$$\frac{d\lambda_2}{dt} \cong x^2 \frac{m^x}{\tau_n} + \frac{m^\delta (\delta^2 \sigma + 2\delta \lambda_1)}{\tau_g} + \frac{1}{\tau_c} \left(\sum_{i=x}^{\infty} i a_i \sum_{j=x}^{\infty} \kappa_{i,j} j a_j \right) \quad (\text{A6})$$

Eq. A4 and A6 are approximate only in that they neglect the terms $\sum_{i=x}^{\infty} a_i^2$ and $\sum_{i=x}^{\infty} (i a_i)^2$ respectively. These terms are due to the self association reaction $a_i + a_i \rightarrow a_{i+i}$ where two same-sized aggregates are consumed, and are negligible when the aggregate size distribution is not close to monodisperse, as is the case when nucleation is slow compared to growth via chain or condensation polymerization.

References

- (1). Fink AL. Folding & Design 1998;3:R9. [PubMed: 9502314]
- (2). Roberts CJ. Biotechnology and Bioengineering 2007;98:927. [PubMed: 17705294]
- (3). Chi EY, Krishnan S, Randolph TW, Carpenter JF. Pharmaceutical Research 2003;20:1325. [PubMed: 14567625]
- (4). Weiss WF IV, Hodgdon TK, Kaler EW, Lenhoff AM, Roberts CJ. Biophys J 2007;93:4392. [PubMed: 17704182]
- (5). Li Y, Weiss WF IV, Roberts CJ. J Pharm Sci. Submitted
- (6). Weiss WF IV, Young TM, Roberts CJ. J Pharm Sci. 2008
- (7). Cromwell MEM, Hilario E, Jacobson F. AAPS Journal 2006;8:E572. [PubMed: 17025275]
- (8). Wang W. International Journal of Pharmaceutics 2005;289:1. [PubMed: 15652195]
- (9). Dobson CM. Seminars in Cell & Developmental Biology 2004;15:3. [PubMed: 15036202]
- (10). Uversky VN, Fink AL. Biochimica et Biophysica Acta, Proteins and Proteomics 2004;1698:131.
- (11). Rosenberg AS. AAPS Journal 2006;8:E501. [PubMed: 17025268]
- (12). Purohit VS, Middaugh CR, Balasubramanian SV. J Pharm Sci 2006;95:358. [PubMed: 16372314]
- (13). Philo JS. Aaps J 2006;8:E564. [PubMed: 17025274]
- (14). Goetz H, Kuschel M, Wulff T, Sauber C, Miller C, Fisher S, Woodward C. J Biochem Biophys Methods 2004;60:281. [PubMed: 15345296]
- (15). Andrews JM, Roberts CJ. Biochemistry 2007;46:7558. [PubMed: 17530865]
- (16). Pallitto MM, Murphy RM. Biophys J 2001;81:1805. [PubMed: 11509390]
- (17). Modler AJ, Gast K, Lutsch G, Damaschun G. J Mol Biol 2003;325:135. [PubMed: 12473457]
- (18). Speed MA, King J, Wang DIC. Biotechnology And Bioengineering 1997;54:333. [PubMed: 18634100]
- (19). Andrews JM, Weiss WF IV, Roberts CJ. Biochemistry 2008;47:2397. [PubMed: 18215071]
- (20). Andrews JM, Roberts CJ. J Phys Chem B 2007;111:7897. [PubMed: 17571872]
- (21). Roberts CJ. Journal of Physical Chemistry B 2003;107:1194.

- (22). Roberts, CJ. Non Native Protein Aggregation: Pathways, Kinetics, and Shelf Life Prediction. In: Murphy, RM.; Tsai, AM., editors. Misbehaving Proteins: Protein Misfolding, Aggregation, and Stability. Springer; New York: 2006. p. 17
- (23). Ferrone F. Methods in Enzymology 1999;309:256. [PubMed: 10507029]
- (24). Oosawa, F.; Asakura, S. Thermodynamics of the Polymerization of Proteins. Academic Press; London: 1975.
- (25). Mahler HC, Friess W, Grauschopf U, Kiese S. J Pharm Sci. 2008
- (26). Ramkrishna, D. Population Balances: Theory and Applications to Particulate Systems in Engineering. Vol. 1st edition. Academic Press; New York: 2007.
- (27). Lee CC, Nayak A, Sethuraman A, Belfort G, McRae GJ. Biophys J 2007;92:3448. [PubMed: 17325005]
- (28). Chen SM, Ferrone FA, Wetzel R. Proceedings Of The National Academy Of Sciences Of The United States Of America 2002;99:11884. [PubMed: 12186976]
- (29). Powers ET, Powers DL. Biophys J 2006;91:122. [PubMed: 16603497]
- (30). Chi EY, Krishnan S, Kendrick BS, Chang BS, Carpenter JF, Randolph TW. Protein Science 2003;12:903. [PubMed: 12717013]
- (31). Gibson TJ, Murphy RM. Biochemistry 2005;44:8898. [PubMed: 15952797]
- (32). Kim JR, Gibson TJ, Murphy RM. Biotechnol Prog 2006;22:605. [PubMed: 16599584]
- (33). Chi EY, Kendrick BS, Carpenter JF, Randolph TW. J Pharm Sci 2005;94:2735. [PubMed: 16258998]
- (34). Kurganov BI. Biochemistry (Mosc) 1998;63:364. [PubMed: 9526133]
- (35). Liu J, Andya JD, Shire SJ. AAPS Journal 2006;8:E580. [PubMed: 17025276]
- (36). Bourhim M, Kruzel M, Srikrishnan T, Nicotera T. J Neurosci Methods 2007;160:264. [PubMed: 17049613]
- (37). LeVine H 3rd. Protein Sci 1993;2:404. [PubMed: 8453378]
- (38). Kendrick BS, Cleland JL, Lam X, Nguyen T, Randolph TW, Manning MC, Carpenter JF. J Pharm Sci 1998;87:1069. [PubMed: 9724556]
- (39). Webb JN, Webb SD, Cleland JL, Carpenter JF, Randolph TW. Proc Natl Acad Sci U S A 2001;98:7259. [PubMed: 11381145]
- (40). Hiemenz, PC. Polymer Chemistry: The Basic Concepts. Marcel Dekker; New York: 1984.
- (41). Wen J, Arakawa T, Philo JS. Anal Biochem 1996;240:155. [PubMed: 8811899]
- (42). Roberts CJ, Darrington RT, Whitley MB. J Pharm Sci 2003;92:1095. [PubMed: 12712430]
- (43). Barzykin AV, Shushin AI. Biophysical Journal 2001;80:2062. [PubMed: 11325710]
- (44). Smoluchowski, M. v. Z. Phys. Chem 1917;92:129.
- (45). Sandkühler P. AIChE Journal 2003;49:1542.
- (46). Hilbe, JM. Negative Binomial Regression. Cambridge University Press; Cambridge, UK: 2007.
- (47). Walpole, RE. Probability & Statistics for Engineers & Scientists. Vol. 8th ed.. Pearson, Prentice Hall; Upper saddle River, NJ: 2006.
- (48). Tsai AM, van Zanten JH, Betenbaugh MJ. Biotechnol Bioeng 1998;59:273. [PubMed: 10099337]
- (49). Ignatova Z, Gierasch LM. Biochemistry 2005;44:7266. [PubMed: 15882065]
- (50). Buswell AM, Middelberg APJ. Biotechnology And Bioengineering 2003;83:567. [PubMed: 12827698]
- (51). Meakin P. Annual Review of Physical Chemistry 1988;39:237.

Figure 1.

Reaction scheme with associated model parameters for the six key stages in the LENP model. The steps shown in each panel are treated as elementary irreversible (single arrow) steps, or as pre equilibrated or steady-state (double arrow) when translating them to mass action kinetic equations.

Figure 2.

Illustrative profiles of limiting behaviors produced by the LEMP model under conditions of fast chain polymerization relative to nucleation, based on simulations of Eq. 1-5 with $\beta_{gn} = 1000$, $x = 6$, $\delta = 1$, and $\bar{\kappa}_n = \bar{\kappa}_w \equiv 1$ for different regimes of n^* and β_{cg} . Qualitatively distinct behaviors are labeled according to the text in Section 3. Types **Ia** (solid gray), **IVa** (dash black),

and **II** (both solid black and dotted gray) correspond to $\beta_{cg} = 10, 0.5, 0.05$ and 0 , respectively, with $n^* \rightarrow \infty$. Type **Ib** (dotted black) corresponds to $\beta_{cg} = 10$ and $n^* = 10$. The panels show (A) monomer loss kinetics, (B) M_w^{agg} as a function of the extent of reaction ($1-m$), (C) dimensionless number concentration (σ) of soluble aggregates available for further growth; (D) polydispersity of the aggregate size distribution as a function of ($1-m$). Polydispersity of type **Ib** is not shown, as experimental polydispersity results would be convoluted by the presence of insoluble/precipitating particles under type **Ib** conditions.

Figure 3.

Analogous profiles to those in Figure 2, but under conditions of slow chain polymerization compared to nucleation; $\beta_{gn} = 0.1$, other parameters are the same as in Fig. 2. Types **Ia** (solid gray), **IVa** (dash black), and **Ic** (both solid black and dotted gray) correspond to β_{cg} at 1000, 50, and 0.01 and 0, respectively.

Figure 4.

Kinetic state diagrams illustrating the placement of types **Ia**, **Ib**, **Ic**, **II**, and **IVa/b** within the space of model parameter values ($x = 6$, $\delta = 1$, $\bar{\kappa}_n = \bar{\kappa}_w \equiv 1$ for all points). Panel A: varying β_{gn} and β_{cg} at fixed $n^* \rightarrow \infty$. Panel B: varying β_{cg} and n^* at fixed $\beta_{gn} = 1000$. $n^* \rightarrow \infty$ (denoted as 'inf') is also included for comparison; Different symbols denote different limiting case behaviors illustrated in Fig. 2, 3: **Ia** (open triangle), **Ib** (filled triangle), **Ic** (filled diamond), **II** (open circle). Types **IVa/b** (blank space) are intermediate behaviors that bridge different limiting cases (see also discussion in text).

Figure 5.

Illustrative size distributions of soluble aggregates as a function of the extent of monomer conversion, based on simulations with $x = 6$, $\delta = 1$, $\beta_{gn} = 1000$, and $n^* \rightarrow \infty$. Panels A ($\beta_{cg} = 0$) and B ($\beta_{cg} = 10$) correspond to simulations with Eqs 1-3 and $\bar{\kappa}_n = \bar{\kappa}_w \equiv 1$. Panels C ($\beta = 10^{-4}$) and D ($\beta_{cg} = 20$) correspond to simulations with Eq. 7-10 and size dependent condensation rate coefficients (Eq. 10-14). Labels indicate the monomer mass fraction remaining for a given curve.

Figure 6.

Comparison of values for τ_g (gray), τ_n (white), and τ_c (black) obtained by regression of Eq. 7-9 against simulated experimental data from Eq. 1-3 (see text for additional details). (A)

$\bar{\kappa}_n = \bar{\kappa}_w \equiv 1$ in both simulated data and fits; simulated data span four half lives. (B) same as A, but fits assumed $\tau_c \rightarrow \infty$ to imitate condensation free models. (C) same as B, but with simulated data sets truncated at the extent of reaction indicated by the label beside each set of bars. Error bars represent 95% confidence intervals from nonlinear least squares fits.

Figure 7.

(A) Representative simulated aggregation kinetics (symbols) with size dependent condensation (Eq. 7-14, $\beta_{gn} = 1000$, $\beta_{cg} = 20$, $x = 6$, $\delta = 1$); curves are fits to the size-independent model (Eq. 7-9, with $\bar{\kappa}_n = \bar{\kappa}_w \equiv 1$). (B) comparison of fitted τ_g (gray), τ_n (white), and τ_c (black) values from the size-independent model versus the true values, based on results analogous to panel A but for a range of β_{cg} values. Asterisks indicate simulated data sets that were truncated at $m = 0.95$ before regression (see also details in text).

Figure 8. (adapted and reproduced with permission from ^{ref. 5}) Illustrative fits of the LENP model to two cases of experimental aggregation kinetics for aCgn. For both cases, the protein concentration (c_0) is 1 mg mL⁻¹ aCgn, and buffer conditions are pH 3.5, 10 mM sodium citrate buffer. The conditions differ in terms of incubation temperature and NaCl concentration: 60 °C with no NaCl (squares); 50 °C with 0.1 M NaCl (triangles). The curves are best-fits from least-squares regression vs. Eq. 7-10 with a size-independent condensation mechanism. The best-fit parameter values⁵ for the first data set (squares) are $x = 3$, $\delta = 1$, $\tau_g = 0.1 \pm 0.01$ min, $\tau_n = 10^3 \pm 10^2$ min, and $\tau_c > 10^{12}$ min; the corresponding parameter values for the second data set (triangles) are $x = 3$, $\tau_g = 0.8 \pm 0.1$ min, $\tau_n = 500 \pm 200$ min, and $\tau_c = 0.1 \pm 0.01$ min. Panels

A and B show the same data in two different formats, for easier comparison of the qualitative features from simulated data in Figures 2 and 3. The open symbols in panel A are M_w^{agg} values from light scattering; the filled symbols are the corresponding m values from chromatography. Details of the experimental protocols are given elsewhere.⁵

Table 1

List of key symbols

Name	Definition
A_j	Agg. composed of j monomers ^(a)
A_x	Nucleus ^(a)
a_j	$[A_j]/C_0$
C_0	Initial monomer concentration ^(a)
C_{ref}	Std. state monomer conc. ^(a)
K_{IU}	Eq. const, for $I \leftrightarrow U$
K_i	Eq. const, for $iR \leftrightarrow R_i^{(b)}$
K_{NI}	Eq. const, for $N \leftrightarrow I$
K_{RA}	Eq. const, for $A_j + R \leftrightarrow A_j R^{(c)}$
k_a	Monomer assoc. rate coeff. ^(d)
$k_{a,x}$	k_a for nucleation step ^(d)
k_B	Boltzmann's constant ^(e)
k_d	Dissociation rate coeff. ^(f)
$k_{d,x}$	k_d for $R_{x-1} + R \leftrightarrow R_x^{(f)}$
k_g	Growth rate coeff. ^(d)
k_{nuc}	Nucleation rate coeff. ^(f)
k_{obs}	Observed rate coeff. for monomer loss ^(f)
ν	Apparent reaction order for monomer loss
δ	# monomers added per chain-polymerization step
k_r	Rearrangement rate coeff. ^(f)
$k_{r,x}$	k_r for nucleation step ^(f)
$k_{i,j}$	Condensation rate coeff. ^(d)
$\kappa_{i,j}$	$k_{i,j}/k_{x,x}$
τ_g	Characteristic timescale of chain polymerization
τ_n	Characteristic timescale of nucleation
τ_c	Characteristic timescale of condensation polymerization
N	Native monomer ^(a)
I	Intermediate state (monomer) ^(a)
U	Unfolded state (monomer) ^(a)
R	Reactive monomer ^(a)
f_R	Fraction reactive monomer
n^*	Size at which precipitation occurs
R_i	Reversible oligomer of i monomers ^(a)
R_x	Reversible prenucleus ^(a)
x	Nucleus stoichiometry
m	$([N]+[I]+[U])/C_0$

Name	Definition
M_n^{agg}	Number-avg. aggregate MW ^(g)
M_{mon}	Monomer MW ^(g)
M_w^{agg}	Weight-avg aggregate MW ^(g)
β_{gn}	τ_n/τ_g
β_{cg}	τ_g/τ_c
σ	$\Sigma[A_j]/C_0$
λ_1	1 st moment of soluble agg. size distribution
λ_2	2 nd moment of soluble agg. size distribution
θ	Dimensionless time (t/τ_n)
μ	Mean of a_j/σ distribution
σ_μ	Variance of a_j/σ distribution
$\overline{\kappa}_N$	Number average $\kappa_{i,j}$
$\overline{\kappa}_W$	Weight average $\kappa_{i,j}$
$\tau_g^{(0)}$	τ_g at C_{ref} and $f_R = 1$
$\tau_n^{(0)}$	τ_n at C_{ref} and $f_R = 1$
$\tau_c^{(0)}$	τ_c at C_{ref}

Abbreviations: agg. = aggregate, aggn. = aggregation, conc. = concentration, const. = constant, eq. = equilibrium, MW = molecular weight, unf. = unfolding

(a) [mol/volume]

(b) [(mol/volume)¹⁻ⁱ]

(c) [(mol/volume)⁻¹]

(d) [(mol/volume)⁻¹·time⁻¹]

(e) [energy/K]

(f) [time⁻¹]

(g) [mass·mol⁻¹]

(h) [Kelvin]

(i) [time]

(j) [energy/mol]

Table 2

Summary of key experimental signatures and scaling behaviors for each kinetic type produced by the LENP model. Examples of illustrative profiles are given in Figures 2 and 3. Expanded from Ref. 20

Kinetic Type	$\gamma^{(a)}$	$\nu^{(b)}$	$\alpha^{(c)}$	Polydispersity	M_w^{agg} / M_{mono}
Ia					
Low β_{gr}	x^{-1}	x	x	High	increases strongly and non-linearly with $(1-m)$
High β_{gr}	$x/2$	$x/2+1$	$x/2+1$	High	
Ib	x^{-1}	x	x	Insoluble Agg. / Precipitates Present	
Ic	x^{-1}	x	x	Low	ca. constant ($\approx x$) versus $(1-m)$
Id	Similar to Ic with small increase of polydispersity and M_w^{agg}				
II	$(x + \delta - 1)/2$	δ	$(x + \delta)/2$	Low	linear in $(1-m)$
III^(e)	x^{-1}	$x, 1$	$x, 1$	Insoluble Agg. / Precipitates Present	
IVa	Early time ($< \text{ca. } t_{50}$) similar to Ic (low β_{gr}) or II (high β_{gr})				
IVb	Early time ($< \text{ca. } t_{50}$) similar to II ; precipitates present at longer times.				
Type	Physical Scenario				
Ia	Low β_{gr} : nucleation-controlled monomer loss, condensation-dominated growth High β_{gr} : slow nucleation, competing condensation and chain polymerization				
Ib	polymerization to low- M_w , soluble aggregates with immediate precipitation				
Ic	nucleation only; negligible chain polymerization, condensation, or precipitation				

Type	Physical Scenario
I_d	nucleation with slow chain polymerization; negligible condensation or precipitation
II	slow nucleation, fast chain polymerization; negligible condensation or precipitation
III	type Ib with aggregates saturated at (finite) solubility limits
IV_a	slow nucleation; condensation appreciable after $t > ca. t_{50}$
IV_b	slow nucleation; fast polymerization; precipitation appreciable after $t > ca. t_{50}$

(*a*) γ defined by $1/50 \sim k_{obs} \sim (C_0)^\gamma$

(*b*) ν defined by $dm/dr = -k_{obs}m^\nu$ over *ca.* one or more half lives

(*c*) α defined by $k_{obs} \sim fR^\alpha$

(*e*) Ref. 21